Predictive Model for Assigning Exercises to Students in Spreadsheet Functions Using Artificial Neural Networks

Edwar Saire-Peralta [1]

[1] *Universidad Nacional de San Agustín de Arequipa, Av. Independencia s/n, Arequipa, Perú*

Abstract

The objective of the article is the development of a model that allows to predict the exercises that the student can solve, and on the other hand the exercises that the student cannot solve in the course of ®Microsoft Excel basic level with the topics of functions. For the development of the process, artificial neural networks have been used. The model is fed with data such as sex, age, academic grade, parents' level of education, type of school, previous grades of the topics that the student obtains while advancing in the course. The research approach is quantitative, experimental, applied and the population was represented by 85 students. The result shows that the model achieves 72% probability of prediction in the assignment of exercises to students. These exercises could not be solved were provided with an aid for their resolution.

Keywords

Artificial neural networks, Supervised learning, Data mining, Cross validation

1. Introduction

The teaching-learning process is integral, according to [1] points out that, if the conditions of the students are always different, such as the rhythms, ways of learning and starting points of each student, then, what is learned and what is evaluated cannot be standardized, but must be differentiated according to the individual characterization. [2] Indicates that student's process information according to their capacity, motivation, environment and the guidance provided by the teacher in their learning. Learning rhythms are linked to academic performance, which is determined by personal, family, social and educational factors [3, 4]. A learning session in the classroom is represented by several moments, one of them represents the practice, which mostly aims to have students solve exercises regarding the topic developed. It has been observed that many students have doubts and certain fears when interacting with new learning topics; they are students who find it difficult to adapt to the pace of progress and understanding imposed by the majority of students and even by the teacher. This reality is measurable through the results of the evaluations. [5] Propose that the teacher should work at a safe level of demand, which does not cause discouragement and low grades. According to [6] indicates that it is a mistake to use the same contents, rhythms and evaluation to students, this is a problem because it can cause frustrations and influences the relationship with other students. The situation described is very common in classrooms, and many researches have used predictions to find the most suitable ways to know the student based on certain data about them and give help.

2. Related work

CISETC 2021: International Congress on Educational and Technology in Sciences, November 16-18, 2021, Chiclayo, Peru
EMAIL: esaire@unsa.edu.pe (A.1)
ORCID: https://orcid.org/0000-0002-9526-0205 (A.1)
© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3037/paper4.pdf

Related and pre-research work is characterized by the use of one or more classification algorithms. Some research uses as input data those traces or interactions that students have with virtual platforms, others use as inputs those data that are collected through instruments and that are designed at the time. The literature has been reviewed and the opportunity to make predictions with data that arise from the teaching-learning process is observed. In Table 1 we can see a summary of works related to the research.

Table 1
Summary of previous Works

Authors	Title	Contribution	Opportunities for improvement
[7]	ICT for education: adaptive system based on automatic learning mechanisms for the appropriation of technologies in high school students.	A system was built that enables the initial recommendations of educational content appropriate to the individual characteristics of students, administered according to their performance and the characteristics of the territory.	The system was fed with data collected from virtual courses, both from students and educators, data analytics and automatic learning to make predictions and initial recommendations, is limited by the classroom courses, since we do not have the necessary data.
[8]	Model to predict academic performance based on neural networks and learning analytics	A predictive model of academic performance was proposed using data provided by a virtual interaction system with students, using learning analytics through artificial neural networks, patterns were found that were determinant in the academic performance of students.	The data collected and used to make the predictions are from the virtual system of courses they have, however, there is still an opportunity for improvement if personal, social and other data of interest to the research are included.
[9]	Predicting academic performance by applying data mining techniques	Shows a range of predictions to classify (pass, fail) prospective students enrolled in a course. Data mining techniques were used and results were compared using logistic regression,	To obtain the classification, the data of the students enrolled in the General Statistics course at UNALM were used; however, the factors or

		decision trees, neural networks and Bayesian networks. A prediction effectiveness of 70% was achieved.	predictor variables were selected based on the data they already had, without taking into account, according to research, that there are very influential variables in academic performance, which was not taken into account in their research.
[10]	Development of a computerized evaluation system using neural networks through R and Shiny	Through the use of artificial neural networks, an environment of attention to the needs of each student was created with the use of correct materials through exercises in their evaluation. This allows to reduce the feeling of dissatisfaction and to avoid in many cases the abandonment of the courses.	Data generated by the same project have been used. The results obtained were different levels of difficulty for the students in the exercises; however, there is still room to analyze other determining factors and to take into account the levels reached by students in the previous topic, since this process is changing.

