Data Mining for the Analysis of Student Assessment Results in Engineering by Applying Active Didactic Strategy

César Baluarte-Araya, Oscar Ramirez-Valdez, Ernesto Suarez-Lopez, Percy Huertas-Niquén
Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú

***Abstract*—To make improvements in the teaching-learning process in educational institutions such as universities, it is necessary to analyse the results obtained and recorded from applying Active Didactic Strategies and, based on this, to propose improvements that will help to achieve the Student Outcomes established for the subject in question; the problem to be solved is thus defined, and the results to be obtained from the analysis are relevant for the improvement of student performance. The objective is to analyse the results of the student assessment, the basis for the calculation of which is based on the recording of the qualification achieved through the performance indicators defined for each criterion, of the competencies involved and aligned with the Student Outcomes of the problems proposed to the student, applying various data mining techniques. Data mining is used to treat large amounts and types of data to obtain hidden information and reveal states, patterns and trends; as well as in Education to study the behaviour of students in terms of their performance. The methodology used for the development of the work is based on the Cross-Industry Standard Process for Data Mining methodological model, which is widely used in data mining projects. The results obtained reveal that the Student's t-test and Snedecor's F-test are highly significant, as well as the determination of the lowest performance indicators in order to plan future improvement actions towards better student performance and achieve a high level of learning. Concluding that if the same teaching and learning process is applied the result will be very similar, therefore, the students have finished learning very well.**

Keywords*—Student outcome; problem based learning; assessment; performance indicator; data mining

I. INTRODUCTION

Considering that the main benefit of Data Mining (DM) is the power to identify patterns and relationships in the gigantic volumes of data found in numerous sources.

Also, other aspects such as the growth of data from educational sources ranging from university admissions, academic information systems, monitoring, assessment, graduation, as well as IT and teaching and learning support platforms is enormous.

The analysis of the data recorded is essential to obtain information, reach conclusions and therefore also to see the correlation in the performance of students that allows to produce multiple results, such as passes, fails, dropouts, grades, among others, and to make decisions for improvement. This is how nowadays the further development of DM to discover the knowledge that may be hidden in databases, has allowed to demonstrate its effectiveness in different contexts, being the educational one the one that is growing in its application giving rise to Educational Data Mining (EDM).

The objective of the research work is to apply EDM to analyse the results of the evaluation of students in the area of Engineering when applying Active Didactic Strategies, which in the case of the study is Problem Based Learning in the course of Electronic Business in the Professional School of Systems Engineering (EPIS) [1] of the Universidad Nacional de San Agustín de Arequipa (UNSA) [2]; and thus discover trends, patterns, behaviours based on the available data and analyse the performance of the students according to the results of the evaluations reflected in the grades obtained, as well as the grades given to the different performance indicators determined for each defined criterion and included in the Reports of Deliverables and Formative Research developed by the students during the academic semesters covered in this work. Additionally, combined with a severe statistical analysis to determine patterns of usage.

UNSA's Educational Model is based on the professional training of students based on competences and the application of Active Didactic Strategies [3] to achieve the objectives and competences established for training.

The EPIS 2022 achieved official accreditation certification by the Accreditation Board for Engineering and Technology (ABET) [4], which has as one of its objectives that the University demonstrates that its graduates and trainees achieve the desired competencies and student outcomes.

The research produced is applied and descriptive, the methodology used is based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodological model [5].

The main result obtained is that the models used have a high significance according to the t-tests and Snedecor's F-tests to which they are subjected.

It is concluded that, if the same teaching and learning process is applied in the course, the result obtained will be very similar, which allows us to visualise that what is planned and executed is well designed; and therefore the student ends up learning very well.

II. Theoretical Framework

A. Active Didactic Strategies

The EPIS professors plan as one of their objectives to apply a certain Didactic Strategy for the development of the course, taking into account that the students are trained through the competence-based approach being the centre of attention as referred to by [6], by applying an established or agreed active didactic strategy contemplated by [7], [8] without leaving aside what is dealt with by [9] which is improved with technology.

The training of students with Active Didactic Strategies implies autonomous and continuous learning, as concluded by [10], ordering their learning, being a permanent work in the professional training.

In the experience of using the active didactic strategy of Problem-Based Learning (PBL), as a learning methodologies, students work in small groups of three or four members executing the planning, organisation, planning the solution, developing it, validating it and thus satisfying the requirements of the problem or problems of reality.

The application of the aforementioned Active Didactic Strategies encourages students to work in teams, increasing their critical capacity, broadening their knowledge of the problem area, developing the competences of the course and therefore those of their professional training, increasing their abilities as well as their procedural skills, valuing their achievements and thus achieving a certain experience in problem-solving for the good of society.

The adaptation and improvement in the application of the Active Didactic Strategy and its complementary techniques are of vital importance to improve students' performance, and teachers are the ones most called upon to adapt them to the teaching and learning methods for the benefit of the students.

B. Indicator-Based Learning Assessment

The evaluation, reflected in a grade that the student obtains to continue to the next level of their study plan, must be based on data and complementary information, to know the student's knowledge deficiencies, weaknesses, the application of methodologies, methods, techniques, tools, as well as understanding the technologies and their limitations in the application of the probable solutions to the problem being dealt with.

