

How Do Students Behave When Using A Tutoring System? Employing Data Mining to Identify Behavioral Patterns Associated to The Learning of Mathematics

<https://doi.org/10.3991/ijet.v15i22.17075>

Roberto Angel Melendez-Armenta
Universidad Veracruzana, Veracruz, México
Tecnológico Nacional de México/Instituto Tecnológico Superior de Misantla, Misantla, Veracruz, México

N. Sofia Huerta-Pacheco
Universidad Nacional Autónoma de México, Ciudad de México, México

Luis Alberto Morales-Rosales
Conacyt-Universidad Michoacana de San Nicolás de Hidalgo, Morelia, México

Genaro Rebolledo-Mendez (✉)
Transferencia de Tecnología, Monterrey, Nuevo León, México
g.rebolledo@tec.mx

Abstract—The inclusion of technology in the academic processes has led to constant innovation and investment of resources to offer a first-level educational service, with international standards, methodologies, study plans, and last generation laboratories. This paper focuses on how teachers can make use of an educational technology tool that will allow them to identify patterns that are associated with learning based on human-computer interaction. We present evidence of learning outcomes based on the detection of behaviors in Intelligent Tutoring Systems. These patterns pave the way for the automatic identification of patterns in association to students' learning while using an educational technological tool. The results suggest a model for student's behavior identification when interacting with the technological tool "Scooter". This model identifies students with prospective better learning outcomes as well as students with difficulties to solve math problems. Work for the future will analyze data that comes from different settings apart from solving exercises in Scooter to prove the hypothesis that there are patterns of behavior associated to learning outcomes in different problem-solving situations presented by educational technology.

Keywords—Student-computer interaction, educational data mining, pattern recognition, educational technology

1 Introduction

Technological innovations have modified the traditional learning approaches, revitalizing the educational system. In traditional education, the detailed observations of how students learn require invasive methods. These methods are commonly applied by external evaluators who give value judgments on the work done by teachers and students, causing them to not develop adequately in the classrooms. Moreover, there is great reluctance from teachers to use novel technological developments in the classroom. This is due to the fact that teachers think that, at some future point, technology could replace them [1]. In conventional classrooms, the teachers monitor the students' learning processes analyzing their performance through observation; however, the use of technology has made it easier to acquire the knowledge in relation to the students' behavior based on the information that is generated in the human-computer interaction.

Education technology is the incorporation of ICT in education to support the teaching and learning processes in different educational contexts, formal and informal. Besides, it seeks to improve those processes by means of reaching the objectives looking into effectiveness and meaning or learning. Systems based on artificial intelligence for education have been developed with the purpose of improving the learning process and in benefit of students [2]. Artificial Intelligence in Education (AIED) is a research area aiming at creating educational technology-based resources. It investigates learning wherever it occurs, in traditional classrooms or in workplaces, with the purpose of supporting education. It is an interdisciplinary field working alongside the learning sciences (education, psychology, neuroscience, linguistics, sociology, and anthropology) to promote the development of learning settings [3]. One of the most important educational technologies in this field are Intelligent Tutoring Systems (ITS).

There is some research focused on applying educational technology as a support to the learning processes inside and outside the classrooms. In the ITS context some researchers [4] have determined the usefulness of students' response time in the prediction of performance. Others [5] have studied the link that exists between the objectives pursued and the carelessness in the consecutive trials (problem solving) inside ASSISTments. Moreover, [6] look for patterns to predict student performance in university based on their previous personal and educational characteristics; [7] allowed observing the prediction of student performance when exploring multiple relations between the students' tasks by means of multi-relational matrix factorization methods. In this research a complexity in the identification of patterns in data that comes from education settings has been identified, for which it is necessary to design models, tasks, methods, and algorithms that allow exploring them [8]; other research like that from [9-11] look into academic performance, students' behavior, learning according to gender, obtaining profiles and typologies in the use of ITS individually, besides in some research they have focused in students' collaborative work during human-computer interaction [12-14].

This paper focuses on how teachers can make use of an educational technological tool that will allow them to identify patterns that are associated with learning based on

the human-computer interaction, it is oriented to teaching math to a group of students in southeast Mexico. The objective of this article is to provide a methodological proposal for the detection of behaviors in Intelligent Tutoring Systems, which allows identifying patterns associated to students' learning while using an educational technological tool. For that purpose, the research is designed to answer the following questions.

1. Research question 1: Which are the steps teachers should follow to identify learning behaviors in the student-computer interaction?
2. Research question 2: What kind of patterns can teachers identify when evaluating the students' learning process during the interaction with an educational technological tool?

Therefore, the hypothesis that leads this research is the following: the methodological proposal of processing and analyzing the students' log files while using the computer allows establishing generalizable behaviors focused on their learning.

There is some research that has focused on the use of educational technological tools, for example, [15] present advantages of learning that are statistically significant between the initial and final evaluation of computer thinking abilities in children based on the examination of learning effects with the use of the educational tool called Bee-Bot. In addition, [16] analyze the students' behavior in a multiagent Intelligent Tutoring System that allows reaching their goals, the results also revealed a significant interaction between the objectives achievements and the condition in the achievement results, in a way that students who adopted a dominant performance approach showed greater achievement in the quick response condition; [17] show the results obtained by the teaching assistants when they provide help to children with learning problems and recommend scaffolding as the key theory to carry out their practice. Finally, it is greatly important to highlight that the contribution of this research is obtaining a methodology that uses artificial intelligence by means of data mining algorithms that allow identifying patterns associated to learning, which contribute directly to the feedback and constant personal instruction of teachers towards students.

2 Theoretical Framework

The Educational Systems Based on Computer are developed under the artificial intelligence approaches, used inside classrooms in many countries around the world. An important topic within Artificial Intelligence in Education are the Intelligent Tutoring Systems, whose purpose is making education more accessible, efficient, and be a means that provides useful measures in learning [18]. Among the first Intelligent Tutoring Systems we can find SCHOLAR which was presented by [19] as one of the first evidence. This program carried out instruction dialogues with a student about South American geography. In SCHOLAR, the natural language was used to answer the student's question or to make a question and give feedback regarding the accuracy of the student's answer. [19] emphasized the fundamental differences between

SCHOLAR and other types of computer assisted instructions that were designed in the 70's. Particularly, he discussed how a domain representation can function as a basis to model the student's knowledge.

