

Improve the Performance of Students in the Mathematics Learning through Bayesian Model

Luis Barba-Guamán¹⁺ and Priscila Valdiviezo-Díaz¹

¹ Universidad Técnica Particular de Loja, Ecuador

Abstract. This paper shows the use of data mining techniques and math tools software as a supplementary educational resource in the understanding of mathematics. The performance process of mathematics in the first and third level of Computer Science and Electronics has used these data mining techniques and math tools as part of a formation process. Mathematics is fundamental in the formation process of educating the future professionals. Students from experimental groups used together with the teacher the Wolfram Mathematica software. The Bayesian model showed the prediction of the approval rate of the students. Part of the experience in this research was also getting the perception from students through a survey. As a result, we determined the need of this math package as a supplementary educational resource, which supports the capacity of calculus, and the interpretation of the non-trivial problems.

Keywords: data mining; wolfram software; bayesian model; education

1. Introduction

There are researchers who analyze the different mathematical software tools in the engineering careers [1]. The mathematical software tools are used in higher education with a didactical focus [2], [3].

Other studies have been done in this area incorporating useful mathematical aspects in an educational data mining approach in order to improve student's retention [4] [5].

These researchers [4], apply a learning analytics approach to identify how the students' retention during the first-year of Mathematics could be improved. Another study, presented by [6] analyzes data collected through the interactions of students and teachers in a virtual learning environment [19] [20] [21] and it is collected through the results recorded in the Institutional Academic System of Evaluations. In this study, authors obtained a predictive model to forecast the performance rate of a new student in each course. Similarly [7], use data mining techniques to predict students' final grade based on the student's database. For this reason, some attributes were collected from the student's database for data classification, using the Decision Tree Method.

Data Mining (DM) is a new technique, which refers to the discovery of information from a large collection of data. This technique allows people to learn efficiently with historical data, discovering hidden patterns and using that knowledge for predicting future behavior of related areas [7]. One application of DM is in Educational Data Mining (EDM), which refers to the use of data mining for automatically extracting meaning from immense repositories of data about people's learning activities in academics. There are increasing research interests in using data mining techniques in education; for instance, the use of data mining tools to facilitate a better resource utilization in terms of student performance and identifying those students which need special attention in a course and taking appropriate action.

Based on this background, this article presents the analysis and the prediction of the approbation rate of Mathematics and Calculus' students of Computer Science and Electronics. This is based on data mining

⁺ Corresponding author. Tel.: + 5930984819880
E-mail address: lrbarba@utpl.edu.ec

techniques to improve the teaching of mathematics using Wolfram software.

1.1. Understanding of mathematics through technology-support

Currently, there is a lot of free and paid software to perform mathematical analysis and calculations in different areas of knowledge. Mathematics is a powerful and vital discipline for solving problems in real life. According to [8] some students are required to solve math problems as part of their studies. Most of the given problems are well structured; but there is an issue in these types of problems. They present all their elements to the learners. These problems require a limited number of rules and principles and have knowable solutions [9]. However, working with problems helps students to connect Mathematics to reality. This aspect makes learning more interesting for students. This mathematic learning process can be developed through individual and group interactions conducted in the classroom. The use of Wolfram software supports the idea of students solving problems and interpreting responses; thus, the students would acquire mathematical enhanced performance and they could have active participation and satisfactory learning outcomes.

The use of Wolfram software supports the idea where students can solve a problem as well as look for an answer to it. In this case, they acquire mathematical knowledge, thus, students can learn best by actively constructing their own knowledge. In this reason, the constructivist learning Mathematics can be produced through individual and group interactions within the classroom. This involves that the teacher needs to provide learning opportunities for students, such as: creating problems, organizing group work, sharing information and transferring knowledge.

These researchers [10] presented an experience about teaching Mathematics for engineers with the aid of Wolfram Alpha software. Students were able to use it in classes and during the written part of examination. The researchers [11], [12] used the software Wolfram Alpha, and Mathematica software as a complement to the theoretical teaching of Calculus and Algebra in core engineering courses. Another similar research is presented by [13] who studied the effects of the computer interventions on the mathematical achievement [14], attitudes towards search engines and Internet self-efficacy in different ways. In this experiment, the first group of teachers used activities provided by a web- based semantic search engine (i.e. Wolfram Alpha). The same content was provided to a second group of teachers through a computer algebra system (i.e. Mathematica). The third and final group of teachers followed the same content through linear web pages. In this study [15], the teachers' reactions about a statistics innovation project (called Wolfram Project), involves four steps of using: (posing asking the right question, mathematical form of the question, calculation, verification and interpretation of the answer). Thus, the learning takes place in a computer-based environment, it does not make the role of the teacher less important. The teacher's role is essential during all these steps to teach lessons and guide students toward the solutions.