In his proposal [11] he contemplates the Evaluation System based on Performance Indicators for Problem Based Learning, which contemplates in its structure: a) General Competence(s) (of the graduate profile), b) Course Competence(s), c) Student Outcomes, d) Process(es), e) Criterion(s), f) Performance Indicator(s), g) Indicator(s) rating scale that allow a more objective, fair evaluation; and have data that allows its analysis to reach conclusions; to its initial structure was adapted the inclusion of the element ‘Student Outcomes’, which is related to the other elements and application for the rating in the evaluation.

As a result of the application of the same in study [12] from which the set of stored data of the academic semesters involved in the present work is derived.

The Student Outcomes defined for the NE course based on those of the EPIS for the NE course are:

7.2 Acquire new knowledge in learning using appropriate strategies in the context of technological changes.

8.2 Selects appropriate tools, skills, techniques, methods and/or methodologies specific to the discipline with an understanding of its limitations.

Fig. 1 shows some of the performance indicators graded according to a grading scale, from which the mark obtained by the student for each problem proposed for development is calculated.

UNIVERSIDAD NACIONAL DE SAN AGUSTIN DE AREQUIPA EVALUATION OF THE DELIVERABLE REPORT AND FORMATIVE RES
FACULTY OF PRODUCTION AND SERVICE ENGINEERING
PROFESSIONAL SCHOOL OF SYSTEMS ENGINEERING Couse : Business Electronics
Semestre : 2021 B (VIII)

			Competences	r	r	r	r	r	r	r	r	p	p	p	p	
			Student Outcomes	7.2	7.2	7.2	7.2	7.2	7.2	7.2	7.2	8.2	8.2	8.2	8.2	
			Criteria	Planni	Organ	Problem		Framework		Generation of possible solutions (A					Prototype	
			Indicators	1	1	1	2	3	15	1	7	2	5	2	1	
				Objetivo General	Organización del equipo	Descripción del contexto	Identificar el problema	Conceptos nuevos	Antecedentes Investigativos	Enunciado	Selección de la mejor alternativa	metodologías relacionadas al	herramientas relacionadas al	Exposición de los resultados del	Requerimientos de TI	Note = (DBnn *DD$1 1)/33 RE.7.2
			Grading Scale	4	4	4	4	4	4	4	4	4	4	4	4	50
				=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	20
Teo	Laborato		Minimum amount (=good)(sc 4-5=Mbu)					8	8	3		2	3			
Grou	Grou	Sub	Surname and first name													
A	A	1	Estudiante 1	2	3	2	3	3	4	2	3	4	3	3	4	18
A	A	1	Estudiante 2	2	3	2	2	1	0	0	0	4	3	3	0	11
A	A	1	Estudiante 3	2	3	2	3	3	4	2	3	4	3	3	4	18

Fig. 1. Grading tool for the deliverable report.

Every evaluation must generate complementary information which must be organised, consolidated, analysed, interpreted, communicated to whoever it may concern, generate knowledge and thus be useful for planning, taking actions and decisions. Thus, it is important to analyse the academic performance of students in order to take timely action individually, as a team and globally for all participants.

C. Data Mining

For study [13] ‘Data Mining is the process of extracting useful information from an accumulation of data, often from a data warehouse or from the collection of linked data sets’ and also for study [14] ‘is the process of discovering patterns and other valuable information in large data sets’.

In the DM process [15] refers to the set of methods, techniques and technologies to explore large databases, automatically, in order to find patterns, trends that help to interpret the behaviour of the data in a given context. Also in research [13] notes that it uses statistical analysis, machine learning algorithms in the search in large databases, analysing the data from different angles and thus identifying trends, patterns and unknown associations.

The application of data mining occurs in different contexts of organisations, as for example in study [16], the Data Mining Model for Predicting Customer Purchase Behavior in e-Commerce Context is proposed; it is also in the context of education that DM is increasingly taking place to extract information from data and be used to plan projects, improve processes in teaching and learning.

Given the growth of the application of DM in the educational context, it has given rise to Educational Data Mining (EDM) treated as such in study [17] by contemplating that the usual approaches and assessments are not suitable to discover the revealing information of the student's scores. Therefore, EDM deals with cases of student data research used to investigate associations not preliminarily detected in a student database. Also in study [18] state that EDM focuses on developing methods for discovering that Learning Management System (LMS) data is used to understand and make improvements in virtual learning environments, as well as building analytical models to discover patterns and trends. Many works deal with student performance as in studies [19], [20], [21], [22], [23], [24], and also in study [25] where the detection of factors associated with performance from educational assessment databases is discussed; as well as somehow dealing with KPI indicators to map strategies and performance indicators applying predictive modelling as applied by study [29].

What is normally followed in a DM process is the application of the CRISP-DM methodology in its six phases or stages that ensures that the expected results are obtained.

D. Data Mining Techniques and Tools

Data mining techniques are being increasingly adopted in studies related to data processing in organisations, which use them for data processing, analysis, generation of information for decision making [14].

Data mining tools contain powerful statistical, mathematical and analytical capabilities to examine huge data sets to identify trends, associations, patterns, correlations, intelligence and strategic organisational information to support planning and decision making [13].

In study [19], they have used DM techniques to generate predictive models of academic performance: decision trees and multivariate regression.

