INNOVATIONS

Hybrid Data Analytics System for Higher Education: A Proposed Model for System Prototype

Ervin L. Rodriguez
College of Computing and Information Sciences, Northwest Samar State University, Calbayog City,Samar, 6710 Phillipines
Corresponding Email :ervin.rodriguez@nwssu.edu.ph

Abstract: The pandemic transformed the management of work in the lives of people differently, and the internet and online systems have likely become a necessity. The study aimed to develop a hybrid data analytics system for Universities. The said study applied a development type of research method. An Agile development methodology was implemented in the software development process using a Server-side scripting language and a Model View Control(MVC) framework. K-Nearest Neighbor(KNN) classification algorithm concept, through PHP String Scripts and other libraries, was used to automate recommendation functionalities. The storage of the records were securely hosted by a Cloud-based Hosting and Database Server. The result of the study allowed the system to be connected both the client and the personnel through its systematic reminders and email notifications. The accessibility were transformed from local to world wide. Automated recommendations functions and other descriptive data analysis to necessary record transactions were generated according to the needs of the user. Using ISO/IEC 9126 software quality models, a software evaluation were conducted in three (3) different type of users had an overall result interpretation of "highly acceptable" in almost all criteria which generally means that the system met the software quality standards. Through the said system, the transactions of the University will be able to operate 24/7 with the assurance that the people are safe from the present pandemic.

Keywords:1.Information System 2. Human Resource 3. HEI System 4. Decision Support System 5. Descriptive Analytics

1. Introduction

Since the beginning of the year 2020, the Pandemic has transformed the management of work in people's lives differently. Various IT applications have performed seminars, conferences, and other essential meetings at a lower cost. Workers accomplish office work and relevant reports at the most comfortable and safe place at home. The internet and online systems have likely become a necessity.

The new usual way of conducting transactions required administrators, office heads, directors, or even presidents to handle its management in a distant location, including making critical decisions to allow its operations to continue working effectively. However, decision-making insights must still come from an analysis generated from relevant existing data.

According to Peersman (2014), it is essential to maximizethe use of existing data. It begins with a planning process of collecting data by conducting several reviews to understand the degree to whichare handled. On the other hand, there are signs of indeed a sort of

emptiness.Some researchers believe that the glass of Big Data, Data Analytics, and Artificial Intelligence(AI) is half-full and still being gradually filledup. Big Data and AI are enabling companies to expand services beyond their reach and serve customers better if utilized creatively, wisely, and of course, effectively (Davenport and Bean, 2019).

With the advent of the Data Analytics System and its innovations, people have become dependent on easy access and navigation. There are a lot of applications that use data analytics to be more efficient and effective. Analytics Vidhya Content Team (2015) shared 13 excellent applications of data science. The most fundamental application is the Internet Search Application. It allows people worldwide to find and locate similarly or, most of the time, accurate information in a wide variety of site locations. The other is Digital Advertising and Recommender Systems, which allow the world of digital marketing to almost spoon-fed customer services to people instantly in almost anywhere and anything.

Meanwhile, private and government schools and universities have also started using data analytics to utilize a better teaching methodology, sustain operational activities, produce and maintain student to being employee productivity. The boundless possibilities of data analytics are already in the grasps of the people in the world. Conclusively, almost everything in the world aims to enhance its operation and traditional processes.The generated data is better done through a cloud-based information system and with data analytics.

According to Yass (2017), cloud-based information systems play an essential role in organizational business value and performance. Perceived value was represented by the perceived improvements in information system processes, as indicated by its performance. The sources of information systems' value, accuracy, usability, comparability, relevance, and transparency, which links to the capabilities of the information systems' adapted from, were represented by the following qualifications: automation, information processing, geographical, and tracking capabilities.