2. Methodology

In this paper, we propose the use of data mining techniques to analyze the student performance in Mathematics and Calculus courses in the first and third semester of Computer Science and Electronics from the Universidad Técnica Particular de Loja. Each semester is divided in two bimesters (I Bim and II Bim) The UTPL has a competency-based teaching model. In this model, the student is the main focus of the educational process. The process is mediated by a teacher, educational resources and new technologies. The source of data to be used to carry out the analysis contains both historical and current data of these courses. In order to improve the teaching of Mathematics our research comprises two phases: 1) to know if there was an increase or decrease in the number of students admitted, and 2) to predict the passing rate in the mathematics courses in Computer Science and Electronics. We apply data mining techniques such as Naïve Bayes, based on historical data from previous applicants used as a dataset of 987 scores [16]. The results obtained from the research were supported by a survey used to get the perception level of students about mathematics classes.

3. Data Analysis

Figure 1 shows the number of admitted students vs the students who were not admitted in Mathematics in the fall semester 2015 and spring semester 2016.

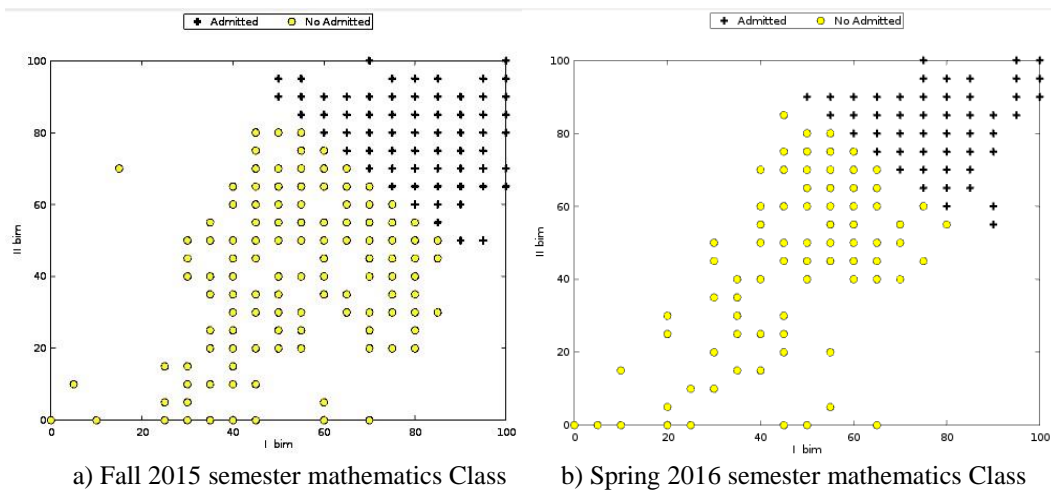


Fig. 1: Students admitted and students who were not admitted in Mathematics between 2015 and 2016. (Academic systems from Universidad Técnica Particular de Loja)

As shown in Figure 1, there is a higher percentage of non-admitted students (yellow) and a low percentage of admitted students (black) in the mathematics courses. The training set used for this model corresponds to 80% of the data and 20% to the test.

The metrics used to measure the model performance were: Precision and Recall. The results were:

- Precision: 97%, which is an acceptable rate due to this percentage of the students who passed the course. These results have been classified correctly.
- Recall: 98% of the students who will pass the courses will be included in the prediction.

Next, we presented the use of predictive model of the promotion students' rate. In the development of model, we used the same dataset mentioned above. In this study, the data mining techniques play an important role to uncovering information about the performance of students. Particularly, the predictive model is constructed using Naïve Bayes.

The experiment was conducted in Orange software [17] with a dataset of 987 instances. For training is considered 80% of this dataset and the rest 20% to the test. Attributes considered for this experiment are shown in Table 1.

Table 1: Attributes used in the model

Attribute	Description
Semester	This attribute indicates the semester of study
Career	Name of career
Component	Name of educational component
Credits	Credit number of component
I bim	Score of the first two months
II bim	Score of the second two months
Total	Final score
Promotion rate	The rate of students passing: value 1 if the student passes the course and value 0 if the student does not pass the course

Table 2: Matrix confusion

A	B	<-- Classified as
377	53	Disapprove (0)
42	513	Approved (1)

The data shows the passing rate of students in the course, according to their scores that they received during the semester. The variable to predict was the promotion rate, which has two values: passed (1), failed (0). Additionally, the confusion matrix obtained is shown in Table 2. As we can see 90% of the instances are classified correctly and 9,4% corresponds to the ones that were classified incorrectly. Based on this matrix, we proceed to calculate the precision and recall metrics to measure the model performance.

- Precision: 90% of the approval predictions are correct.

- Recall: 92% of the students who will pass the courses are included in the prediction.