On the other hand, in study [17] they use the statistical package for social sciences (SPSS V26) for both descriptive and inferential statistics, using ANOVA as a method of statistical analysis; [21] uses IBM SPSS MODELER 18.1 through classification techniques.

Also, in [20] they use the best known, WEKA classifiers; so also in study [22] additionally use it with clustering; and also [25] use it to develop statistical analysis of massive information from evaluation databases using association, clustering, classification.

In study [23] they use classification with machine learning algorithms; also in study [26] they used Machine Learning techniques.

Also, in research [24], the need to consider qualitative and quantitative elements to predict and evaluate the academic performance of students is highlighted. Through machine learning, educational data is refined, discovering valuable patterns and simplifying complexities through feature selection methods using clustering, classification and regression.

III. Related Work

The field of action and application of DM is very broad in many contexts of activity in organisations; thus, [27] in the field of IT project management, when defining a new project, it is necessary to choose the development team considering characteristics, roles, results of previous projects, experience of the developers, suitability and affinity. As well as in study [16] in the e-commerce environment to predict customer buying behaviour.

In the context of education, the study [19] performs an analysis of the academic performance of students entering a degree programme, relating performance to socio-economic and academic characteristics stored in a database; as well as in study [17] they analyse the performance of students in TOEFL reading, listening and writing scores; study [20] investigates whether there are patterns in the available data that can be useful for predicting performance based on personal and pre-college characteristics, in research [21] as the aim is to automatically classify students based on their academic performance and to identify profiles and trends as well as student attrition; the study[22] uses classification techniques to study and analyse student performance, as well as to help to advise students; the study [25] develops statistical analyses of massive information from student assessment databases, also presenting to the scientific community statistical procedures and techniques that can be valuable and replicable in other educational and/or social spaces; also in study [23] proposes a model for predicting student performance, allowing accurate identification of final grades and providing information to guide educational interventions. Also in study [26], the analysis of students' academic performance is carried out using the Power BI tool, evaluated different machine learning techniques and found Random Forest to be the most efficient algorithm, with the highest accuracy of academic performance. In study [24] their systems collect data from various sources: exams, virtual courses, enrolments, e-learning platforms; the analysis of this data leads to the application of machine learning techniques to predict and evaluate student performance.

IV. Methodology

The methodology used in this work is based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodological model. For study [13] there are as many approaches to DM as there are data mining workers. The approach defined depends on the type of questions asked, as well as the content and organisation of the database.

In this regard [5] summarises the phases or stages of CRISP-DM implementation developed from the knowledge discovery processes widely used in organisations, responding directly to user requirements; namely: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Implementation. Fig. 2 shows the six phases of CRISP-DM.

Fig. 2. CRISP-DM model. source (Taylor, 2017).

In many organisations and in research work carried out, the application of CRISP-DM implementation phases is generated according to the problem to be addressed; thus, the study [13] proposes the phases of: Understanding the problem, or at least the research area, Data collection, Data preparation and understanding, User training; also in study [28] in his guide proposes the phases of: : Understanding the business, Understanding the data, Data preparation, Modelling, Evaluation, Virtual, which is closer to the essence of CRISP-DM.

Already in the application of CRISP-DM we have that [20], develops the work with the phases of: Business Understanding, Data Understanding, Data Preprocessing, Modelling, working then the results; in study [18], [21], they use the phases of: Business Understanding, Data Understanding, Data Preparation, Data Preparation, Modelling, working then also the results; in study [26], they deal with the phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Visualisation.

The data mining project uses the tools of: IBM SPSS Modeler V.21 which supports decision tree models, neural networks and regression models; to do statistical analysis of the data, determine relationships between Student Outcomes and their assessments. It is used to perform data capture and analysis, create tables and graphs with complex data; known for its ability to handle large volumes of data as well as perform text analysis among other formats.

IBM SPSS Modeler software offers advanced statistics in addition to many basic statistical functions, including cross-tabulation, frequencies, dual variable statistics such as t-tests and ANOVA, correlation, linear and non-linear models.

RAPIDMINER is a free software tool used for data mining tasks in both research and organisations, supporting single and multidimensional predictive models. It works from nestable operators to compose the models to work with; to do data analysis, determine patterns.

It allows to accelerate the creation, delivery and maintenance of predictive analytics; dealing with large volumes of data from multiple different sources, managed by more business-oriented than technical profiles.

Rapidminer software by the use of its workflow system decreases the use of code for data modelling, speeding up the analysis.

The data mining task is to analyse the university performance of students from their assessments.

V. Development of the Proposal

A. Research Context

The experience is developed in the subject of Electronic Business (NE) that correspond to the V semester respectively of the EPIS Curriculum.

From the academic year 2020 to the year 2023, the tool of qualification of the Deliverable Report and Formative Research Report is applied, the result of which by Feedback Report is given to the students so that in work team sessions they analyse, reach conclusions on the qualification of each Performance Indicator, propose improvements and apply them in the development and solution of the following problems.

In the development of the course it is contemplated and ensured that there is alignment between the Student Outcomes, the contents, the evaluation method, the use of the grading tool that encompasses what is proposed by study [11].

B. Development of the Model

The data mining work is based on the CRISP-DM research approach as it is a non-proprietary, application-neutral standard for data mining projects and widely used by users.

Phase 1: Understanding the business

The objectives and requirements of the business, which in this case is the education sector, are understood.

