Big Data Predictive Analytics in Educational Database Systems

Miguel-Angel Quiroz-Martinez[*], Gabriele-Enmanuel Torres-Secaira, Santiago-Teodoro Castro-Arias[1] and Maikel-Yelandi Leyva-Vazquez[1]

[1] Computer Science Department, Universidad Politécnica Salesiana, Guayaquil, Ecuador

ABSTRACT

Through computer algorithms, daily face-to-face or online classes knowledge can be transmitted efficiently by improving student performance in knowledge assessments. The critical success factor is a management term for an element necessary for a project to achieve its mission, and it is essential to model and analyze them in big data projects. This document presents a critical framework for success factor analysis for Big Data Predictive Analytics in educational databases defining an appropriate architecture and the availability of data in that order are the main factor. The document concludes with the conclusion and recommendation of future work.

Keywords: Predictive analytics · big data · educational databases · fuzzy decision maps

INTRODUCTION

Including advanced technology in educational institutions is now essential to improve your education methods and increase the academic level and student performance. To this end, educational stakeholders have the challenge of identifying the student's difficulties in the subjects and preventing them from being approved with good grades.

One possible solution is to conduct surveys on the difficulty of the subjects and evaluate the responses of the students; an alternative is the Big Data program, which through a practical predictive analysis, provides the ease of changing the teaching method so that the student captures the content of the subject and can also obtain good scores when evaluating it. That is why Big Data has become a necessity for human beings, not only in the educational area but also in the medical and environ-mental branches, where it assists in the wellbeing of humanity.

Hypothesis question: Could the academic level in schools be improved by applying Big Data with their predictive analyses?

Through computer and human algorithms, they confer the most didactic method to efficiently transmit class knowledge, either in person or online, and get students to improve their performance in knowledge assessments.

The objective is to explore the results obtained by educational institutions in Latin America where they applied Predictive Analytical Big Data in Educational Database Systems. Their academic level and study methodology could be improved for stu-dents who are pursuing a career they chose.

PRELIMINARY

This segment consists of informing what Big Data consists of a small piece of knowledge of what it is and its function, then the rest of the segment consists of ap-plying it with the Lala Project software.

BIG DATA

The need to apply Big Data in higher education institutes to update effective educational techniques and teaching methods were identified; at different universities, with their students, a survey was conducted to collect data; therefore, if it was favorable to apply them, as he predicted that the mistakes they made in teaching; this was able to evolve learning methods for students.

Companies were not known that by getting customer data daily when selling prod-ucts, they could use it to increase profits; through social groups, the desire of the peo-ple in which the product was purchased was observed, turning them into valuable information; obtained the data, through an analysis, companies predicted that the products sell to customers; A significant advantage when using data to predict large sales.

Currently, chronic non-communicable pathologies have a high incidence in the world's

population, including diabetes mellitus; Through an open-source tool, the age of onset of diabetes symptoms could be predicted in patients with a predisposition to diabetes, being decisive for early therapeutic measures that improve the quality of these.

They wanted to optimize with Big Data cloud storage data that had previously been physically archived, as manual annotation had been associated, due to a lack of time to detail the entire procedure, lack of medical care, and increased mortality rates; at the time of the triage, through the code, they could take care of the patient quickly, focus the attention time based on gravity, with this classification tool avoid deceased.

Given the enormous amount of information collected in different areas, it was necessary to use the process to reduce and classify data, which would not otherwise have been classified, often discarding data that could be useful when performing an analysis to predict environmental or economic catastrophes; these old flaws and errors could be eliminated thanks to the systematized data classification.