Another ITS is BIP, which was developed by [20] and programming tasks were assigned to students with the same needs and learning competences. The interest on ITS has increased significantly since 1980, starting with the research carried out by [21]. However, since 1997 evaluation studies that compared the learning results of students who use ITS to those who use other learning instruction methods have been published, for example, the ITS presented by [22-27]. In this sense, [28] present results that determine how students are assigned collaborative activities with the objective of optimizing the learning experience, and they identify the elements that influence in the collaboration and how they relate to cultural issues. This research discovers that an association between collaborative and cultural elements exist, such association allows an improvement in the learning advantages in a student when he/she is using an ITS.

In addition, [29] analyze the cultural aspect and the relation that exists with the affective tendency while using the ITS. This research lays out the hypothesis that the students' cultural dimension can indicate an affective tendency during the use of the ITS. The cultural dimensions that were analyzed were the uncertainty avoidance (UA), the individualism (IND), the masculinity (MAS), the power distance (PD), and the intermediate level in the long-term orientation (LTO), and the results suggest different behavior patterns, which were observed during the use of the Math Intelligent Tutoring System "Scooter" in the middle school level in Mexico. The information about the student's cultural criteria affects the Mexican students' academic environment.

Another aspect in the use of ITS is the research presented by [30], that identifies learning gains in students that work collaboratively when using an Intelligent Tutoring System for math called "Scooter". Having as a result that a student can learn based on his/her social environment, on his/her interaction with learning partners when carrying out a collaborative activity and the use of educational tools that contribute to the teaching-learning process. Moreover, the results show that the students with the same learning style have more learning gains once the most successful pair in a collaborative activity has been found in contrast with those that worked in pairs with different learning styles.

Nowadays, the development of Intelligent Tutoring Systems has been applied in web platforms, [31] present the validation of an ITS in its web version regarding the original desk version by means of a hierarchical analysis process. The results show that both versions offer similarities in their functionality, making it possible to increase the results in the students' individual performance when interacting with an ITS. In conclusion, the use of technology in learning experiences is an area that has had unprecedented growth throughout the years. Each person has his/her own ideas about what students need to reach their goals in terms of education. The Intelligent Tutoring Systems can be a fundamental support, whose use is useful to acquire and increase knowledge, but because of technological development, the use of web platforms has increased making it possible to mark attainable objectives, for example, ITS

“Scooter” in a material resource used by students to learn math and that generates log files, which are analyzed to detect patterns associated to learning.

However, students’ modeling oriented to represent and anticipate performance is one of the approaches of data mining in education. Many indicators of performance are worth being modeled, such as: efficiency, evaluation, learning styles, collaborative activities, lapse of time in the use of the system, help request, and even, cultural differences. In general, data mining contemplates two models: descriptive and predictive. The descriptive models generally apply functions for non-supervised learning to produce patterns that explain or generalize the intrinsic structure, the relations, and the interconnection of data [32]. The predictive models estimate unknown values or dependent variable futures based on the characteristics of related independent variables [33].

The implementation of a type of a model depends on the specific task, for example, association rules, [34], correlational analysis [35] and regression [36]. Moreover, [37] apply learning through reinforcement to improve an ITS design used for teaching basic counselling abilities. In this research, trust and gains in learning were part of the reward function used to train learning through reinforcement, and the results show the success in the improvement of an ITS to teach interpersonal abilities between students.

[38] argue that boys interact differently with ITS than girls. This finding is evident in the results of Bayesian Knowledge Tracing and Learning Curve Analysis models. Additionally, they refute what is essential of considering the gender differences in the way that students learn, the context in reading material that was presented, especially for boys. In more recent years [39] present that the ITS personalizes learning for students with different backgrounds, abilities, behaviors, and knowledge. In this research a method that uses diffuse decision trees is proposed to build a series of diffusion predictive models that combine all the dimensions of the learning styles model of [40]. The results show that diffusion models increased the predictive precision to four dimensions of learning style and made it easier to discover some interesting relations between the behavior variables.

In education terms, the use of data mining for research matters has increased, generating the emergence of a new area named Education Data Mining (EDM). The EDM is defined as a scientific research area centered in the development of methods to make discoveries within the unique types of data that come from educational settings, and the use of those methods to better understand the students and the setting in which they learn [41,42]. Education data mining more often concentrates the predictive models regarding a suppressor variable in the students’ learning when using a Learning Management System (LMS).

[43] identify slow students and they show it by means of a predictive data mining model using algorithms based in classification. Several classification algorithms are applied like the multilayer perceptron, Naive Bayes, SMO, J48 and REPTree to a group of academic data from a student. As a result, statistics based in all the classification algorithms are generated and the comparison of the precision of five classifiers is also carried out, and they find the best performance classification algorithm. Finally, a knowledge flow model is presented among the five classifiers showing the im-

portance of mining data algorithms. Moreover, [44] use different classification methods, such as statistical methods, decision trees, fuzzy logic, and artificial neural networks. In this research experiments were carried out using all the available data, filtered to obtain a greater precision. Finally, some examples of discovered models are shown, and it is also explained that a classifying model appropriate for an educational environment has to be both precise and comprehensible so that teachers and course administrators can use it to make decisions.

3 Methodological Proposal

In Figure 1 one can observe the diagram of the proposed methodology in this research, which allows identifying patterns associated to learning in students who make use of an educational technological tool.

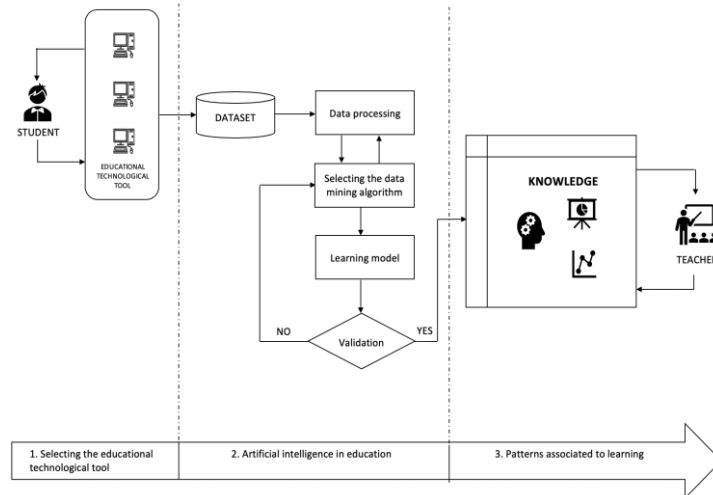


Fig. 1. Methodological proposal for the identification of patterns based on human-computer interaction

3.1 Selecting the educational technological tool

Nowadays, a great variety of educational settings exists (see Table 1) and each of them provides different sources of data that should be processed depending on its nature, the problems, and the specific tasks that will be solved with data mining techniques. Particularly, educational settings store a great quantity of possible data (raw, original or primary) that are often not properly stored allowing to use them to solve a specific problem [45-46].