The base location of study, presently located in Samar, is a University that molds people to be Engineers, Educators, and Businessmen. Even though the University has an existing and operational outsourced Management Information System (MIS), unfortunately, the MIS office has a hard time maintaining its sustainability. The common problem is implementing new processes in the University Information System (UIS) because the accessibility of the outsourced system is limited to accessing the database only and is currently dependent on the local area network setup. The outsourced system includes many modules, such as Medical, Dental, Archiving, Research, and Extension. Unfortunately, the only modules that are operational are Registrar and Enrolment Modules. Other modules,such as Human Resource modules,were accepted but still require the programmer from the company to revisit and perform the actual customization process inside the campus.

Moreover, the MIS's current system is somehow useful in conducting a day-to-day operation of the University. Still, unfortunately, the said system is incapable of providing comprehensive data analysis support. The management must understand the current and future status of data and how it can support both standard and critical University decisions. Despite the existence of reliable data, some vital decisions were unsuitable and inappropriately made. Most of the data are stored but not correctly utilized.

Information systems likeMIS require a proper product evaluation before the actual acquisition proceeds. Moreover, a standard way of examining the current operation is to plan if it remains consistent in providing the expected output and eventually implemented.

The study utilized ISO/IEC 9126 Software Engineering Product Quality Model.Recently,ISO/IEC 25010:2011 Systems and software Quality Requirements and Evaluation (SQuaRE) is more suitable in all software engineering related softwares. It presents two models that define the characteristics and subcharacteristics relevant to all software products and computer systems. These characteristics and subcharacteristics ensure consistency with terminologies utilized to specify, measure, and evaluate system and software product quality and completeness.

The current operation of the University requires its client to be physically available for a particular transaction. With this current Pandemic situation, both people inside and outside of the University should continue having their transactions online 24/7. In this way, people are safe, and transactions are more efficient.

This paper aimed to determine the problems encountered in the current and future data utilization, and system operation. Moreover, it also aimed to determine what hybrid data analytics system should be designed and developed along with the administration HR core systems. In addition, a system evaluation process was needed to determine the performance of the developed system prototype.

2. Methodology

The proposed study applied a developmental type of research method. It employed a 1-to-1 interview, FGD, and observation to gather data, identify the problems, anddesign and develop asystem prototype. The conduct of software evaluation process utilized a questionnaire based on the ISO/IEC 9126 Software Product Quality Standards. The type of development applies the Rapid Application Development (RAD) in creating the prototype. A Hybrid Data Analytics System wasdesigned and developed. Specifically, the system focused on the core systems of the Program to Institutionalize Meritocracy and Excellence in Human Resource Management (PRIME HRM). Ideally, the said system will have four (4) sub-modules and will be integrated as one (1) Human Resource Management Hybrid Data Analytics System.

Moreover, the program's development used the Server-side Scripting Programming Language, Model View Control (MVC) Framework for rapid growth, and Web Packages or Libraries for data analysis. The said programming language and framework were selected because the researcher believed that the development process would be more straightforward considering previous web development experiences. Moreover, it includes necessary features such as 1) ease ofwriting a program; 2) high-level programming language; 3) fast run time; 4) significant data handling capabilities; 5) flexible language for building new web applications from scratch, and 6) suitable libraries for Data Analytics. To manage the database of the prototype of the proposed system, the researcher used an open-source, cross-platform web server solution stack package and Cloud Database Server.

Furthermore, the researcher used the K-Nearest Neighbor (KNN) Classification Algorithm concept through PHP String libraries and Scripts to execute recommender system functionalities and process data analysis procedures for every core of the PRIME HRM. For Recruitment, Selection, and Placement(RSP) offices, the said algorithm allowed the office to efficiently analyze the existing datasets or records to produce shortlisted applicants for a particular available position offered in the University and other relevant reports. Meanwhile, in the Learning and Development (L&D) office, the capabilities of the said algorithm to analyze the

existing data were also utilized. For example,the recommender system functionality that automates the identification of the employee's training, seminars, or conferences for specific personnel was done quickly. Similarly, the same functionalities were implemented for the necessary data analysis of the existing data for thePerformance Management(PM) and, Recognition and Rewards(R&R) offices. In future development, asimilar concept will be implemented to other necessary modules in the University. Other relevant Integrated Development Environment software applications and tools to develop the proposed system was utilized to secure a user-friendly interface.