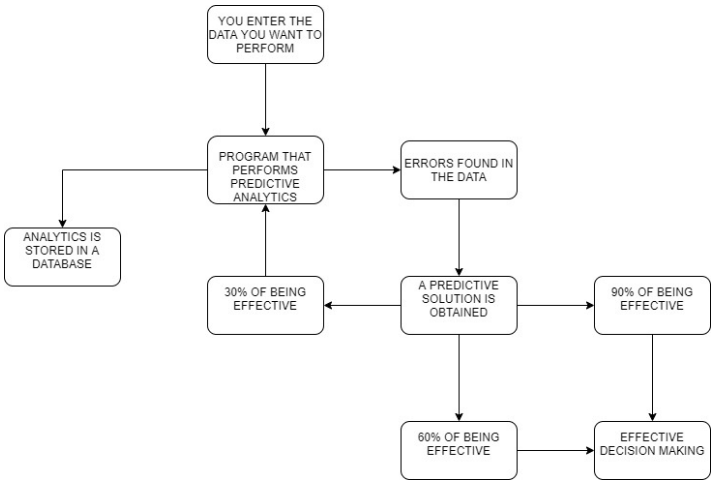

Fig. 1. Architecture on the operation of predictive analysis

Fig.1 you see a predictive analysis of the data that you want to know which effective solution to take:

- At first, problem data is digitized to analyze.
- This data provides predictive analytics.
- All analyses are stored in a database for future study.
- The analysis looks for what error is, whether it needs to be changed, or whether it is feasible

to proceed to the next step.

- In the end, several possible solutions are obtained with their percentage of effectiveness in making good decisions to the problem.

- If the feasibility rate is more than 60% than 90%, decision-making is obtained for the effective solution.

- If the feasibility rate is less than 30%, the problem is reanalyzed

USING THE LALA PROYECT

Lala is a Hispanic organization where she carries out projects about analytical predictions where they want to improve instructional learning in educational institutions in Latin American countries; safeguards and secures the information that, easily adapts its execution technique in institutes, can bring institutions together so that they can choose more ideas in the to run the Lala project at a school, consider:

- Belong to the Lala community by accessing the necessary permissions to be able to run in the desired institution.

- The information is collected from educational institutes to conduct a study on the educational strategies they use.

- A sketch is made about the implementation of the database with the data collected from the institution.

- Have a master server so that you can control the users and the information that is collected.

- Make Client Server for those teachers easy to enter and observe the re-quested information.

- The codes implemented in this project are in the NodeJS language because it collects the data to structure the link to the database.

- It adapts to any operating system and, in turn, takes up enough space de-pending on the amount of information collected from an educational institute.

Table 1. Comparisons of Lala projects at different universities

Data **Comparative**	Viewing the program	Analysis of subject scores	Analysis of approved, disapproved topics	Results from previous analysis
Espol	See the student's academic history	Compare the student's grade with that of other students in bar chart format	Using colors, define the approved or failed theme with the score you earned.	Determines the difficulty of the subject and the level of difficulty of the university degree you chose to study
Catholic University of Chile	See the student's academic history	Track's student scores daily on the topic you are taking inline graphic format	Using a pastel chart with a respective color defines the approved or failed theme with the score obtained	alerts the student of the difficulty of the subjects and advises him how to improve his performance for the future
University of Cuenca	See the student's academic history	Compare the notes the student got in the first part with the end in bar chart format	Using colors, the approved or failed subject is defined with the score obtained.	Determines the difficulty of the subject for each elective year
Universidad Austral de Chile	See the student's academic history	The student's overall and partial performance is tracked in these subjects	Using colors, define the approved or failed theme with the score you earned.	Compared with the grades obtained by the student in the previous years with the current year, the degree of difficulty of the degree is determined.

FUZZY DECISION MAPS

Diffuse cognitive maps are a generalization of cognitive maps, both are directed graphs,

whose vertices represent concepts, and their edges represent the causal relationships between these concepts; these were presented by B. Kosko in 1986.

Fig 4. shows an example of a cognitive map on the left and a diffuse cognitive map on the right; the main difference is in the values assigned to the edges that mean the degree of relationship between the vertices. In Cognitive Maps, these values are –1, 1, which means a reverse or direct correlation between concepts. While Fuzzy Cognitive Maps takes values in the range [-1, 1], which includes a gradation between the relationships of