Depending on the type of educational setting, different kinds of data can be chosen to solve several education problems. All this data can come from different sources, including administrative data, field observations, motivation questionnaires, meas-

urements taken from controlled experiments, and even, final grades. The first educational technological tools allowed the analysis of data that students generate in a virtual learning setting, both individually and collaboratively. The Computer-supported collaborative learning settings (CSCL) are a tool that can be used due to the fact that they generate log files based on the technological tools in which they are implemented. A specific case in the technological tool WorkMates, implemented in a web page, that provides collaborative tools to share files, discussion forums, and most importantly, files that register students' participation.

Table 1. Educational technological tools description

Educational setting	Description
Learning Management System (LMS)	Software that provides functions of administration, documentation, and information on courses (both in traditional classrooms and online). They also offer a great variety of channels and workspaces to facilitate the information exchange and the communication between all the course participants. They register any activity of the student: like reading, writing, taking exams, presenting homework, and commenting on events with classmates.
Intelligent Tutoring System (ITS)	The Intelligent Tutoring Systems provide personalized instruction in a direct way and feedback to students, while modeling their behavior and even changing the way each one interacts, this based on their individual model. They register all the interactions between the student and the tutor.
Adaptive and Intelligent Hypermedia (AIH)	They try to adapt more by building a model of objectives, preferences and each particular student's knowledge, and use this model throughout the interaction with the student to adapt to his or her necessities. The data stored by the AIH are similar to the ITS data.

Nowadays, the great advance in technology developed in the web pages helps to have more access to the e-learning platforms. The e-learning systems developed in a web server are capable of adapting the users' preferences in terms of their learning styles. These systems are able to store the users' behaviors in web log files. For example, Moodle platform provides browsing reports from the students who take the course. These reports include the different variables that are associated to the browsing path of each student in the platform when he/she carries out the corresponding activities. Finally, a specific case is the Intelligent Tutoring System, defined as an educational technological tool which has the capacity of storing log files, in a specific format for each system, which generally include students' behaviors while doing the activities requested by the teachers inside the educational platform. The compilation and integration of this non-processed data are tasks for data mining, making a previous step necessary, which is the data processing.

3.2 Artificial intelligence in education

To correctly apply techniques of artificial intelligence in education, a sequence of steps that allow obtaining better results must be laid out. These steps are listed below:

Data processing: The main purpose of preparing data is to manipulate and transform raw data so that the information content can be exposed and made more accessible [46]. Preparing data is highly important because data without processing can lead

to the extraction of not very useful data and rules as these can contain incomplete data, with noise, and even, inconsistent. Generally preprocessing reduces the initial data to a smaller set.

Selecting the data mining algorithm: Data mining is the automatic extraction of patterns that are found implicit in large collections of data, and according to research carried out by [48-53] the data mining models are divided in:

1. **Predictive:** The term prediction is generally used to develop models that can infer only one aspect of the data (predictor variable) of one combination of characteristics (predictor variables) of the data. Once the models are designed, they will work to predict new results or settings based on new observations.
2. **Finding structures:** The objective is to find structures in the data without a previous idea (initial knowledge) about what should be found. It is attempted to determine what structure naturally emerges from the data.
3. **Relational mining:** The objective is to discover the relations between variables in a set of data which has a great number of variables.
4. **Finding by through models:** The results of a data mining analysis are used inside another data mining analysis. The objective is to obtain a model, generally through prediction; and the predictions of this model are used as an entry for another model.

Learning model: Once data has been collected, cleaned, transformed, and integrated in a file with a specific format; in this phase, the patterns of interest to the research are extracted, based on the implementation of mining data algorithms. There are a number of functionalities inside data mining: classification and regression, associations and correlations, extraction of frequent patterns, grouping and atypical values analysis. These functionalities are used to specify the types of patterns that can be identified based on data analysis [54]. The objective is to identify those patterns that frequently occur in data, including frequent elements, subsequences, and frequent substructures. A frequent element usually refers to a group of elements that often appear together in a group of transactional data, for example, milk and bread that many clients often buy in supermarkets. A frequent subsequence is the pattern in which clients tend to buy first a laptop, followed by a digital camera and afterwards a memory card, it is a sequential pattern (frequent). A substructure can refer to different structural forms (for example, diagrams or trees) that can be combined with groups of elements or subsequences. If a substructure is often produced, it is called structures pattern [1].

Validation: Phase in which methods and techniques are implemented to validate the results, which are briefly described below:

1. **State-of-the-art revision:** Relevant research work, which is gathered from varied bibliographical sources, is analyzed. These works allow establishing basic concepts that support the research, as well as the algorithms used in its development.
2. **Descriptive and inferential analysis:** The data analysis in the log files is carried out applying descriptive statistics. To do that central tendency and dispersion measures can be implemented, as well as bar diagrams, boxes, and distributions.

Moreover, parametric and non-parametric inference tests can be used which allow identifying significant differences in data.

3. **Validation criteria:** The validation criteria used in data mining can include the Holdout method, in which the available data is divided randomly in two groups: a) The estimation group used to select the algorithm is 2/3 of the total data, and, b) The validation group used to test or validate the algorithm is 1/3 of the total data. Nevertheless, in some cases when there are low proportional values, it is recommended to increase the sample value in the training set and decrease the validation group. For example, 80% estimation and 20% for validation.

3.3 Patterns associated to learning

The most interesting patterns are presented to the final user and can be stored as new knowledge that can be used for decision making. These activities include information recollections from the log files, variable selection, data cleaning, information transformation, and finally, data mining algorithms to identify patterns associated to learning. Generally, patterns are shown in a graphic way that facilitates the comprehension from teachers and that are useful to make decisions inside and outside the classroom.

4 Study case

The results presented are based in a study case that uses and validates the proposed methodology by means of its application in the context of use of an Intelligent Tutoring System in Xalapa, Veracruz, Mexico.