3. Problem

Students in educational centers are characterized by being unique, singular and belonging to heterogeneous groups. In each learning session the teacher tries to improve his work with the students, especially when developing the practical part, where the teacher usually leaves a set of exercises during the class, which the whole group must solve in a certain time. [11] indicates that a school model where teachers teach the same contents, with the same level of complexity and at the same speed, this school is not attending to the differential needs of the students. It has been observed during the classes that students, when solving the battery of exercises, need support, tutoring, help in some formulas, in their application and syntax. The teacher is regularly confronted with two situations: first, when the student asks the teacher for help or tutoring, time is always pressing, and second, many students need help, but do not ask for it. Diversity refers to heterogeneous groups of students in the classroom. Students are unique and different, through their learning styles, ways of thinking and speed of learning within their limitations [2].

4. Proposal

To address the stated problem, a model has been implemented that allows predicting the assignment of exercises to students on an individualized basis, providing textual help in the exercises that the student cannot solve. In the proposal, a predictive model based on artificial neural networks will be implemented. This algorithm has a set of interconnected elements, where its processing capacity is stored in weight units, this thanks to the adaptation and learning of a set of patterns [12]. The proposed Model has as input the personal and social factors and the academic performance qualifications of the students, which will allow us to classify the students with data mining. With mining we can obtain models that allow discovering patterns and trends regarding student information [13]. The outline of the proposal is shown in Figure 1.