The objective of the research is to analyse the results of the student evaluation by applying different data mining techniques.

Among the main requirements are:

- Determine the relationship between Student Outcomes and their Evaluation Process.
- To analyse the results of the application of different data mining algorithms.
- Determine trends or patterns.
- Show the lowest performing performance indicators in order to plan and take future improvement actions.

Phase 2: Understanding the data

It comprises the tasks of initial data collection, establishing its main characteristics such as: structure, quality, identifying likely subsets of the data of interest.

The initial data collection is given by the recording of the ratings made in the respective assessment of each problem addressed through the tool used.

The description of the data is given by the composition of the 1681 records of the grades recorded in the tool in Excel, the data are grouped into two categories:

- Course General Data with the Final Results of the evaluation of each student: Faculty, Professional School, course, credits, type, year, semester, theory group, laboratory group, laboratory subgroup, analysis code, surname and first name, gender; marks for: final, exam and continuous averages, exams 1 2 3, continuous 1 2 3, student results 7.2 and 8.2; dropout.
- Data of the Results of the Grading of the Deliverable Reports made by the students: Problem, problem name, problem score, student performance indicators score, problem assessment criteria scores, student outcomes scores 7.2 and 8.2 of the problem.

The exploration of data by understanding its structure and characteristics.

In the Final Assessment Result 5 categories are considered to appreciate the behaviour of the assessment; Fig. 3 shows the behaviour of the Final Assessment Results variable of the students in the course, appreciating that category 4 with a rating of Very Good with a mark of 14 to 16 prevails, and category 5 with a rating of Outstanding with a mark of 17 to 20 has a significant presence not so in the year 2021.

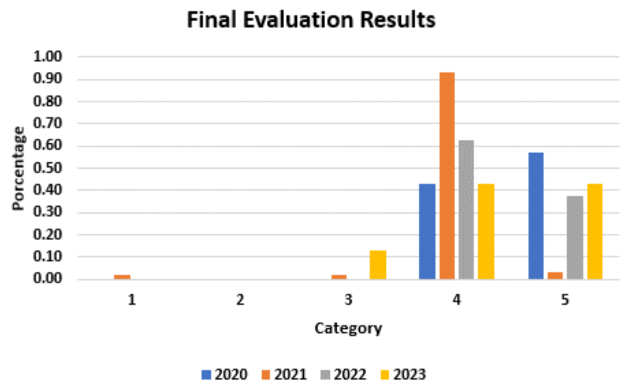

Fig. 3. Final evaluation results.

Fig. 4 shows the behaviour of the variable: Gender of the students, showing that the presence of men prevails, which is derived from the application and entry to study in the study programme.

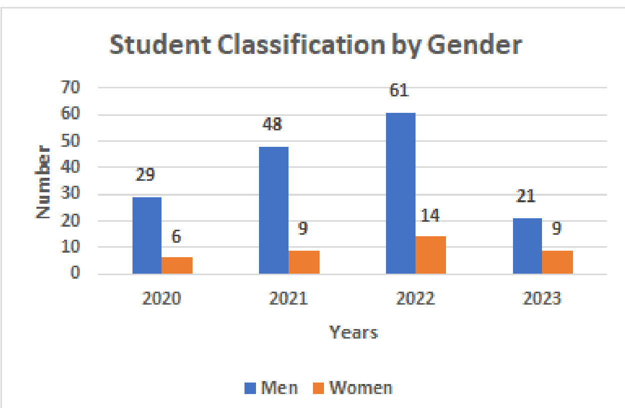

Fig. 4. Classification of students according to gender.

Table I shows the composition of the Theory Groups, Laboratory Groups and Subgroups.

TABLE I. CONFORMATION OF GROUPS AND SUBGROUPS

Year	Groups Theory	Groups Laboratory	Subgroups Laboratory
2020	A	A	5
	B	B	5
2021	A	A	4
	B	B	6
		C	5
2022	A	A	5
	B	B	5
		C	5
		D	5
2023	A	A	4
	B	B	4

Table II shows the problems dealt with in the development of the course by applying Problem Based Learning.

TABLE II. PROBLEMS

Coding	Description
1	Virtual Stores
2	Customer Relations Management - CRM
3	Supply Chain Management - SCM
4	e-Marketplace
5	e-Learning
6	e Employee
7	e-Government
8	m-Business

The data quality check checked the consistency of the data values.

Phase 3: Data preparation

This phase involves the activities to extract the student data from the assessment record and build the final dataset into a dataset described with 112 data for each instance or occurrence that serves for modelling.

The Selection of the data covers the student assessments for the years 2020 to 2023 which are 1681 records of the grades recorded in the grading tool.

Data Cleaning deals with the inclusion of missing data, such as gender; or derived from existing data such as dropout/dropout; and data correction such as replacing blank value with zero in performance indicators.

The Construction and Integration of the data is given by the elaboration of the file containing the two categories: General Course Data with the Final Results of the evaluation of each student and the Data of the Grading Results of the Deliverable Reports made by the students, and thus obtaining the final dataset.

Data Formatting converts the data types for the application of the specific DM technology. The 66 performance indicators are nominal variables with four or five different values.

Phase 4: Modelling

In this phase, the modelling techniques that best enable the analysis of the data of the data mining project are selected. The data are entered into the data mining software, the runs are performed and the results are studied.