4.1 Participants

The participants in this research were middle school students from Escuela Federal No. 2 “Julio Zárate” in the city of Xalapa, Veracruz, Mexico. The evaluation was carried out in the school year 2016-2017 with students chosen from a population of registered students (N=132) in the morning shift. The selection of students required a simple random test, which was considered based on the regular attendance to the math course (90% anticipated proportion), and it was evaluated with a 95% level of trust and a maximum permissible error (tolerance) of 6.5%, providing in this way a sample size of n=50 participants to obtain the minimum information needed during experimentation. The director’s approval from Escuela Secundaria Federal No.2 “Julio Zárate” was obtained to carry out this study, and the ethical norms for experimentation with human beings were followed [55], such as not recollecting personal or sensitive information from students.

4.2 Educational technological tool

The Intelligent Tutoring System “Scooter” is a software developed by [56] who combines the learning given by the teacher with the problems that each student should solve in the system and it is one of the methods used in teaching by means of a computer in the classroom. “Scooter” clearly specifies the objectives that each student should reach, knows the students’ needs, and can correct a student immediately after making a mistake. Moreover, “Scooter” covers a problem called “gaming the system” which consists in the fact that students use the help tools to obtain the answers to the exercises.

To avoid this problem, the tutor makes a question to the student and it should be answered correctly to continue providing help. The Intelligent Tutoring System “Scooter” has two working modalities: 1) reactive, in which the system tutor “Scooter” appears and reacts during all the student’s interaction; and 2) non-reactive: where the tutor is not shown to the student so that it does not interfere the learning process; this ITS has five parts (see Figure 2): 1) “Scooter”, 2) Problem to be solved, 3) Abilities, 4) Data, and 5) Answer sheet.

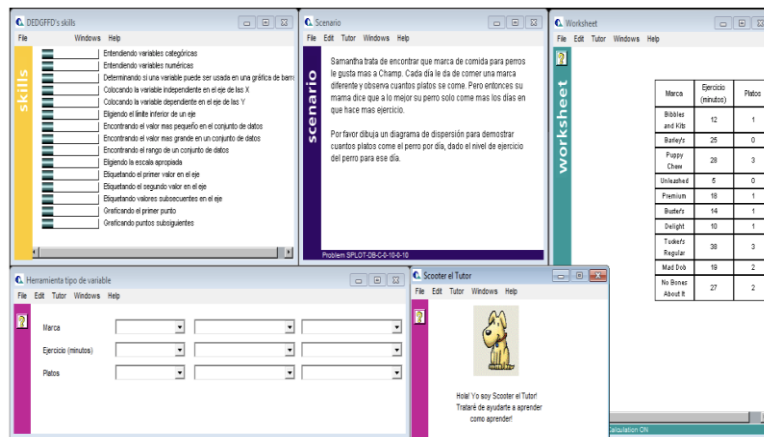


Fig. 2. Intelligent Tutoring System “Scooter”

Moreover, “Scooter” has two versions: 1) desk (original) and 2) web platform, which is an original tutor version implemented under the direction of a development group from the Tecnológico Superior de Alvarado, this is an external development to this research, and it was validated by [31].

The log files show the actions which students carry out with the Intelligent Tutoring System “Scooter” and they are stored in a file with a special format and extension defined by [54]. This file with extension p-0 contains too much information regarding the performance between the ITS “Scooter” and the student. These actions include: 1) if he/she answers correctly or not to a question, 2) requests for help, 3) amount of time to make the scenario, among others. The number of instances per student varies depending on the times that they interact with the system to solve an exercise and it is

stored based on a time sequence. Figure 3 is an example of log file, which contains too much information that needs to be transformed in a suitable format to be used for a selected data mining tool. File HEADER contains general student data, for example: name of student, lesson, section, starting time, name of the file, and the initial probabilities in each of the defines goals in the Bayesian Knowledge Tracing algorithm; and the file BODY contains the results from the student when solving an exercise presented by ITS “Scooter”.

Information filtering: It is highly important to filter the data because if this step is not done, non-processed data can lead to the extraction of patterns and rules that are not very useful, as they can contain: incomplete data, with noise, and even inconsistencies. The preprocessing (generally) reduces the initial data into a smaller set. The data analysis is centered in two numerical variables (registered time during the use of the tutor and learning probability) and two discrete variables (type of answer and success in the activity). In table 2 the statistics of the continuous variables for the 2,764 registers of the interactions between the students and the Intelligent Tutoring System “Scooter” are observed.

Table 2. Descriptive statistics of the continuous variables identified in the log files

	Mean	Standard deviation	Min	25%	50%	75%	Max	Coefficient of variation
Learning probability (p-know)	0.72	0.28	0.17	0.43	0.79	0.99	1.00	34.34%
Registered time	12.14	22.62	0.013	2.59	4.85	11.71	329.48	186.32%

4.3 Generation of structured data

The criteria to generate structured data is based on the results presented by [57] who define the main activities made by students when interacting with ITS “Scooter”, such variables integrate the basis of knowledge and are described in Table 3.