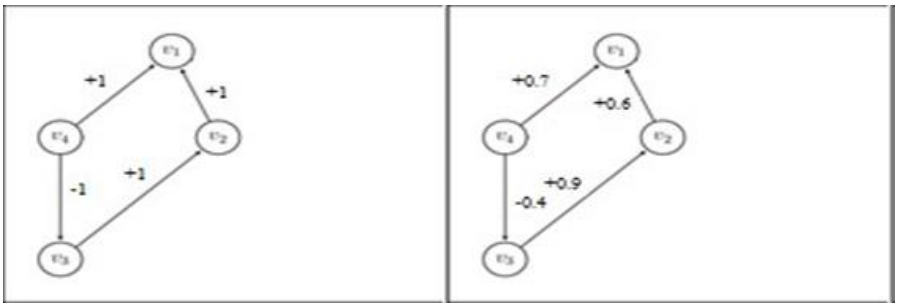

Fig. 2. Visualization of the watershed university advisory system

Fuzzy decision maps are based on the idea of Fuzzy Cognitive Maps' representation, to delete decision-makers considering self-independence and feedback. These allow you to model dependency and feedback between concepts easily.

MEHTODS-FORMULAS

A framework is proposed to prioritize the critical success factor in the review of Big Data Predictive Analytics in educational database systems, especially for Lala project implementations.

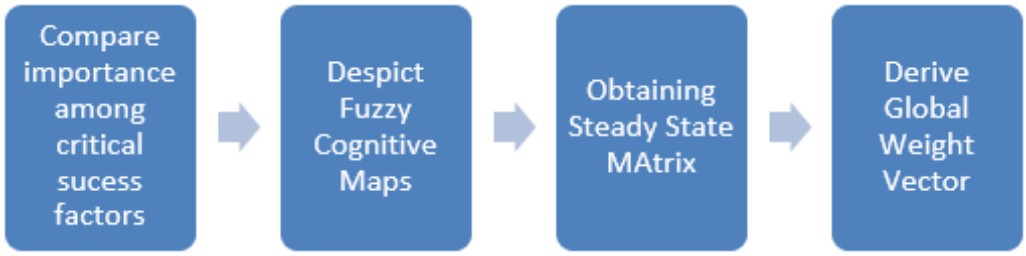

Fig. 3. Proposed framework

DERIVE LOCAL WEIGHTS

The importance between nodes will be derived from the local weight vector by com-paring the value focus.

Table 2. Scale for the relationship between the criteria.

Description	Numerical value
Egalitarian importance	1
Moderate importance	2
Strong importance	3
Very strong importance	4
Extreme importance	5

The Eigenvalue method is used to obtain local weights. This must be done by domain experts.

REPRESENT FUZZY COGNITIVE MAPS

A diffuse cognitive map (FCM) was developed to indicate the influence between the criteria by the expert. Causal interdependencies are modeled this way.

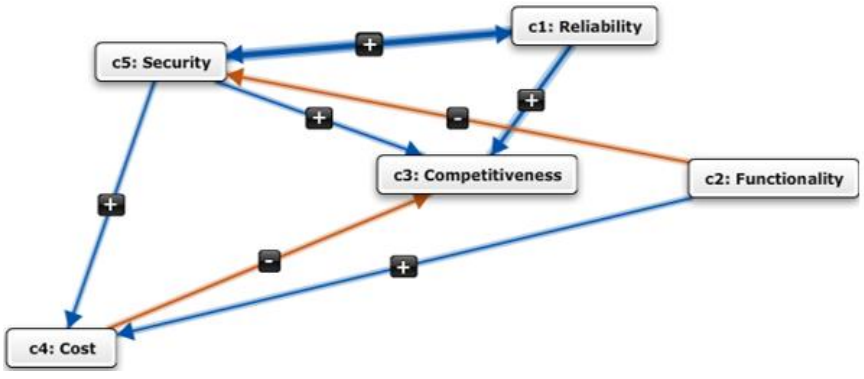

Fig. 4. The proposed framework of represent fuzzy cognitive maps

This stage consists of the formation of the FCM using experts. Nodes represent CSF interrelationships and vertices. It is a group of experts involving k experts, the final MCF adjacency matrix (E) is obtained as

$$W = \frac{W_1 + W_2 + \cdots + W_k}{k}$$

Relationships between concepts are described using the degree of influence.