In the modelling, the following options have been used:

- Linear Regression, supported by correlation coefficients, Student's t-test and Snedecor's F-test.

The Student's t-test is used when the population does not follow a normal distribution. In this case we do not know the behaviour of the population data so we understand that there is no normal distribution because this t-test is robust to deviations from the normality of the population.

The F-test is used to assess the overall significance of a regression model. This test gives strength to the analysis of variance of linear regression.

According to the systematic review of the research works discussed or reviewed, we did not find a general model that deals with student performance indicators for each of the criteria proposed in an evaluation based on the definition of the rubric that allows us to evaluate the level of performance or performance of a task, which in this case are the proposed problems; Our proposal differs because we reach the level of qualification in the evaluation of the performance indicators defined for each criterion; and we conceptualise a scheme based on periods, groups and subgroups, problems, managing to reach estimated models; therefore, it is determined to use the splines for the determination of patterns.

Phase 5: Evaluation

The models created are measured against the defined objectives; according to the result feedback is given and the resulting modifications are part of the knowledge discovery process.

Phase 6: Implementation and Visualisation

In this phase, the generated models are used by other users to produce business intelligence. Based on the results obtained, there are tables and graphs that allow visualisation, the generation of reports and improvements in the procedures of the processes.

VI. Results

The main objective of the study is firstly to perform the statistical analysis of data to determine the relationship of Student Outcomes and their Assessment Process; secondly to perform the data mining process to determine the existence or not of patterns, correlations or trends in the assessment data.

The results of the statistical analysis of the SPPS Modeler application are shown below.

There is a high relationship between the predictor variables and the R72 model, in the same way, there is similarity with the R82 model, which implies that in the R82 model there is a greater effort on the part of the students to learn. Highly significant results by analysis of variance (F-test). These results are reflected in Table III and Table IV.

On the other hand, the constants of the variables of the estimated models show a highly significant Student's t-test, supporting the result of the correlation coefficient. These results can be seen in Table V.

TABLE III. Results of the Correlation Coefficient for the R72 y R82 Models

Model	R	R^2	R^2 corrected	Predictor variables
R72	0,963	0,928	0,926	(Constant), CO3, EX3, EX1, EX2, CO1, CO2
R82	0,944	0,891	0,887	(Constant), CO3, EX3, EX1, EX2, CO1, CO2

TABLE IV. Analysis of Variance

Model	Sum of squares	gf	Root mean square	F	Significance
R72	547,947	6	91,324	409,448	,000
R82	1002,675	6	167,112	258,653	,000

TABLE V. Student t Test

Predictor variables	Dependent variable			
	Model R72		*Model R82*	
	t	*Significance*	*t*	*Significance*
Constant	3,127	,002	-1,990	,048
EX1	13,743	,000	-3,108	,002
EX2	11,510	,000	1,829	,069
EX3	11,527	,000	2,398	,017
CO1	10,451	,000	10,830	,000
CO2	3,724	,000	13,905	,000
CO3	11,895	,000	5,619	,000

From the application of the Rapidminer tool we have the results of applying splines for the determination of patterns.

Fig. 5 shows that in the RE82 model, students are assertive in the handling of techniques, methods and tools, but they have difficulties in the RE72 model in coping with learning. It can be seen that between grades 13 and 17 students make the greatest effort in learning and applying information technologies; it can also be seen that from grade 18 onwards they culminate with their learning.

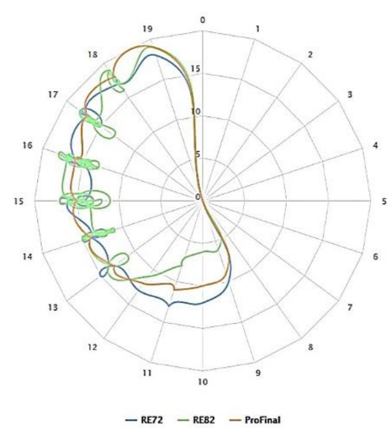

Fig. 5. R72_R82_PFINAL.

Fig. 6 shows the Student Outcome model 7.2 - RE72 which deals with the average exam marks and average continuum marks, indicating that students between marks 13 and 15 have greater practical skills and greater difficulty in theoretical learning; from the upper limit mentioned above, both theoretical and practical aspects are combined to reach the desired level of learning.

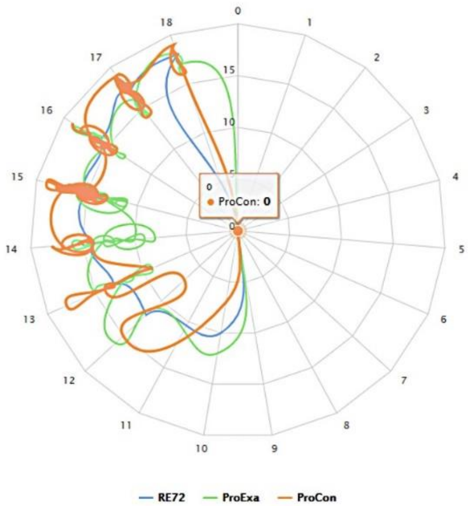

Fig. 6. R72_PEXA-PCON.