Table 3. Variables used in the predictive model

Variable	Type	Encoding
p-know	ordinal	{low, medium, high}
registered time	ordinal	{low, medium, quick}
input type	nominal	{blank, choice, number, point, string}
success	nominal	{right, wrong, help, bug}
class	nominal	{lacking, satisfactory, advanced}

```

*000:00:02 START
STUDENT; "EJGEFDD",
USER-DATA; ((DIAGRAM-CHOICE "Diagram Scatterplot Generation") (VARIABLE-TYPE-TOOL-CHOICE "Variable Type Tool") (WORKSHEET-CHOICE "Title Row Only Worksheet") (SPLOT2-INTERP-TOOL-CHOICE "Scatterplot Interpretation Tool2") (DA-INTERP-TOOL-CHOICE "Scatterplot Interpretation Tool") (SPLOT-SCALING-TOOL-CHOICE "Scatterplot Scaling Tool") (GAMING-AGENT-TOOL-CHOICE "Scooter the Tutor")),
PROBLEM; SPLOT-DB-C-0-10-0-10,
NPRESERVED; (0 0),
START-TIME; "18/27/10 23:05",
FILE-NAME; "C:\Users\jamy\Desktop\TAR\merge-all\Student Files\Data Analysis Course\Students\Mid School Splot Gen Gaming\EJGEFDD\1NSPLOT-DB-C-0-10-0-10%0.p-0",
LESSON; SCATTERPLOT-GENERATION-GAMING-2005,
LESSON-PRINT-NAME; "Mid School Splot Gen Gaming",
SECTION; SCATTERPLOT-GENERATION-FADING-SEC,
SECTION-PRINT-NAME; "Scatterplot generation problems",
*000:00:04 INITIAL-PROBABILITIES
p/p-1; (0.2 META-CHOOSE-OK-BG),
p/p-2; (0.2 META-CHOOSE-X-QUANTITATIVE),
p/p-3; (0.2 META-CHOOSE-Y-QUANTITATIVE),
p/p-4; (0.2 META-DETERMINE-MIN),
p/p-5; (0.2 META-DETERMINE-RANGE),
p/p-6; (0.2 META-DETERMINE-SETPAX),
p/p-7; (0.2 META-DETERMINE-SETPMIN),
p/p-8; (0.2 META-PLOT-FIRST-POINT),
p/p-9; (0.2 META-PLOT-POINT),
p/p-10; (0.2 META-QUANTITATIVE-VALUING-DETERMINE-ARBITRARY-SCALEKNOWN),
p/p-11; (0.2 META-QUANTITATIVE-VALUING-FIRST-BIN),
p/p-12; (0.2 META-QUANTITATIVE-VALUING-SECOND-BIN),
p/p-13; (0.2 META-VALUING-CAT-FEATURES),
p/p-14; (0.2 META-VALUING-NUM-FEATURES),
p/p-15; (0.2 META-VALUING-SCALE),
*000:00:05 WINDOW-SELECT
WINDOW-TITLE; ALGEBRA-2-TRANSLATOR::ALGEBRA-2-TRANSLATOR,
*000:32:177 READY
*000:33:946 APPLY-ACTION
WINDOW; ALGEBRA-2-TRANSLATOR::VARIABLE-TYPE-MODEL,
CONTEXT; SPLOT-DB-C-0-10-0-10,
SELECTIONS; ({var=0val=1}),
ACTION; SUBSTITUTE-TEXT-INTO-BLANK,
INPUT; ("CASA@CASA"),
*000:33:950 UPDATE-P-KNOW
META; META-VALUING-CAT-FEATURES,
PRODUCTION; (CHOOSE-VAR-TYPE-CAT MIDSCH-VARIABLE-TYPING),
SUCCESS?; T,
P-KNOW; 0.333333333333326,
*000:33:951 GOOD-PATH
*000:33:952 HISTORY
P-1; (FOCUS-ON-VARIABLE-TYPE-TOOL-SP-MODEL),
P-2; (CHOOSE-VAR-TYPE-CAT MIDSCH-VARIABLE-TYPING),
*000:33:953 READY
*000:44:517 APPLY-ACTION
WINDOW; ALGEBRA-2-TRANSLATOR::VARIABLE-TYPE-MODEL,
CONTEXT; SPLOT-DB-C-0-10-0-10,
SELECTIONS; ({var=0val=2}),
ACTION; SUBSTITUTE-TEXT-INTO-BLANK,
INPUT; ("No OK para gr: fcos de dispersión"),

```

HEADER

BODY

Fig. 3. Log file between student and the Intelligent Tutoring System “Scooter”

4.4 Selection of algorithm

The nature of the data allowed selecting the inductive learning in decision trees, which allows obtaining a set of rules to classify objects based on knowledge of a subset of objects with values given in their variables. A total of 1,384 registers for the generation of a predictive model were used, in this research the following variables were used: success of the activity, type of answer, p-know, time registered during the use of the tutor, and the class (associated learning); the corresponding values can be observed in Table 3.

Learning model: In the decision tree from Figure 4 one can observe that there are six possible final solutions: three for the lacking class, two for advanced, and one to classify satisfactory. Moreover, if the student responds: choice, number, or string, he/she will be classified as lacking. Another important aspect in the tree corresponds to the 977 registers in the training group, from which, 88% are students classified with a lacking learning which are equivalent to approximately 860 registers.

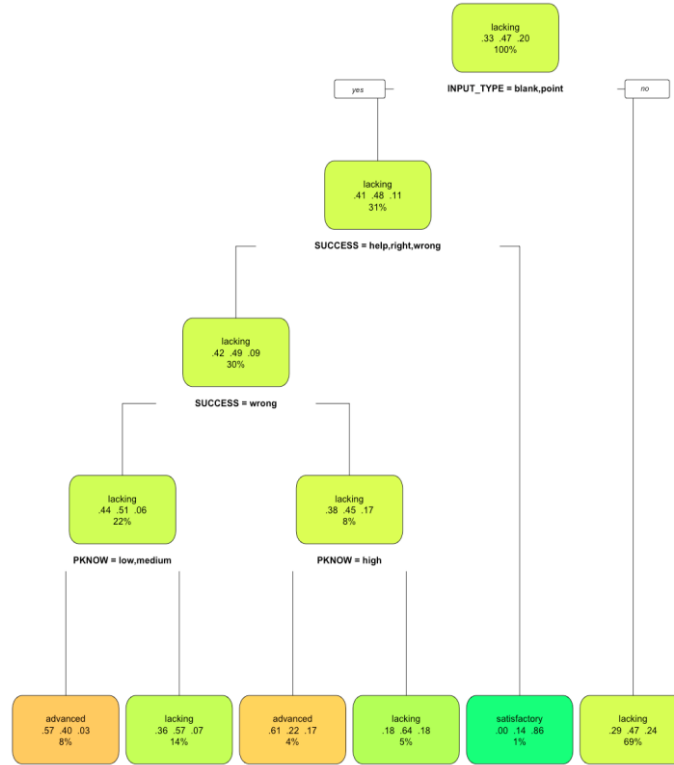


Fig. 4. Decision tree to identify patterns associated to learning

Validation of model: In the evaluation of the model approximately 70% of the total registers for the training of the model is used and the rest of the registers for its test (30%). In table 4 the confusion matrix algorithm applied to the knowledge basis is observed, based on this matrix the precision model is obtained; in the decision tree applied algorithm 44.47% precision is obtained, this is equivalent to all those that were lacking and were classified as lacking (160), as well as satisfactory-satisfactory (1) and advanced-advanced (20).