Consequently, the results are acceptable considering the proposed target. This model allows to project a passing rate for upcoming semesters. The results of model we can see in Table 3 that the Bayes model presents a higher percentage in precision and recall of the passing rate of mathematics students. Thus, we can use this model to predict if a particular student would pass the course. On the contrary, the study conducted by [18], where both models have similar performance with the same size of dataset that it used in our research.

Table 3: Bayes model

Sample Size	Naïve Bayes	
	Precision	Recall
987	90%	92%

Based on the results, we selected two courses (i.e., Mathematics and Calculus) of a new semester from the first and third level of Computer Science and Electronics. These two courses were under the direction of two professors from different disciplines to get the perception of students about the mathematics course. The experiment involved 23 mathematics students and 17 calculus students.

To get the perceptions of students we applied a survey about the features of the subject:

- The academic subject is difficult to understand.
- It is interesting for most of the students
- It is theoretical and abstract.
- Most of the students consider this subject really important in the development of their careers.
- It is significant to have prior knowledge in the fundamentals of Mathematics.

Table 4 shows the results of this survey for the first group, which was developed through a traditional methodology.

Table 4: Details of the first group using traditional methodology.

	1.strongly agree	2. agree	3. fairly agree	4. disagree	5. strongly disagree
To consider the subject difficult.	8,70%	8,70%	43,48%	30,43%	8,70%
The academic subject seems interesting so far.	26,09%	56,52%	17,39%	0%	0%
The practice and the theoretical content relate to one another.	47,83%	30,43%	17,39%	4,35%	0%
This subject is important for professional training.	60,87%	21,74%	17,39%	0%	0%
My prior knowledge in Mathematics is sufficient for the subject.	39,13%	17,39%	30,43%	13,04%	0%
The contents are abstract.	4,35%	26,09%	43,48%	13,04%	13,04%
The practice activities are in Mathematics and Calculus.	39,13%	43,48%	17,39%	0%	0%

A Linkert-type scale was used to validate the results. It comprises the following assessments: (1) strongly agree, (2) agree, (3) fairly agree, (4) disagree and (5) strongly disagree. The result of the first group is shown in table 1. With a total of 23 students, 74% of the students are male and 26% female. From this group, 12 students were from 17 to 19 years old, 9 were from 20 to 22 years old and 2 were older than 22 years old. 21 students have studied the course for the first time and only 2 students for the second time. In this research, teachers presented the need of using this math package as a resource, which supports the capacity of calculus and the interpretation of the non-trivial problems.

Thus, a second course was developed based on a traditional methodology. However, they had to validate the answer of the given problem by using the mathematics software. The teachers and students from the experimental groups used the software

Wolfram Mathematica. Table 5 describes the data obtained by the group that used the mathematics software. All of them had the same survey. The participants were 17 students, 82% male and 18% female, 12

were from 17 to 19 years old, 5 from 20 to 22 years old. 95% of the students have studied this course for the first time and only 5% of them, for the second time.

Table 5: Details of the first group using traditional methodology.

	1.Strongly agree	2. Agree	3.Fairly agree	4. Disagree	5. Strongly disagree
To consider the subject difficult.	23,53%	29,41%	29,41%	17,65%	0%
The academic subject seems interesting so far.	17,65%	41,18%	41,18%	0%	0%
The practice and the theoretical content relate to one another.	0%	70,59%	17,65%	11,76%	0%
This subject is important for professional training.	64,71%	29,41%	0%	0%	5,88%
My prior knowledge in Mathematics is sufficient for the subject.	0%	41,18%	41,18%	17,65%	0%
The contents are abstract.	0%	64,71%	29,41%	5,88%	0%
The practice activities are in Mathematics and Calculus.	35,29%	47,06%	17,65%	0%	0%

It is clear that the percentages presented in Tables 4 and 5 do not differ much between them. However, there is a slight growth because the subject seems interesting, easier and the theoretical/abstract part is important in professional training.

4. Conclusion

In this paper, we presented the results of the analysis of one model of data mining as Naïve Bayes. A significant finding of this study is a statistical difference in global terms in Naïve Bayes. This is when the passing rate of students of is determined. Although, in the Bayesian model there were more students who passed than failed. But the number of students predicted to fail was considered high. The goal is to improve the learning process with the support of mathematical tools.

In addition, a survey was applied to know the perception of students about teaching Mathematics. Tables 3 and 4 show the results. They showed on average that students have considered that the academic subject seems interesting despite the abstract and theoretical topics in the engineering area. Students mention that the use of the mathematical tools in their formative process and professional training are significant, this can be perceived as an upgrade in the teaching-learning process in the future mathematical courses for engineering. A positive aspect of this study was to increase the interaction between the teacher and the student. The theoretical concepts were developed and analyzed graphically, which improved learning.