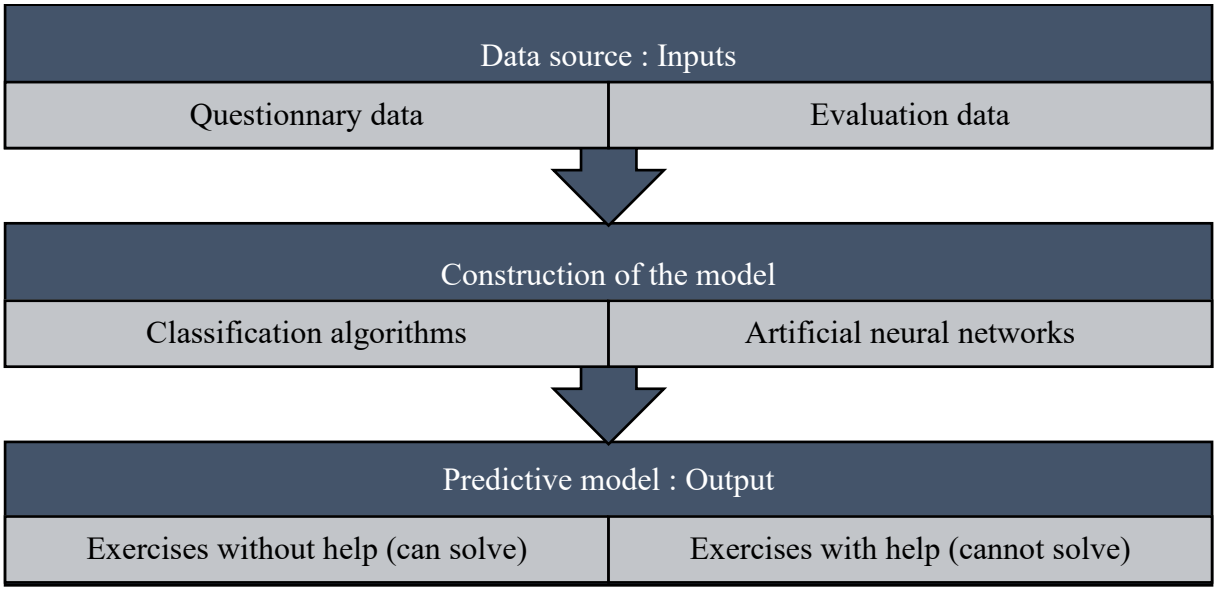

Figure 1: Structure of the Model

4.1. Population

The students who participated in the research took the ®Microsoft Excel basic level course. The topics that were developed in the course are mathematical operators, mathematical functions, and statistical functions, among others. There is no filter to enroll students, anyone can take it. We work with heterogeneous groups. In total we worked with 85 students as the population, which also represents the sample.

4.2. Questionnaire data

An instrument based on the survey technique was constructed. For the elaboration of the questionnaire instrument, the literature on those factors that influence students in the handling of function subjects was reviewed, in addition to adding other factors contextualized to the problem being addressed. Academic achievement, being multicausal, according to [14] can be grouped into social, personal and institutional determinants. Table 2 shows the factors taken into account for the questionnaire.

Table 2
Classification of attributes

Individual	Academic	Socioeconomic	Institutional
	Person with numerical skills	Current occupation	
	You were taught	How many hours a day do you work	
Age	computer skills in	Father's academic level	
Sex	school	Mother's academic level	Type of school
Marital status	Experience with Excel	Reason for studying	
	Academic degree	Excel	
	Hours of study during the day		

The instrument was validated with a psychologist in Education and a professional in Educational Sciences, the reliability of the instrument was calculated, applying cronbach's alpha, where the result was 0.733, which represents a value of good in reliability. With the validity and reliability obtained, the questionnaire was applied to 85 students who took the course.

4.3. Evaluation data

The data collection regarding evaluation represents data from 58 students. The evaluation grades is an indicator that determines the academic performance of students, as stated by [15], where it indicates that academic performance is the level of knowledge that a student has which is reflected in a numerical value, where it measures the result of the teaching and learning process in which the student is the main actor. For each of the 8 topics of the course, exercises were prepared. For each topic, 10 types of exercises were designed. A total of 80 types of exercises were prepared.

5. Application of the Methodology

To develop the proposal, the KDD (Knowledge Discovery in Databases) data mining process was followed, as described in [16]. The KDD process is a rare process that allows obtaining information from the data, which is present in a hidden way, initially anonymous and very useful for users or companies [17].

5.1. Integration and Collection Phase

The data sources were merged. The first data source was obtained by applying the questionnaire and the second data source was formed by the evaluation data collection (it was recorded for each topic and type of exercise whether the student could or could not solve the exercise). In the Figure 2, we can see the evaluations recorded. The value of 1 indicates that a student was able to solve one type of exercise and a value of 0 indicates that the student was not able to solve that type of exercise. As mentioned for each topic, 10 types of exercises were designed with labels from letter A to J.

	Tema 01											Tema 02											Tema 03											Tema 04										
CodEstudiante	A	B	C	D	E	F	G	H	I	J	Nota1	A	B	C	D	E	F	G	H	I	J	Nota2	A	B	C	D	E	F	G	H	I	J	Nota3	A	B	C	D	E	F	G	H	I	J	Nota4
1	1	0	1	1	1	1	1	1	1	0	16	1	1	1	1	1	1	0	1	1	1	18	1	0	1	1	1	1	1	1	1	0	16	1	1	1	1	1	1	1	1	1	1	20
2	1	1	1	1	0	1	1	1	1	1	18	1	1	1	0	1	0	1	1	1	1	16	1	1	1	1	1	1	1	1	1	1	20	1	1	1	1	1	1	1	1	1	1	20
3	1	0	1	1	0	1	1	1	1	0	14	1	1	0	1	1	1	1	0	0	0	12	0	1	1	1	1	1	0	1	1	0	14	1	1	1	1	0	0	1	0	0	0	10
4	0	1	1	1	0	1	0	0	0	0	8	1	1	0	1	1	0	1	1	1	1	16	1	0	1	1	0	1	1	1	1	1	16	1	1	1	1	1	1	1	1	1	1	20
5	1	1	1	1	1	1	1	1	1	1	20	1	1	1	1	0	1	1	1	1	1	18	0	1	1	1	1	1	1	1	1	1	18	1	1	1	1	1	1	1	1	1	1	20