On the other hand, Fig. 7 shows the Student Outcome model 8.2 - RE82 which deals with the average marks of exams and the average of continuations, indicating that the student understands the practice better through the use of techniques, methods and information technology tools that work in parallel with the proposed RE82 model; and the theory is adapted according to the practice until mark 18, at which point they are combined with the model, reaching the desired level of learning.

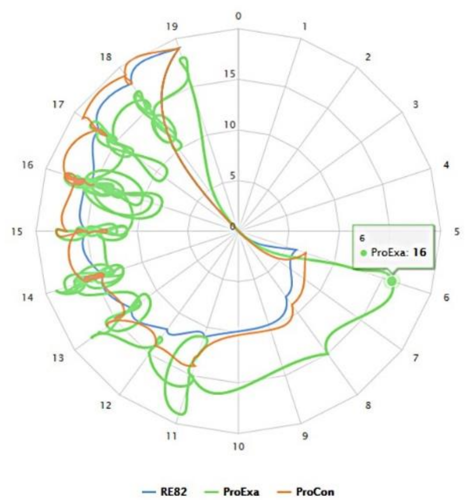

Fig. 7. R82_PEXA-PCON.

From the result of the treatment of the qualification of the performance indicators of the criteria contemplated for the evaluation of the problems proposed and developed by the students, those performance indicators (66 in total) are shown that in the qualification scales have the scale value of Absent (>= to 15%) and/or Regular (>=10%) that allows the teachers to pay attention to reflect weaknesses to improve, and to complement the analysis of the results. and based on the conclusions to raise the improvement actions to raise the level of performance of the students. These actions, when dealt with in the first class, session of the following semester, also encourage students to take into account the fulfilment of the tasks to be completed in all the indicators and thus have a high level of learning.

Table VI shows the performance indicators involved.

TABLE VI. PERFORMANCE INDICATORS INVOLVED

Indicators Involved				
Indicator		***Percentages Scale***		
Cod	***Name***	***Absent***	***Regular***	***Total***
12	Organisation of information	6.70	11.30	18.00
13	New concepts	4.50	17.60	22.10
18	Search for models related to the problem	5.80	13.50	19.30
20	Search for methods related to the problem	7.60	14.90	22.50
21	Search for techniques related to the problem	7.40	13.00	20.40
22	Search for tools related to the problem	7.30	18.30	25.60
23	Search for the necessary skills	8.60	13.60	22.20
25	Research background	11.70	15.10	26.80
26	Configuration of tools (SW)	16.60	9.40	26.00
27	Installation of tools (SW)	17.30	6.20	23.50
28	Sharing information worked on	16.00	12.60	28.60
30	Drawing up the comparative table of methodologies	15.10	3.30	18.40
31	Drawing up the comparative table of methods	16.90	3.20	20.10
32	Drawing up the comparative table of techniques	16.80	3.50	20.30
34	Drawing up the table of necessary skills	16.30	7.90	24.20
35	Propose and support the IT collected.	16.40	7.70	24.10
43	Others.	30.40	4.90	35.30
44	Selection of the best alternative	18.70	7.00	25.70
50	IT requirements	15.10	13.20	28.30
51	Configuration	15.10	8.10	23.20
52	Installation	17.10	6.40	23.50
53	Prints Screen of results	12.10	11.50	23.60
56	Discussion or comments	30.60	3.50	34.10
58	Writing consistent with objectives	11.50	18.00	29.50
62	Other	20.60	14.80	35.40

It is also emphasised that many efforts are being made to increase the learning levels of Peruvian students, such as the effort being made by UNSA to apply Active Didactic Strategies in the teaching-learning process in order to increase student learning levels.

VII. DISCUSSION

There are several works in Education such as that of [27] in the area of Data Mining; in the selection of computer project teams; as well as in the area of those related to students' academic performance from those of [19], [20], [25], [23], [26],

[24], or to extract academic behavioural profiles such as that of [21] or also that of [22] to help advise students and predict their academic performance.

Having that the prediction of academic performance shows challenges due to the varied factors that influence it as discussed by [24] by referring to [H.A.A. Hamza, P. Kommers, 2018, and M.M.A. Tair, A.M. El-Halees, 2012], having that through prediction by machine learning is instrumental in improving education in various ways, by allowing early identification of deficiencies, academic difficulties, enabling timely interventions and personalized learning plans.

In view of the above and nowadays the availability of information technologies with sufficient capacity to process large and varied types of data has allowed Data Mining techniques to evolve and allow obtaining, processing and detecting information from large amounts of data tools Nghe et al., 2007, cited in [25].

This study uses IBM SPSS Modeler v. 21 software for statistical analysis of the data to determine relationships and Rapidminer software to determine the presence of patterns, showing the potential of techniques and algorithms that respond to the stated objective by applying simple data analysis tools.

Continuous monitoring promotes quality assurance, accountability and a competitive advantage for institutions. Overall, it empowers educators to provide targeted support to plan and make improvement decisions, leading to better student outcomes and a more responsive education system as manifested in [24].

A variety of techniques are also used that based on different algorithms allow the processing of data from a given storage source and the evaluation of the results provided to make appropriate decisions.