Lastly, the design of the system includes Cloud integration. Primarily, the Cloud was used to centralize the access of the system in the University. This phase also allows the system to utilize a more considerable amount of space to store the information equivalent to a costly physical server and eventually provides a massive advantage to the system to handle historical, current, and upcoming data of the University.

Block Diagram

The diagram shows that the Input of the proposed system will be the required records of every office. The Processes block will be the core transactions of a module. The output block shows the determined expected result of every module and sub-modules of each office.

Furthermore, the storage block was included to highlight that the proposed system will be implemented in a cloud environment.

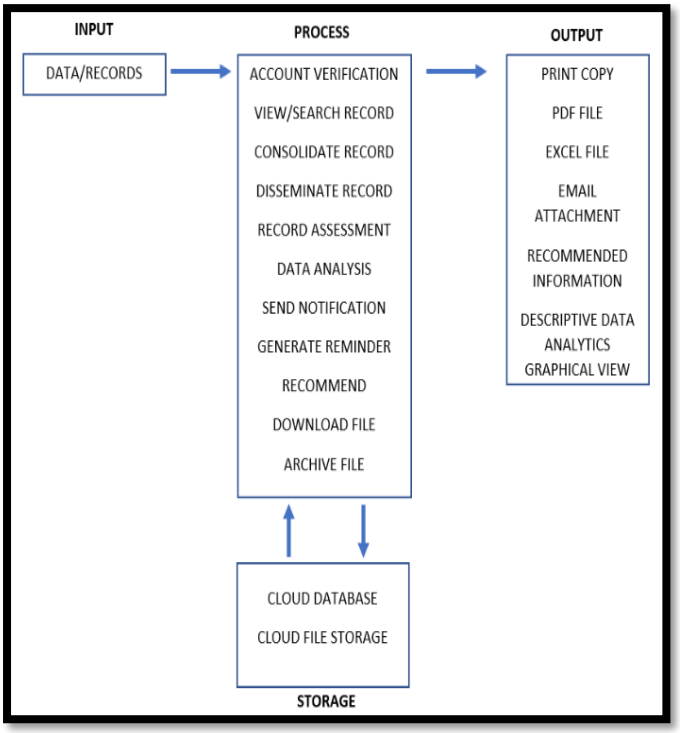

Figure 1. Input, Process, Output, and Storage of the Proposed System

Flowchart

The flowchart shows the necessary procedure in the system development process and system implementation. Furthermore, the proposed system will have different primary functions according to the available offices in the entire University. Every office will have one(1) module and several sub-modules.

After identifying the modules and sub-modules, a prototype will be developed and tested and evaluated for final functionality. If the operation of the module is at least 95% approved, the said module will be uploaded to the acquired Cloud hosting for its process to be available online.

Every module and sub-modules will follow the development and implementation process.

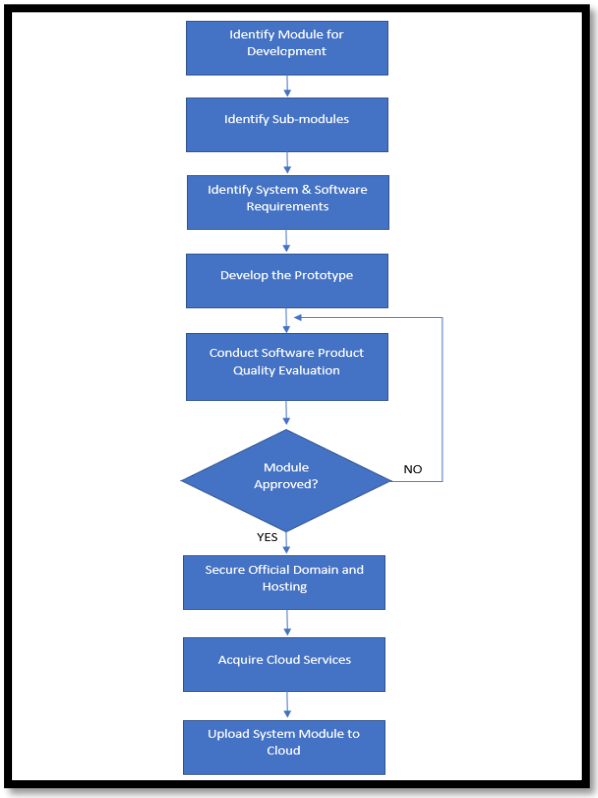

Figure 2. System development and Implementation Process

Materials and Components

The study requires materials for the developed software to be operational. The following are the required materials and components:

Table 1 shows the specification of the cloud services that should be acquired for the system to be operational in the Cloud. The said specifications will assure both the user and the developer to have continuous and sustainable operation.

Table 1. Cloud Service Acquisition

Cloud Service	Minimum Requirement	Recommended
Cloud Web Server	• Fixed Business Type • vCPUs: 2 • RAM: 1024MB • Disk Space: 20GB • Bandwidth: 750GB • Port speed: 100Mbits/sec	• Enterprise Type • vCPUs: 2 or higher • RAM: 2048 or higher • Disk Space: 40GB or higher • Bandwidth: 750 or higher • Port speed: 150Mbits/sec • Plesk/Web Admin Service
Database Management System	• MySQL (MariaDB cloud VM)	• MSSQL Server (Cloud VM)
PHP Version	• PHP 7.1 Version(Cloud VM Compatible)	• PHP 7.3 Version(Cloud VM Compatible)
Security	• SSL	• SSL
Admin Panel	• Plesk/Web Admin Service	• Plesk/Web Admin Service
Technical Support	• Cloud Service Technical Support	• Cloud Service Technical Support

Table 2 shows the sample characteristics, capabilities, or capacities of a particular user of the system. The said Table is required to ensure accountability of the transactions done in every module.

Table 2. User Requirements Sample

User	Capacity	Designation
RSP Personnel	• RSP Designated • Computer Literate • MS Office Application Literate • Internet Operation Literate	• Job Order / Permanent
L&D Personnel	• L&D Designated • Computer Literate • MS Office Application Literate • Internet Operation Literate	• Job Order / Permanent
Teaching & Non-teaching Personnel	• Computer Literate • MS Office Application Literate • Internet Operation Literate	• Temporary / Permanent

Table 3. shows the five (5) point scale, range, and verbal interpretation that serves as the basis for the understanding of the software product quality evaluation result. The said scale will interpret the evaluation result and determine the fulfillment of the corresponding software quality standards criteria of ISO/IEC 9126.

Table 3. 5 Point Scale, Range, and Verbal Interpretation

Range	Verbal Interpretation	
4.21-5.00	Highly Acceptable	If the condition is highly extensive and functioning excellently
3.41-4.20	Acceptable	If the condition is extensive and functioning well
2.61-3.40	Moderately Acceptable	If the condition is lightly extensive and functioning moderately
1.81-2.60	Slightly Acceptable	If the condition is lightly extensive but functioning fairly
1.00-1.80	Not Acceptable	If the condition is not extensive and functioning poorly

3. Results and Discussion

The proposed system has been tested and evaluated by selected users in three (3) Universities of Samar. The activity allowed the researcher to observe several positive feedbacks and significant recommendations.

The system's RSP module was found to be more helpful to speed up the processes and or transactions done by the RSP personnel to accomplish and produce meaningful reports accurately and efficiently. The online conduct of the RSP transactions allows the office to be more reliably productive to its regular works and accomplishments.

On the other hand, the Learning and Development module was also found to be more effective compared to the traditional processes of the said office using conventional tools for producing results. The data analysis and automated functions of theproposed system allowed the personnel to simplify and improve the accuracy and efficiency of the current system. Through the proposed module for L & D, it was realized that the processes that involve necessary information from all the personnel of the University will be done in a more synchronized manner.