Figure 2: Recorded evaluations

5.2. Selection, cleaning and transformation phase

The selection phase involved the use of all the data from the questionnaire with the totality of the records collected. The cleaning phase was applied to the assessments with students who did not have assessment scores. Finally, in the transformation phase all the questionnaire data were replaced by numerical codes in order to process the data. The only field calculated was grade with a value from 0 to 20. The two data sources were integrated and linked. Finally, the data were normalized, since there were blunt values, for better processing quality.

5.3. Data mining phase

The data mining technique applied to the proposed project is classification. The collected data were separated, where 92% were assigned for training and 8% for model validation. The Artificial Neural Networks algorithm with supervised learning Backpropagation was used. This supervised learning algorithm is based on the repetition of the adjustment of the synaptic weights in the network, with the aim of minimizing the difference in error between the expected and observed results, achieving the most optimal [18]. To obtain the predictive model we used the free tool based on artificial Neural Networks, which is called "Simbrain". In Figure 3 shows the network topology for the first subject.

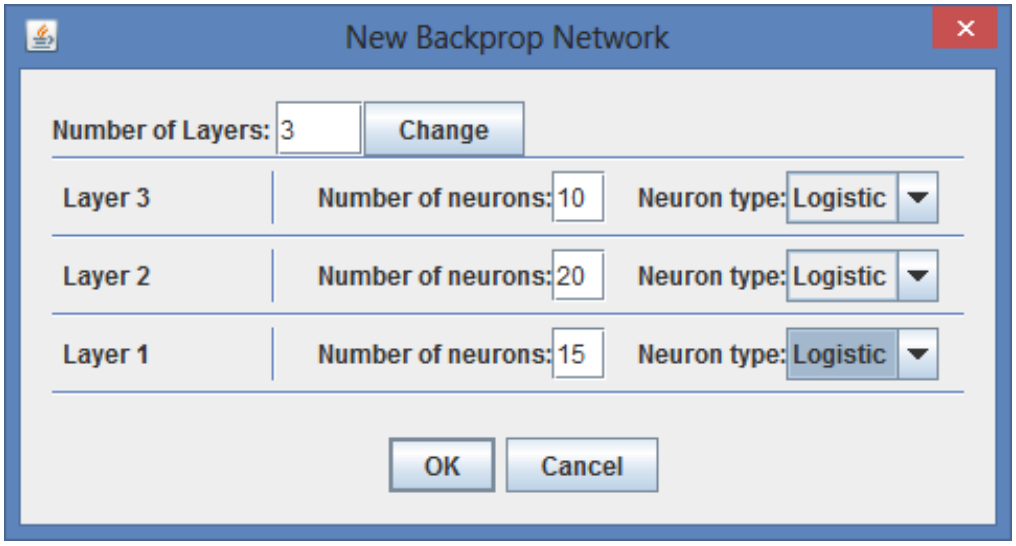

Figure 3: Topology of the first subject

Layer 1 represents the input layer (15 attributes of the questionnaire), layer 2 refers to the hidden layer with 20 neurons and finally layer 3 refers to the output layer with 10 answers. To train the model for the second topic (mathematical functions), the topology must now have 16 inputs, which represents the 15 student questionnaire data and the grade obtained for the academic performance of the previous topic (first topic) and so on will increase the inputs for the other topics. There are numerous researches such as those of [19, 20, and 21] have found evidence that the previous performance of their academic performance could condition future results. Table 3 shows results for the first 5 topics.