VIII. Conclusion

The following conclusions are reached:

The R72 model considers the Student's t-test which is highly significant, which means that the prognosis that can be obtained, the results are totally valid; that is to say that if the same teaching-learning process is applied the result is going to be very similar. Because there is a high relationship between the RE72 (Student Outcome 7.2) and its evaluation process, where the Snedecor's F-test is highly significant. That is, in the teaching and learning process the students have finished learning. Similarly, the R82 model shows similarity with respect to the RE72 model.

According to the result of the analysis students have some difficulty in creating results (in the Deliverable Report) of creating the product. There is a problem in the starting point of the course; and they start to understand in Co1 (Continuous Assessment 1) that deals with the development of the first two (2) problems; which is not enough to pass the EX1 (Exam 1); and from CO2 (Continuous Assessment 2) they correct themselves noticing the improvement in the learning and they manage to finish the learning.

The students in the RE82 model show assertiveness in the handling of techniques, methods and tools; however, in the RE72 model they present certain difficulties in dealing with theoretical learning and at the end of the period they complete their learning.

The Student Outcome model 7.2 - RE72 PEXA-PCON, indicates that the student, in certain sections, presents greater skills for practice and in other sections greater difficulty for theoretical learning; however, from the upper limit, both the theoretical and practical aspects are combined, reaching the desired level of learning.

The Student Outcome Model 8.2 - RE82 PEXA-PCON, indicates that the student understands the practice better through the use of techniques, methods and information technology tools that are performed in parallel with the proposed RE82 model in many sections; and the theory is adapted according to the practice until the upper section, where they are combined with the model, reaching the desired level of learning.

It is possible to determine the highest and lowest performance indicators to which attention should be paid by complementing the factors that influence the lower performance of students and thus plan future improvement actions that allow better performance of students and their level of learning to be at a high level. This is an advantage over other forms of assessment, as it leads to the rating of each Performance Indicator involved in each Criterion, which in its analysis and evaluation makes it possible to determine the status level according to its rating scale and to make decisions to improve student performance.

It allowed us to see the additional particular consideration that can be given to a system by considering, for example, the Hierarchy of levels that the evaluation system contemplates from Faculty, Professional School, Course, Academic Year and Semester, Theory Group, Laboratory Group and Subgroup, Problem.

It allowed to visualise the future applicability of the analysis that can be carried out by the use of other Active Didactic Strategies that are used in other courses in the teaching-learning process.

It is recognised that the scope of future application can be extended to other courses of the EPIS Syllabus, to other Professional Schools of the Faculty, as well as Professional Schools of other Faculties and thus be able to obtain results at Faculty Level, University of the application of the Evaluation by Performance Indicators determined for each of the Criteria contemplated, which differs as a working method with respect to the treatment of students‘ performance dealt with by other works, for the development of students’ Competences when applying Active Didactic Strategies in the teaching-learning process.

Acknowledgment

Thanks to the Universidad Nacional de San Agustín de Arequipa for the support it provides for the development and implementation of proposals that benefit the continuous improvement of student performance through the application of Active Teaching Strategies, with beneficial results for students as well as the teaching-learning process.

References

[1] Escuela Profesional de Ingeniería de Sistemas. http://www.episunsa.edu.pe.

[2] Universidad Nacional de San Agustín de Arequipa. http://www.unsa.edu.pe.

[3] C Ma. Cristina Sánchez Martínez, M. Aguilar Venegas, J.L. Martínez Durán and J. L. Sánchez Ríos, Estrategias didácticas en entornos de aprendizaje enriquecidos con tecnología (antes del Covid-19), UNIVERSIDAD AUTÓNOMA METROPOLITANA-XOCHIMILCO, México, 2020, pp. 11-15.

[4] ABET. Why ABET Accreditation Matters. https://amspub.abet.org/aps/category-search?countries=PE.

[5] Ultimo acceso julio 2023.

[6] Connolly, Thomas; Begg, Carolyn, Sistemas de Bases de Datos, cuarta edición, Pearson Educación S.A., España, 2005, págs.1120-1122.

[7] B. Restrepo Gómez, Aprendizaje basado en problemas (ABP): una innovación didáctica para la enseñanza universitaria, Educación y Educadores, vol. 8, 2005, pp. 9-19, Universidad de La Sabana, Cundinamarca, Colombia.

[8] Subdirección de Currículum y Evaluación, Dirección de Desarrollo Académico, Vicerrectoría Académica de Pregrado, Universidad Tecnológica de Chile INACAP. (2017). "Manual de Estrategias Didácticas: Orientaciones para su selección", Santiago, Chile: Ediciones INACAP.

[9] R. Rodriguez, "Compendio de estrategias bajo el enfoque por competencias", Instituto Tecnológico de Sonora, México, 2007. http://www.itesca.edu.mx/documentos/desarrollo_academico/compendio_de_estrategias_didacticas.pdf.

[10] Ma. C. Sanchez, M. Aguilar, J.L. Martinez and J.L. Sanchez, "Estrategias didácticas en entornos de aprendizaje enriquecidos con tecnología (antes del Covid 19)", Universidad Autónoma Metropolitana, No. 146 series académicos, Mexico, 2020.

[11] A. León, E. Risco del Valle, and C. Alarcón, "Estrategias de Aprendizaje en educación superior en un modelo curricular por competencias", Revista de la Educación Superior, Vol. XLIII (4); No. 172, pp. 123-144, octubre-diciembre del 2014. ISSN: 0185-2760.