MATERIALS

A review of Latin America's literature and experiences in other areas[17]identified the following critical success factors.

Table 3. Critical success factors identified.

Id	**Factor**	Description
F1	**Motivation**	The motivation of teachers to participate in the project
F2	**Training academic staff**	Training performed on predictive-applied analytics cents to the educational database
F3	**Data availability**	Data availability, as well as compliance with academic data standards and quality
F4	**Manager support**	Managers' support for training processes, organizational changes, and necessary systems
F5	**Proper architecture**	Definition of an appropriate architecture that facilitates the participation of students and teachers, as well as alert systems and the correct integration of data.

The development of the peer comparison judgment matrix between the CSF is developed as shown in the following table.

Table 4. Diffuse cognitive map modeling was performed using the Mental Modeler tool

	F1	F2	F3	F4	F5
F1	1	1	2	1/4	3

F2	1	1	1	1/3	1/2
F3	1/2	1	1	1/2	1
F4	4	3	2	1	2
F5	3	2	1	1/2	1

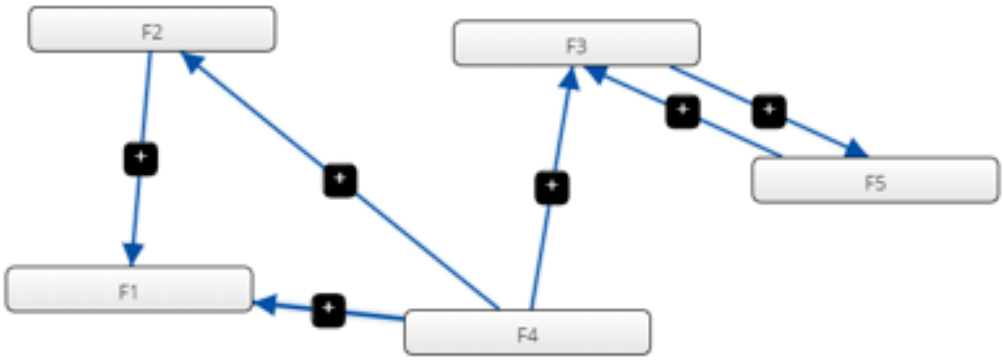

Fig. 5. Diffuse cognitive map

Based on expert knowledge, a diffuse cognitive map was developed. FCM shows the influence between the CSF. after the adjacency matrix Figure 8 was obtained.

	F1	F2	F3	F4	F5
F1					
F2	0.39				
F3					0.48
F4	0.53	0.48	0.39		
F5			0.42		

Fig. 6. Adjacency matrix

Subsequently, the calculations are performed, obtaining the following weight in the factors.

Table 5. Calculations made.

	Weight
F1	0.15
F2	0.11
F3	0.16
F4	0.37

CONCLUSIONS

Knowledge for future generations, new technologies must be used, which are evolving to such an extent that tools have been made that favor higher performance when following a university career; For this reason, it is necessary to integrate software such as the Lala project to improve academic performance and educational methods, currently the Lala project has the optimal references of certain institutions of higher education.

Consequently, a framework was presented to analyze critical factors using a fuzzy decision map. If a case study is carried out, the applicability of the proposal could be demonstrated.

According to the results, the three most important factors support managers, defin-ing an appropriate architecture and data availability are the three most essential fac-tors in this order. Future work will focus on expanding the framework for group deci-sion-making, including the consensus process

Acknowledgments. This work has been supported by the GIIAR research group and the Universidad Politécnica Salesiana.

REFERENCES

Peñaloza Báez, M.J.: Big data y analítica del aprendizaje en aplicaciones de salud y educación médica. Investig. en Educ. médica. 7, 61–66 (2018). https://doi.org/10.1016/j.riem.2017.11.003

Matas Terrón, A., Leiva Olivencia, J.J., Núñez Rojas, N.: El Big Data desde la perspectiva de una muestra de estudiantes de universidades del norte de Perú. Propósitos y Represent. 8, (2020). https://doi.org/10.20511/pyr2020.v8n1.376