Meanwhile, the system's module for the Regular User (Teaching and Non-teaching Staff) were also found to be efficient from the transactions and tasks of the current system. The module provided the regular user an easier way to edit, download, and print an up to date Personal Data Sheet record. The PDS function of the proposed system is more adaptable than a stand-alone hard copy or soft copy PDS file since the said function allows the Human Resource office to be updated about the personnel relevant records. Additionally, online transactions such as Training Records inventory, Training Needs Analysis, and Individual Development Plan was found to have a better advantage.

Figure 3 shows the sample generated descriptive data analytics of the yearly or annual training of the personnel of a particular college. Meanwhile, figure 4 shows the example generated descriptive data analytics of the yearly training analysis in different colleges or departments.

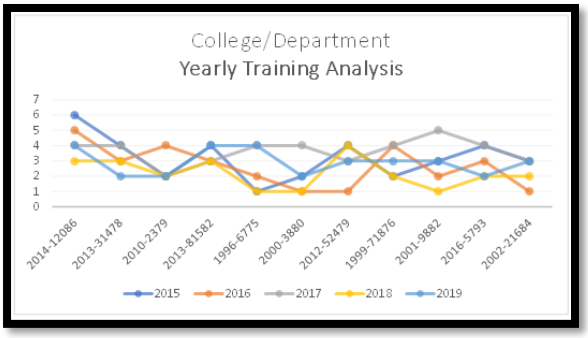

Figure 3. Yearly Training Analysis of a specific college of 2015-2019.

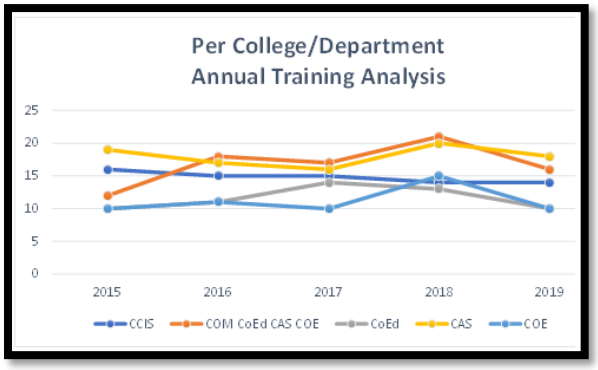

Figure 4. Yearly Training Analysis of different colleges of 2015-2019.

Interpretation of Data

The following responses of the problems encountered were categorized into three namely: 1) Current Data Utilization; 2) Future Data Utilization, and; 3) Present System Operation, and presented throughtables 4, 5, and 6 respectively. Table 4 shows the common problems encountered of the management of the past and present data of the University. Several list of common terminologies were identified in the said list namely: 1) Computation; 2) Consolidation; 3) Safe Keeping or Saving of Records; 4) Dissemination; 5) Duplication; 6) Evaluation; 7) Generation; 8) Inconsistency; 9) Validation; 10) Manual Process; 11) Notifications; 12) Searching; 13) Sharing, and; 14) Sorting.

Table 4. Problems encountered of the Current Data Utilization

Problems Encountered of the Current Data Utilization (*N*=33)	Frequency (f)	Percentage (%)	Rank
MS Excel dependent transactions in record keeping, monitoring, and generation of progress report	33	100%	1
Generating of reports that requires a lot of time of work	30	91%	2
Generating of comparative assessment type of report	29	88%	3

Saving of records by file and folder	28	85%	4.5
Notification or reminders	28	85%	4.5
Current system of faculty records are safe kept by folder but some of the contents were not fully categorized	27	82%	6
IPCR entry validation and computation consistency	26	79%	7.5
Searching history of records	26	79%	7.5
Search and identify student and faculty research projects or thesis that will be helpful to research and extension projects.	25	76%	9
Consolidate personnel trainings and seminars by Category	24	73%	10
IPCR and OPCR management is hard to validate since the entries of the personnel were incorrect.	21	64%	11
Manual checking of personnel who attended a particular training per year	20	61%	13
Manual Computation per record	20	61%	13
Re-encoding of consolidated data	20	61%	13
Computation of NBC Points with regards to seminars or trainings attended and type of training (Local, Regional,etc.)	19	58%	15
Generation of NBC 461 evaluation result based from Local Evaluators	16	48%	17
Sharing of files and or records	14	42%	20
Dissemination of Information with regards to Research and Extension Project activities	13	39%	21
Consolidation of IDP records	10	30%	23.5
Consolidation of reports	10	30%	23.5
Problems Encountered of the Current Data Utilization (*N*=33)	**Frequency (f)**	**Percentage (%)**	**Rank**
Dissemination of research document templates	10	30%	23.5
Reminders or notification for License Renewal or Expired	10	30%	23.5
Duplication of records	8	24%	27
Notification to outsider applicants	8	24%	27
Sharing of files and or records from L & D to RSP is not efficient	8	24%	27
Problem to identify personnel to attend the available training	7	21%	29.5
Problem to identify personnel with training from those without	7	21%	29.5