Table 3
Final topology of the proposed model

Topic	Input layer	Hidden layer	Output layer
Topic 01	15	20	10
Topic 02	16	30	10
Topic 03	17	20	10
Topic 04	18	20	10
Topic 05	19	20	10

5.4. Evaluation and Interpretation

Cross-validation was performed with 8% of the records that were initially separated. The results are show in Table 4 for the first item. Recall that the value 1 represents that the student can solve the exercise and the value 0 represents that the student cannot solve the exercise. The prediction on the set of 5 students had a reliability of 72%.

Table 4
Model prediction for students

Student	Model result	Expected result	Output layer
1	1111111111	1111011100	70%
2	1111111111	1111011101	80%
3	1110000100	1010000111	70%
4	1011011110	1111111101	60%
5	1111011110	1011011111	80%

5.5. Dissemination and Use

At the end of the evaluation, teachers as well as students were satisfied with the results, since an assertiveness of 72% was achieved. The strength of the model is to identify those types of exercises where students show difficulty, and it is in this space where the student will be supported.

6. Application and testing

Based on the predictive model obtained, its effectiveness was tested by selecting experimental groups (group of students new to the course), to which the predictive model was applied for the first two topics. In addition, control groups were selected (groups of students new to the course) where the first two topics were also developed, but applying the traditional model. A total of 3 experimental groups and 3 control groups were used for testing. In Table 5 we can see the averages obtained by the experimental and control groups, where the average increase was from 13.4 to 17.2.

Table 5
Results of the groups

Group N°	N°. topic	Experimental	Control
Group 01	Topic 01	16.3	13.8
	Topic 02	17.4	12.0
Group 02	Topic 01	18.3	15.4
	Topic 02	16.8	12.0
Group 03	Topic 01	17.9	15.2

Average	17.2	13.4

To give more reliability support to the obtained predictive model, the averages of the current results of the predictive model were compared with the averages of students from previous years and months (historical data of averages of 3 months of the previous year) and it was seen that the proposed model also improves the averages from 13.3 to 16.4.

7. Discussion and conclusions

It is reflected and indicated that, in order to obtain efficient predictive models, it is necessary to work not only with a greater amount of data, but also that these data must be of quality, must be selected by experts in this discipline, data that other researches support. Many times institutions already have data in their virtual systems [22, 23], but we must also measure the quantity of these data against the quality. It has been shown that the proposed model can be successfully used to predict the types of exercises that a student can solve and the types of exercises where he/she shows difficulties. The model was exposed to a cross-validation, which had a prediction close to 72% with respect to the expected results. An increase in their average from 13.49 to 17.29 in their evaluations was observed. This research not only validated the model with 8% of the students, but also tested the model with new groups of students in the institution. This model can be improved by working with more students and more that are related to academic performance, since the learning would be more solid, in addition to adding the characteristic that the exercises should be assigned gradually, i.e., classify the exercises by levels such as basic, intermediate and advanced.

8. References

[1] R. Anijovich y G. Cappelletti. La evaluación como oportunidad. Buenos Aires: Paidós, 2017, pp. 21-22.

[2] M. L. Méndez, "Diversidad en el aula", *Revista Digital: Innovación y experiencias Educativas*, 41, 1(9), 2011.

[3] N. Medina, J. Fereira and R. Marzol, "Factores personales que inciden en el bajo rendimiento académico de los estudiantes de geometría", Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales, 20(1), 4-28, 2018. URL: https://dialnet.unirioja.es/descarga/articulo/6436353.pdf

[4] M. Saucedo, S. Herrera-Sánchez, J. Díaz, S. Bautista and H. Salinas, "Indicadores de reprobación: Facultad de Ciencias Educativas (UNACAR)". *Revista Iberoamericana para la Investigación y el Desarrollo Educativo RIDE*, 5(9), 1-11, 2014. doi: https://doi. org/10.23913/ride.v5i9.7

[5] A. Anaya-Durand and C. Anaya-Huertas, "¿Motivar para aprobar o para aprender? Estrategias de motivación del aprendizaje para los estudiantes". *Tecnología, Ciencia. Educación*, 25 (1), 5-14, 2010. https://www.redalyc.org/articulo.oa?id=48215094002.