Table 4. Confusion Matrix of the decision tree algorithm

	lacking	satisfactory	advanced
lacking	160	84	112
satisfactory	0	1	0
advanced	24	6	20

4.5 Identification of behaviors

It is important to deduce the students' actions when interacting with the ITS based on the decision tree in Figure 4, remembering that in the terminal node (called leaf) the conclusions of the rules are formed. For it, the rules found are enlisted, specifically the ones that result in a lacking learning:

1. If the student's answer in the Intelligent Tutoring System= choice or number or string, then the learning associated to the student is lacking.
2. If the student's answer in the Intelligent Tutoring System= blank or point and the student does an activity = bug, then the learning associated to the student is satisfactory.
3. If the student's answer in the Intelligent Tutoring System= choice or number or string, then the learning associated to the student is lacking.
4. If the student's answer in the Intelligent Tutoring System= blank or point and the student does an activity = bug, then the learning associated to the student is satisfactory.
5. If the student's answer in the Intelligent Tutoring System = blank or point and the student does an activity = wrong and the learning probability (p-know) = high, then the learning associated to the student is lacking.
6. If the student's answer in the Intelligent Tutoring system= blank or point and the student does an activity = wrong and the learning probability (p-know) = low or medium, then the learning associated to the student is advanced.
7. If the student's answer in the Intelligent Tutoring System= blank or point and the student does an activity= help or right and the learning probability (p-know) = high, then the learning associated to the student is advanced.
8. If the student's answer in the Intelligent Tutoring System= blank or point and the student does an activity= help or right and the learning probability (p-know) = low or medium, then the learning associated to the student is lacking.

Based on the generated rules, the path to the help value (request for help) of the variable activity success (AS) in the decision tree (Figure 4) concludes that a student that requests help in the Intelligent Tutoring System "Scooter" will have a lacking performance or a satisfactory performance, but it will not reach an advanced learning. Based on the rules, teachers in charge of the class can identify the students with deficiencies in solving math problems and making the decision to take the appropriate correction actions.

5 Discussion

The objective of this research is to identify the behavior patterns which associate to the students' learning, based on the log files that Intelligent Tutoring System "Scooter" generates while students solve the math exercises that correspond to the SPLOT files (files that contain the activities done by students in exercises that correspond to the dispersion diagrams). The analysis of the log files allowed identifying the stu-

dents' relevant characteristics when interacting with ITS "Scooter". Afterwards, the determined variables (registered time in the use of the tutor, learning probability, type of response and activity success) were analyzed for the integration of a knowledge base. From which, a model is presented for the students' behavior identification when interacting with the technological tool Intelligent Tutoring System "Scooter", that identifies students with major advance in the system, as well as students with difficulties to solve math problems. As a result of the model it is concluded that the characteristics found will allow teachers to contribute in the teaching-learning process, due to the fact that students are grouped in 1) lacking, 2) satisfactory, and 3) advanced. Moreover, based on the model decision rules are generated that allow concluding the performance that students will reach when solving exercises regarding the dispersion diagrams.

In this sense, [44] mention how web mining can be applied in learning settings to predict the grades students will obtain in the final exam of a course, and this research is related to the results obtained as the focus on analyzing the behavior students perform in an educational setting and associate it with their learning inside the classroom. The findings in the research by [15-17] suggest that the students benefit from scaffolding techniques, detecting a positive effect in the student's interaction with learning settings. These results are important because they show the strategies students use to deal with the complexity of a specific task [58]. It is important to point that the predictive model identifies patterns that associate to students' behaviors when interacting with the math Intelligent Tutoring System "Scooter", additionally, it contributes to having the knowledge of students' behavior for teachers, assistants, parents and even students regarding how they interact with the ITS "Scooter". Therefore, the results presented in this research show that the proposed methodology allows modeling students' behaviors in a learning situation when interacting with a technological tool.

5.1 Future considerations

The findings In this aspect, it is suggested to explore a different point of view to generate the knowledge base from the log files, on the grounds of identifying new relevant characteristics that allow the application of other mining data algorithms, whose focus is to explore and analyze atypical data that provides learning patterns in students who take longer than the average time while using the Intelligent Tutoring System "Scooter". Moreover, as a future work it is suggested to perform the analysis of data that comes from different settings apart from solving exercises in different technological educational tools, that allows identifying new behaviors in students in regard to the new scenarios. Finally, applying this model in other contexts, with other experimental groups with a larger number of participants and whose objective is to adapt the steps of the development methodology to identify new patterns, providing a complementary vision in the teaching-learning strategies feedback.

6 Acknowledgement

The first author acknowledges the financial support from the Mexican Council of Science and Technology (CONACYT, Scholarship 442422). The corresponding author acknowledges the financial support of Writing Lab, TecLabs, Tecnológico de Monterrey, Mexico, in the production of this work, the second and third author gratefully acknowledges from the CONACYT.

7 References

- [1] Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41: 1432-1462. <https://doi.org/10.1016/j.eswa.2013.08.042>
- [2] R Oliveros, M.O.C., de Castro, A.E., Truyó, V.I.C., Flórez, D.G., Robles, H.S., Plaza, D.D., Fuentes, R.A.R. (2014). Comprensión lectora y TIC en la universidad. *Apertura*, 6: 46-59.
- [3] Self, J. (1998). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education*, 10: 350-364.
- [4] Xiong, W., Litman, D., Schunn, C. (2010). Assessing reviewers' performance based on mining problem localization in peer-review data. *Educational Data Mining 2010-3rd International Conference on Educational Data Mining*, 211-220.
- [5] Hershkovitz, A., Nachmias, R. (2011). Online persistence in higher education web-supported courses. *The Internet and Higher Education*, 14: 98-106. <https://doi.org/10.1016/j.iheduc.2010.08.001>
- [6] Kabakchieva, D., Stefanova, K., Kisimov, V. (2010). Analyzing university data for determining student profiles and predicting performance. *Educational Data Mining 2011*.
- [7] Thai-Nghe, N., Horváth, T., Schmidt-Thieme, L. (2010). Factorization models for forecasting student performance. *Educational Data Mining 2011*.
- [8] Junker, B.W. (2006). Using on-line tutoring records to predict end-of-year exam scores: experience with the ASSISTments project and MCAS 8th grade mathematics. *Assessing and modeling cognitive development in school: intellectual growth and standard settings*. JAM, Maple Grove, MN.
- [9] Pardos, Z.A., Heffernan, N. T. (2010). Using HMMs and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. *Journal of Machine Learning Research W & CP*, 40.
- [10] Baker, R.S., Corbett, A.T., Koedinger, K.R. (2004). Detecting student misuse of intelligent tutoring systems. *Intelligent Tutoring Systems. ITS 2004. Lecture Notes in Computer Science*, 3220: 531-540. https://doi.org/10.1007/978-3-540-30139-4_50
- [11] Wijaya, T.K., Prasetyo, P.K. (2010). Knowledge tracing with stochastic method. *Proceedings of the KDD 2010 cup 2010 workshop: Knowledge discovery in educational data*, 51-53.
- [12] Harsley, R. (2015). *When two heads are better than one: A Critical Review of Four Collaborative Intelligent Tutoring Systems*. Chicago, IL: University of Illinois at Chicago.
- [13] Suebnukarn, S., Haddawy, P. (2004). A collaborative intelligent tutoring system for medical problem-based learning. *Proceedings of the 9th international conference on Intelligent user interfaces*. 14-21. <https://doi.org/10.1145/964442.964447>