[12] C. Baluarte-Araya, " Proposal of an Assessment System based on Indicators to Problem Based Learning – IEEE Conference Publication, Published in: 2020 39th International Conference of the Chilean Computer Science Society (SCCC), 16-20 Nov. 2020, Coquimbo, Chile. DOI: 10.1109/SCCC51225.2020.9281203.

[13] César Baluarte-Araya, Ernesto Suarez-Lopez and Oscar Ramirez-Valdez, "Problem based Learning: An Experience of Evaluation based on Indicators, Case of Electronic Business in Professional Career of Systems Engineering" International Journal of Advanced Computer Science and Applications (IJACSA), 12(9), 2021 http://dx.doi.org/10.14569/IJACSA.2021.0120966.

[14] SAP, ¿Qué es la minería de datos?, 2024 https://www.sap.com/latinamerica/products/technology-platform/hana/what-is-data-mining.html IBM, ¿Qué es la minería de datos?, 2024. https://www.ibm.com/es-es/topics/data-mining.

[15] R. Herrera, Myriam; Ruiz, Susana; Romagnano, María; Ganga, Leonel; Lund, María Inés y Torres, Estela. "Aplicando métodos y técnicas de la ciencia de los datos a datos universitarios". XXI Workshop de Investigadores en Ciencias de la Computación, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan, Argentina, Abril 2019.

[16] Orieb Abu Alghanam, Sumaya N. Al-Khatib and Mohammad O. Hiari, "Data Mining Model for Predicting Customer Purchase Behavior in E-Commerce Context" International Journal of Advanced Computer Science and Applications(IJACSA), 13(2), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130249.

[17] Khaled M. Hassan, Mohammed Helmy Khafagy and Mostafa Thabet, "Mining Educational Data to Analyze the Student's Performance in TOEFL iBT Reading, Listening and Writing Scores" International Journal of Advanced Computer Science and Applications(IJACSA), 13(7), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130741.

[18] Johan Calderon-Valenzuela, Keisi Payihuanca-Mamani and Norka Bedregal-Alpaca, "Educational Data Mining to Identify the Patterns of Use made by the University Professors of the Moodle Platform" International Journal of Advanced Computer Science and Applications(IJACSA), 13(1), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130140.

[19] R. Alcover, J. Benlloch, P. Blesa, M. A. Calduch, M. Celma, C. Ferri, L. Iniesta, J. Más, M. J. Ramírez-Quintana, A. Robles, J. M. Valiente, M. J. Vicent and L. R. Zúnica, Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos, XIII Jornadas de Enseñanza Universitaria de la Informática, España.

[20] D. Kabakchieva, Predicting Student Performance by Using Data Mining Methods for Classification, CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume 13, No 1, 2013, Sofia, Bulgaria. DOI: 10.2478/cait-2013-0006.

[21] Norka Bedregal-Alpaca, Danitza Aruquipa-Velazco, and Víctor Cornejo-Aparicio Tecnicas de Data Mining para extraer perfiles comportamiento Académico y predecir la deserción universitaria, Revista Ibérica de Sistemas e Tecnologias de Informação RISTI N.º E27, 03/2020, Pages: 592–604.

[22] Hosam Alhakami, Tahani Alsubait and Abdullah Aljarallah, "Data Mining for Student Advising" International Journal of Advanced Computer Science and Applications(IJACSA), 11(3), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110367.

[23] Abdellatif HARIF and Moulay Abdellah KASSIMI, "Predictive Modeling of Student Performance Using RFECV-RF for Feature Selection and Machine Learning Techniques" International Journal of Advanced Computer Science and Applications(IJACSA), 15(7), 2024. http://dx.doi.org/10.14569/IJACSA.2024.0150723.

[24] Xi LU, "Modern Education: Advanced Prediction Techniques for Student Achievement Data" International Journal of Advanced Computer Science and Applications(IJACSA), 15(1), 2024. http://dx.doi.org/10.14569/IJACSA.2024.01501126.

[25] Martínez-Abad, Fernando; Hernández-Ramos, Juan Pablo, Técnicas de minería de datos con software libre para la detección de factores asociados al rendimiento, REXE. Revista de Estudios y Experiencias en Educación, vol. 2, núm. Esp.2, 2018, Universidad Católica de la Santísima Concepción, Chile, Disponible en: https://www.redalyc.org/articulo.oa?id=243156768012, DOI: https://doi.org/10.21703/rexe.Especial3201812514512.

[26] Salto- Mero, J. & Cruz-Felipe, M., (2023). Análisis del Rendimiento Académico de Estudiantes de las Carreras Economía y Turismo con Power BI en los Periodos (2021). 593 Digital Publisher CEIT, 9(1), 762-772, https://doi.org/10.33386/593dp.2024.1.2162.

[27] I. Wilford R., Minería de datos: herramienta de apoyo en la selección de equipos de proyectos informáticos, Industrial/Vol. XXVII/No. 2-3/2006, Cuba, 2006.

[28] IBM, Guia de CRISP-DM de IBM SPSS Modeler, 2021. https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf.

[29] Ashraf Abdelhadi, Suhaila Zainudin and Nor Samsiah Sani, "A Regression Model to Predict Key Performance Indicators in Higher Education Enrollments" International Journal of Advanced Computer Science and Applications(IJACSA), 13(1), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130156.