Sorting of research projects according to classification like regular, or special case	6	18%	31
Computation of tally result of TNA per personnel	4	12%	33
Consolidation of TNA records	4	12%	33
Searching history of applicant records and qualification requirements	4	12%	33

Meanwhile, Table 5 shows theproblems that will be encountered in the future data utilization.

Table 5. Problems that will be encountered in the Future Data Utilization

Problems the will be Encountered in the Future Data Utilization (*N*=33)	**Frequency (f)**	**Percentage (%)**	**Rank**
Comparison of new records to old records	33	100%	1.5
Generating of report from Histrory of records	33	100%	1.5
Inventory of trainings and seminars is not easily traced	31	94%	3
Tracing of IDP record for a particular training	30	91%	4.5
Monitoring Teaching and Non-teaching Training records acquired	30	91%	4.5
Generating of summary of training records for the whole University	29	88%	6
Monitoring of research for submission of terminal report	28	85%	7
Summary of Trainings for college or office and the individual personnel	27	82%	8
Problems the will be Encountered in the Future Data Utilization (N=33)	**Frequency (f)**	**Percentage (%)**	**Rank**
Monitoring Number of completed researches per college	24	73%	10.5
Tracing of submission of reports with signatories or approval consumes time per office	24	73%	10.5
Tracing of IPCR and OPCR records	22	67%	12
Generating of annual reports	21	64%	14
Monitoring of requested supplies for research and extension	21	64%	14

Monitoring of research progress like completed or submitted to publication	21	64%	14
Comparison of published research to list of CHED accredited journals	18	55%	16
Monitoring Number of implemented research	17	52%	17.5
Problem to find and locate records like existing research projects consumes a lot of time	17	52%	17.5
Monitoring Number of approved proposal per college	15	45%	19.5
Monthly Monitoring of research for submission of progress report	15	45%	19.5
Monitoring of teachers per subject and monitoring of part-timer per college	14	42%	21.5
Searching of available teaching personnel for subject Loading	14	42%	21.5
Monitoring the personnel of updating eligibility records	13	39%	23
Tracing of Academic Records of the Employed Personnel necessary for Temporary Position Status of employment	9	27%	24
Detection of NBC Points for personnel with ceiling points	5	15%	25

Moreover, Table 6 also shows the problems encounteredin the systemoperation of the University.

Table 6. Problems encountered in System Operation

Problems Encountered in System Operation (*N*=33)	**Frequency (f)**	**Percentage (%)**	**Rank**
Problem of paperworks	30	91%	1
Records are not centralized	28	85%	2
Monitoring of budget for training	24	73%	3
Problem to compare, control and trace budget allotment per personnel with regards to attending trainings and seminars	19	58%	4
Some important history of data were not accessible	18	55%	5

Outsourced System is not that customizable to fit the transactions of the University	16	48%	6
Tracing of activity records from the previous administration	10	30%	7
Current process is overpowered thus to bypass proper step by step process of the system	7	21%	8
No central mechanism to determine personnel for travel	6	18%	9.5
The office of L & D was not aware of the personnal training of the personnel right until the training was completed	6	18%	9.5
Problem to control or limit the number of training per personnel	4	12%	11
No proper filing of Faculty records from the previous administration	3	9%	12
No proper filing of Certificates submitted from research training	2	6%	13.5
The present system or situation of the University is not institutionalized	2	6%	13.5