Subbalakshmi, S., Prabhu, C.S.R.: Protagonist of Big Data and Predictive Analytics using data analytics. Proc. Int. Conf. Comput. Tech. Electron. Mech. Syst. CTEMS 2018. 276–279 (2018). https://doi.org/10.1109/CTEMS.2018.8769141

Prasad, S.T., Sangavi, S., Deepa, A., Sairabanu, F., Ragasudha, R.: Diabetic data analysis in big data with predictive method. 2017 Int. Conf. Algorithms, Methodol. Model. Appl. Emerg. Technol. ICAMMAET 2017. 2017-Janua, 1–4 (2017). https://doi.org/10.1109/ICAMMAET.2017.8186738

Reddy, A.R., Kumar, P.S.: Predictive big data analytics in healthcare. Proc. - 2016 2nd Int. Conf. Comput. Intell. Commun. Technol. CICT 2016. 623–626 (2016). https://doi.org/10.1109/CICT.2016.129

Zainab, K., Dhanda, N.: Big data and predictive analytics in various sectors. Proc. 2018 Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2018. 39–43 (2018). https://doi.org/10.1109/SYSMART.2018.8746929

Hilliger, I., Maldonado, J., Tsai, Y., Ortiz, M., Broos, T., Pesantez, P., Sheihing, E., Whitelock-wainright, A.: Building Capacity to Use Learning Analytics to. 1–133 (2018)

Kosko, B.: Fuzzy cognitive maps, (1986)

Luo, C., Song, X., Zheng, Y.: Algebraic dynamics of k-valued fuzzy cognitive maps and its stabilization. Knowledge-Based Syst. 209, 106424 (2020). https://doi.org/https://doi.org/10.1016/j.knosys.2020.106424

Papageorgiou, K., Singh, P.K., Papageorgiou, E.I., Chudasama, H., Bochtis, D., Stamoulis, G.: Participatory modelling for poverty alleviation using fuzzy cognitive maps and OWA learning aggregation. PLoS One. 15, (2020). https://doi.org/10.1371/journal.pone.0233984

Domínguez, D., Educación, F. De: Analítica del aprendizaje y. 1–14 (2016)

Leyva, M., Hechavarría, J., Batista, N., Alarcón, J., Gómez, O.: A framework for PEST analysis based on fuzzy decision maps. Espacios. 39, (2018)

Tzeng, G.H., Chen, W.H., Yu, R., Shih, M.L.: Fuzzy decision maps: A generalization of the DEMATEL methods. Soft Comput. 14, 1141–1150 (2010). https://doi.org/10.1007/s00500-009-0507-0

Elomda, B.M., Hefny, H.A., Hassan, H.A.: An extension of fuzzy decisi on maps for multi-criteria decision-making. Egypt. Informatics J. 14, 147–155 (2013). https://doi.org/10.1016/j.eij.2013.05.001

Soliman, T.H.A., Mohamed, R., Sewissy, A.A.: A hybrid Analytical Hierarchical Process and Deep Neural Networks approach for classifying breast cancer. Proc. 2016 11th Int. Conf. Comput. Eng. Syst. ICCES 2016. 212–219 (2017). https://doi.org/10.1109/ICCES.2016.7822002

Leyva-Vázquez, M., Pérez-Teruel, K., John, R.I.: A model for enterprise architecture scenario analysis based on fuzzy cognitive maps and OWA operators. CONIELECOMP 2014 - 24th Int. Conf. Electron. Commun. Comput. 243–247 (2014). https://doi.org/10.1109/CONIELECOMP.2014.6808598

Saleh, S.H.: Evaluación de la calidad de software empleando decisión Multicriterio basado en los números de SVN. (2018)

Strang, K.D.: Do the Critical Success Factors From Learning Analytics Predict Student Outcomes? J. Educ. Technol. Syst. 44, 273–299 (2016). https://doi.org/10.1177/0047239515615850

Gray, S.A., Gray, S., Cox, L.J., Henly-Shepard, S.: Mental Modeler: A fuzzy-logic cognitive mapping modeling tool for adaptive environmental management. Proc. Annu. Hawaii Int. Conf. Syst. Sci. 965–973 (2013). https://doi.org/10.1109/HICSS.2013.399