- [14] Walker, E., Rummel, N., Koedinger, K.R. (2009). Integrating collaboration and intelligent tutoring data in the evaluation of a reciprocal peer tutoring environment. *Research and Practice in Technology Enhanced Learning*, 4: 221-251. <https://doi.org/10.1142/s179320680900074x>
- [15] Angeli, C., Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. *Computers in Human Behavior*, 105: 105954. <https://doi.org/10.1016/j.chb.2019.03.018>
- [16] Duffy, M.C., Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, 52: 338-348. <https://doi.org/10.1016/j.chb.2015.05.041>
- [17] Radford, J., Bosanquet, P., Webster, R., Blatchford, P. (2015). Scaffolding learning for independence: Clarifying teacher and teaching assistant roles for children with special educational needs. *Learning and Instruction*, 36: 1-10. <https://doi.org/10.1016/j.learninstruc.2014.10.005>
- [18] Corbett, A.T., Koedinger, K.R., Anderson, J.R. (1997). Intelligent tutoring systems. *Handbook of human-computer interaction*, 849-874. <https://doi.org/10.1016/b978-044481862-1.50103-5>
- [19] Carbonell, J.R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11: 190-202. <https://doi.org/10.1109/tmms.1970.299942>
- [20] Barr, A., Beard, M., Atkinson, R.C. (1976). The computer as a tutorial laboratory: The Stanford BIP project. *International Journal of Man-Machine Studies*, 8: 567-596. [https://doi.org/10.1016/s0020-7373\(76\)80021-1](https://doi.org/10.1016/s0020-7373(76)80021-1)
- [21] Papert, S., Valente, J.A., Bitelman, B. (1980). *Logo: computadores e educação*. Brasiliense.
- [22] Koedinger, K.R., Anderson, J.R., Hadley, W.H., Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8: 30-43.
- [23] Albacete, P.L., VanLehn, K.A. (2000). Evaluating the effectiveness of a cognitive tutor for fundamental physics concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 22.
- [24] Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, 13: 173-197.
- [25] Tsiriga, V., Virvou, M. (2004). Evaluating the intelligent features of a web-based intelligent computer assisted language learning system. *International Journal on Artificial Intelligence Tools*, 13: 411-425. <https://doi.org/10.1142/s0218213004001600>
- [26] Woo, C.W., Evens, M.W., Freedman, R., Glass, M., Shim, L. S., Zhang, Y., Michael, J. (2006). An intelligent tutoring system that generates a natural language dialogue using dynamic multi-level planning. *Artificial intelligence in medicine*, 38: 25-46. <https://doi.org/10.1016/j.artmed.2005.10.004>
- [27] Pinkwart, N., Ashley, K., Lynch, C., Alevan, V. (2009). Evaluating an intelligent tutoring system for making legal arguments with hypotheticals. *International Journal of Artificial Intelligence in Education*, 19: 401-424. <https://dl.acm.org/doi/10.5555/1898071.1898076>
- [28] Lara-Muñoz, E.M., Rebolledo-Méndez, G., Rojano-Caceres, J.R., Huerta-Pacheco, N.S. (2018). Cultural Factors Linked to Collaborative Learning in Intelligent Tutoring System in the Domain of Mathematics. *ITS Workshops*, 41-50.

- [29] Huerta-Pacheco, N.S., Rebolledo-Mendez, G., Aguirre, V., Hernandez-Gonzalez, S., Lara-Muñoz, E.M. (2018). Relevance of the Cultural Dimensions in Affective-Cognitive Behavior During Interaction with an Intelligent Tutoring System. *ITS Workshops*, 33-40.
- [30] Lara Muñoz, E., Rebolledo Méndez, G., Rojano Cáceres, J. (2019). Mejorando el aprovechamiento de las actividades colaborativas por pares de estudiantes utilizando tecnología educativa en matemática. *Revista Complutense De Educación*, 30: 441-460. <https://doi.org/10.5209/iced.57597>
- [31] Rosas Escobedo, A., Lara-Muñoz, E.M., Huerta-Pacheco, N.S. (2019). Validacion de un sistema tutor inteligente web para el aprendizaje de las matematicas. *Komputer Sapiens*, 2: 18-22.
- [32] Peng, Y., Kou, G., Shi, Y., Chen, Z. (2008). A descriptive framework for the field of data mining and knowledge discovery. *International Journal of Information Technology & Decision Making*, 7: 639-682. <https://doi.org/10.1142/S0219622008003204>
- [33] Hand, D.J., Mannila, H., Smyth, P. (2001). *Principles of data mining (adaptive computation and machine learning)*. Cambridge, MA: MIT press, pp. 361-452.
- [34] Hong, T.P., Lin, K.Y., Wang, S.L. (2003). Fuzzy data mining for interesting generalized association rules. *Fuzzy sets and systems*, 138: 255-269. [https://doi.org/10.1016/s0165-0114\(02\)00272-5](https://doi.org/10.1016/s0165-0114(02)00272-5)
- [35] Hardoon, D.R., Szedmak, S., Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16: 2639-2664. <https://doi.org/10.1162/0899766042321814>
- [36] Wu, Z., Li, C. H. (2007). Lo-constrained regression for data mining. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 981-988. https://doi.org/10.1007/978-3-540-71701-0_110
- [37] Georgila, K., Core, M.G., Nye, B.D., Karumbaiah, S., Auerbach, D.; Ram, M. (2019). Using Reinforcement Learning to Optimize the Policies of an Intelligent Tutoring System for Interpersonal Skills Training. *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 737-745. <https://dl.acm.org/doi/10.5555/3306127.3331763>
- [38] Zhuhadar, L., Marklin, S., Thrasher, E., Lytras, M.D. (2016). Is there a gender difference in interacting with intelligent tutoring system? Can Bayesian Knowledge Tracing and Learning Curve Analysis Models answer this question? *Computers in Human Behavior*, 61: 198-204. <https://doi.org/10.1016/j.chb.2016.02.073>
- [39] Taub, M., Azevedo, R., Rajendran, R., Cloude, E.B., Biswas, G., Price, M.J. (2019). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction*, 101200. <https://doi.org/10.1016/j.learninstruc.2019.04.001>
- [40] Felder, R.M., Silverman, L.K. (1998). Learning and teaching styles in engineering education. *Engineering education*, 78: 674-681.
- [41] Baker, R.S., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68: 223-241. <https://doi.org/10.1016/j.ijhcs.2009.12.003>
- [42] Baker, R. S., Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics*, 61-75. https://doi.org/10.1007/978-1-4614-3305-7_4
- [43] Kaur, P., Singh, M., Josan, G. S. (2015). Classification and prediction-based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science*, 57, 500-508. <https://doi.org/10.1016/j.procs.2015.07.372>