The evaluation result of the prototype was found to be functional, reliable, usable,efficient, and accurate. Specifically, the said result was based on the simulation and testing conducted to 15 Teaching Staff, 9 Non-teaching Staff, 3 RSP Personnel, and 3 L & D Personnel from PRIME HRD Office of three Universities of Samar. The software quality evaluation questionnaire was based on the ISO 9126 Software Quality International Standards of ISO/IEC. Each respondent used and tested the proposed system to understand mainly its level of acceptability. A five (5) point scale, range, and verbal interpretation was used to serve as the basis for the understanding of the software quality evaluation result. The said scale is characterized in different variations such as five (5) as Highly Acceptable down to one (1) as Not Acceptable.

The software quality evaluation questionnaire used was collected and interpreted using Weighted Mean and Standard Deviation. Below is the formula used for calculation:

Weighted Mean formula used: $WM = \frac{\sum wx}{\sum w}$

Where:

WM = Weighted Mean

w = weights

x = values

Standard Deviation formula used:

$$S=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

Where:

S = Standard Deviation

$\bar{x}$ = Mean or Average Mean

x = Data or Value

n = Sample Size of Population

Table 7 shows the overall software product quality evaluation result for the Regular Users or the Teaching and Non-teaching personnel of the University. Moreover, Table 8 shows the comprehensive software product quality evaluation for the RSP Module Users. Table 9 shows the overall software product quality evaluation for the L and D Module Users.

Table 7. Regular Users overall result of the software quality evaluation process.

Evaluation criteria	Mean weight	Interpretation
Functionality	4.70	Highly Acceptable
Reliability	4.26	Highly Acceptable
Usability	4.67	Highly Acceptable
Efficiency	4.67	Highly Acceptable
Accuracy	4.74	Highly Acceptable

Table 8. The overall result of the software quality evaluation process for Recruitment, Selection, and Placement (RSP) Users.

Evaluation criteria	Mean weight	Interpretation
Functionality	4.88	Highly Acceptable
Reliability	4.33	Highly Acceptable
Usability	5.0	Highly Acceptable

Efficiency	4.87	Highly Acceptable
Accuracy	4.87	Highly Acceptable

Table 9. The overall result of the software quality evaluation process for Learning and Development (L&D) Users.

Evaluation criteria	Mean weight	Interpretation
Functionality	4.85	Highly Acceptable
Reliability	4.67	Highly Acceptable
Usability	5.00	Highly Acceptable
Efficiency	4.93	Highly Acceptable
Accuracy	5.00	Highly Acceptable

The overall results that reflected to tables 7, 8, and 9 means that the module that was designed and developed proves to have a High Acceptability, since the said modules were done as much as possible customized to users needs.

4. Conclusion and Recommendation

Based on the results and the interpretation of the data, the common problems of the current and future data utilization of the data were answered by the designed and developed prototype of the system. The system operation was improved in the efficiency of the transaction and is highly considered timely amid Pandemic.There was an improvement in the reliability of the data, thus supporting the decision making of the management.A high possibility of a solution to the problems encountered in the current and future data utilization and system operation if the proposed system will be implemented in the University.

The researcher recommends that the evaluation of the software should be re-evaluated using the latest ISO standards of Software Engineering using ISO/IEC 25010:2011 System and Software Product Quality Standards. Moreover, the researcher also recommends that the study to be utilized as a basis for conducting similar studies.

5. References

1. *Davenport, T.H., Bean, R., 2019. NVP - New Vantage Partners. Big Data and AI Executive Summary 2019. Executive Summary of Findings. Available at newvantage.com*
2. *Peersman, G. (2014). Overview: Data Collection and Analysis Methods in Impact Evaluation. Methodological Briefs Impact Evaluation No. 10. UNICEF, September 2014.*
3. *Yass, Ahmed, 2015. The Impact of Cloud BasedIn formation Systemson Organization's Performance. Retrieved on September 10, 2020. Retrievedfrom www.researchgate.net*