[6] J. Tourón, R. Santiago and A. Díez. The Flipped Classroom: Cómo convertir la escuela en un espacio de aprendizaje. Grupo Océano. 2014.

[7] A. Otero, W. Rivera, C. Pedraza, and J. Canay. (2019), "TIC para la educación: sistema adaptativo basado en mecanismos de aprendizaje automático para la apropiación de tecnologías en estudiantes de educación media", *Telos: Revista de Estudios Interdisciplinarios En Ciencias Sociales*, 21(3), 526–543, 2019. doi: https://doi.org/10.36390/telos213.03.

[8] N. Salgado Reyes, J. Beltrán Morales, J. Guaña Moya, C. Escobar Teran, D. Nicolalde Rodriguez, and G. Chafla Altamirano, "Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje", *Revista Ibérica de Sistemas y Tecnologías de Información*, 1, 258–266, 2019. https://uvirtual.uce.edu.ec/

[9] C. H. Menacho Chiok, "Predicción del rendimiento académico aplicando técnicas de minería de datos", *Anales Científicos*, Vol 78(1), 26, 2017. doi: https://doi.org/10.21704/ac.v78i1.811

[10] J. M. Gutiérrez Cárdenas and F. Casafranca Aguilar. Implementation of a Computerized Assessment System by using Backpropagation Neural Networks with R and Shiny, 2015. http://hdl.handle.net/11354/1087
[11] J. Tourón and R. Santiago, "El modelo Flipped Learning y el desarrollo del talento en la escuela", *Revista de Educación*, 368, 196-231, 2015. doi: https://doi.org/10.4438/1988-592X-RE-2015-368-288.
[12] K. Gurney. An introduction to neural networks, London, UK: UCL Press, 1997.
[13] C. Romero, S. Ventura, E. García, "Data mining in course management systems: Moodle case study and tutorial", *Computers and Education*, 51(1), 368-384, 2008. doi: https://doi.org/10.1016/j.compedu.2007.05.016
[14] G. M. Garbanzo Vargas, "Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública", *Revista educación*, 31(1), 43 – 64, 2007. https://revistas.ucr.ac.cr/index.php/educacion/article/view/1252
[15] S. Cueto. "Una década evaluando el rendimiento escolar. Organización Grupo de Análisis para el Desarrollo". Lima: GRADE, 2006.
[16] J. H. Orallo, M. J. R. Quintana, C. F. Ramírez. Introducción a la minería de datos, Person Educación, S.A. Madrid, ISBN: 978-84-205-4091-7, 2004.
[17] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth. From data mining to knowledge discovery in Databases: an overview. Ai Magazine. pp. 37-54, 1996.
[18] D. E. Rumelhart, G. E. Hinton and R. J. Williams. Learning representations by back-propagating errors. Nature, 323, 533-536, 1996.
[19] J. R. Betts and D. Morell, "The Determinants of Undergraduate Grade Point Average. The Relative Importance of Family Background, High School Resources, and Peer Group Effects", *The Journal of Human Resources*, Vol. 34 No. (2), 1999. doi: https://doi.org/10.2307/146346
[20] A. Porto and L. Di Gresia: "Rendimiento de Estudiantes Universitarios y sus Determinantes," Presentado en la Asociación Argentina de Economía Política, 2001
[21] R. A. Naylor and J. Smith, "Determinants of Educational Success in Higher Education," *International Handbook in the Economics of Education*, Elgart, 2004.
[22] R. Timarán-Pereira, J. Caicedo-Zambrano and A. Hidalgo-Troya, "Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°", *Revista de Investigación, Desarrollo e Innovación*, 9(2), 363-378, 2019. doi: 10.19053/20278306.v9.n2.2019.9184.
[23] J. Zárate-Valderrama, N. Bedregal-Alpaca and V. Cornejo-Aparicio, "Modelos de clasificación para reconocer patrones de deserción en estudiantes universitarios", *Ingeniare. Revista chilena de ingeniería*, 29(1), 168-177, 2021. doi: http://dx.doi.org/10.4067/S0718-33052021000100168.