- [44] Romero, C., López, M. I., Luna, J.M., Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68: 458-472. <https://doi.org/10.1016/j.compedu.2013.06.009>
- [45] Romero, C., Ventura, S. (2012). Data mining in education. *Data Mining and Knowledge Discovery*, 3: 12-27. <https://doi.org/10.1002/widm.1075>
- [46] Conde, Miguel Á., Hernández-García, A. (2019). Data Driven Education in Personal Learning Environments—What About Learning beyond the Institution? *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI)* 1: 43-57. <https://doi.org/10.3991/ijai.v1i1.11041>
- [47] Pyle, D. (1999). *Data preparation for data mining*. 1st ed.; Morgan Kaufmann Publishers Inc. San Francisco CA. United States.
- [48] Baker, R.S., Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1: 3-17. <https://doi.org/10.5281/zenodo.3554657>
- [49] Romero, C., Romero, J.R., Luna, J.M., Ventura, S. (2010). Mining rare association rules from e-learning data. *Educational Data Mining*.
- [50] Ferguson, R. (2012). The state of learning analytics in 2012: A review and future challenges. Knowledge Media Institute, Technical Report, 1.
- [51] Siemens, G., Baker, R.S.D. (2012). Learning analytics and educational data mining: towards communication and collaboration. *Proceedings of the 2nd international conference on learning analytics and knowledge*, 252-254. <https://doi.org/10.1145/2330601.2330661>
- [52] Hershkovitz, A., de Baker, R.S.J., Gobert, J., Wixon, M., Pedro, M. S. (2013). Discovery with models: A case study on carelessness in computer-based science inquiry. *American Behavioral Scientist*, 57: 1480-1499. <https://doi.org/10.1177/0002764213479365>
- [53] Suhirman, Jasni, M.Z., Haruna, C., Tutut, H. (2014). Data Mining for Education Decision Support: A Review. *International Journal of Emerging Technologies in Learning*, 9: 4-19. <http://dx.doi.org/10.3991/ijet.v9i6.3950>
- [54] Han, J., Kamber, M., Pei, J. (2011). *Data mining concepts and techniques*. 3th ed.; Morgan Kaufmann Publishers Inc. San Francisco CA. United States.
- [55] Department of Health, E. (2014). The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. *The Journal of the American College of Dentists*, 81: 4.
- [56] Baker, R.S. (2005). *Designing intelligent tutors that adapt to when students game the system* (Doctoral dissertation, Carnegie Mellon University).
- [57] Melendez-Armenta, R., Rebolledo-Mendez, G., Huerta-Pacheco, N., Fernandez-Lambert, G. (2020). Can data mining in education identify students' learning in a mathematics tutoring system? Case study of students at secondary level in Mexico. *Proceedings of INTED2020*. IATED, 7796-7803. <https://doi.org/10.21125/inted.2020.2127>
- [58] Sunday, K., Ocheja, P., Hussain, S., Oyelere, S., Samson, B., & Agbo, F. (2020). Analyzing Student Performance in Programming Education Using Classification Techniques. *International Journal of Emerging Technologies in Learning (IJET)*, 15: 127-144. <https://doi.org/10.3991/ijet.v15i02.11527>

8 Authors

Roberto Angel Melendez-Armenta received the B. Eng. in Computer Sciences from Benemérita Universidad Autónoma de Puebla in 2007 and master's degree in

Artificial Intelligence from Universidad Veracruzana in 2011. Since 2012, he has been full time professor Instituto Tecnológico Superior de Misantla. His scientific interests are focused on intelligent computing, data mining and artificial intelligence in education. He was Ph.D. in Computer Sciences student at Universidad Veracruzana from 2016 to 2019.

N. Sofia Huerta-Pacheco is a Professor of Research at the Universidad Nacional Autónoma de México with a Ph.D. in Computer Sciences from the Universidad Veracruzana and an MSc in Advanced Multivariate Data Analysis from the University of Salamanca. Her research interests comprise Computational Statistics, Multivariate Statistics, Statistics Education, Forensic Science, and Affective Computing in education

Luis Alberto Morales-Rosales graduated in 2009 from the National Institute of Optical and Electronic Astrophysics as a Ph.D. in Computer Sciences. He is part of the Mexican National System of Researchers, distinguished as Level 1. Currently, he works at Michoacan University of San Nicolás de Hidalgo (UMSNH) as Professor Conacyt, assigned to the master's degree in Transportation Infrastructure in the Land Roads Branch. His research includes Computational Models for Transportation Engineering, Intelligent Computing, Security and Distributed Systems with interdisciplinary and inter-institutional collaborations.

Genaro Rebolledo-Mendez obtained his D.Phil degree in Computer Systems and Artificial Intelligence from the University of Sussex, United Kingdom. Genaro has been an active faculty member in academic institutions in the United Kingdom, the United States, and Mexico. He was a full-time researcher and Dean of the Informatics Department at the Universidad Veracruzana as well as a member of Mexico's National Research System (SNI I). He is editorial member of the International Journal of Artificial Intelligence in Education. Currently, he is a Mentor in the Writing Lab, Campus Monterrey, Tecnológico de Monterrey, Mexico. Email: g.rebolledo@tec.mx

Article submitted 2020-07-15. Resubmitted 2020-08-01. Final acceptance 2020-08-02. Final version published as submitted by the authors.