Teachers know how to work with this software; however, there is a need to create new teaching strategies for this system. This means that there is a lack of training in the planning and designing of materials for their classes, therefore, students can enhance interaction with the teacher.

Thus, we paid attention to these results and made a decision to further increase the number of students approved through the use of Wolfram Software as a support to the learning-teaching process.

Wolfram Mathematica had an impact on the student perception, because it is possible to do the direct calculus and solve a problem. This can be something positive in the current academic course. Also, this practice in the classroom allowed students to develop skills such as: programming, researching, and interpreting data.

The results of this study should be analyzed carefully, since they serve as an example of a particular population. Teachers participated voluntarily. Time was one of the limitations because the mathematical software has been used recently and its novelty may have some impact on the results.

Future studies may consider the use of new teaching methods designed for this system.

5. References

- [1] R. Barba-Guamán, P. Quezada-Sarmiento, & C. Calderon-Cordova. Using Wolfram software to improve reading comprehension in mathematics. 2016 International Conference on Sport Science and Social Science (ICSS 2016),

Yokohama, Japon. 2016. In press.

- [2] E. Ascheri, R. Pizarro, G. Astudillo, P. García, & M. Culla. Software educativo en línea para la enseñanza y el aprendizaje de temas de Cálculo Numérico. *Revista Digital: Matemática, Educación e Internet*, 14(2). 2014
- [3] P. Quezada-Sarmiento, L. Enciso & J. Garbajosa. Use of body knowledge and cloud computing tools to develop software projects based in innovation 2016. *IEEE Global Engineering Education Conference (EDUCON)*: 2016, 267-272. Abu Dhabi.
- [4] Y. Fridhan, B Loch & L. Walker. Improving retention in first-year mathematics using learning analytics: 2013, 278–282.
- [5] E. Vichez Quesada. Impacto de las Nuevas Tecnologías de la Información y la Comunicación para la Enseñanza de la Matemática en la Educación Superior. 2015.
- [6] P. Valdiviezo-Díaz, J. Cordero, R. Reátegui & J. Aguilar. A business intelligence model for online tutoring process. *Proceedings - Frontiers in Education Conference (FIE)*. 2015. <http://doi.org/10.1109/FIE.2015.7344385>
- [7] A. Ahmed & I. Elaraby. Data Mining: A prediction for student's Performance using Classification Method. *World Journal of Computer Application and Technology*: 2014, 43-47.
- [8] D.H. Jonassen. Toward a design theory of problem solving. *ETR&D* 2000, 48, 63–85.
- [9] P.K. Wood. Inquiring Systems and Problem Structure: Implications for Cognitive Development. *Human Development*, 1983, 26, 249–265.
- [10] P. Necesal & J. Pošpišil. Experience with teaching mathematics for engineers with the aid of Wolfram Alpha. In *Proceedings of the World Congress on Engineering and Computer Science, San Francisco, CA* (1): 2012, 271-274.
- [11] A. Conceição, J. Pereira, C. Silva, & C. Simão. Mathematica in the classroom: new tools for exploring precalculus and differential calculus. In *Proceedings of the 1st National Conference on Symbolic Computation in Education and Research*. 2012.
- [12] S. Martins. An approach to teach Calculus/Mathematical Analysis (for engineering students) using computers and active learning—its conception, development of materials and evaluation. 2013.
- [13] Ersoy, M. & Akbulut, Y. Cognitive and affective implications of persuasive technology use on mathematics instruction. *Computers & Education*, 75, 253-262. 2014.
- [14] J. Olabe, X. Basogain, M. Olabe, I. Maíz & C. Castaño. Resolviendo problemas de matemáticas y ciencias en el mundo real con una mente computacional. 2014.
- [15] C. Hommik, & T. Hõim. Teachers' Feedback to Secondary-Level Statistics Course Innovation: Computer-Based Education Pilot in Estonia. *Proceedings of the 14th European Conference on e-Learning: European Conference on e-Learning, United Kingdom*: 705–711. 2015.
- [16] L. Barba-Guamán, “Admitted and no admitted in mathematics class.”, *Mendeley Data*, v2. 2016. <http://dx.doi.org/10.17632/2nhvzt23fd.2>
- [17] Bioinformatics Laboratory, U. 2016. Orange Data Mining. [online] [Orange.biolab.si](http://orange.biolab.si). Available at: <http://orange.biolab.si> [Accessed 1 Sep. 2016].
- [18] Rudner, Lawrence. Accuracy of Bayes and Logistic Regression Subscale Probabilities for Educational and Certification Tests. *Practical Assessment, Research & Evaluation*, 21(8). 2016. Available online: <http://pareonline.net/getvn.asp?v=21&n=8>
- [19] D. Liu, P. Valdiviezo-Díaz, G. Riofrio, Yi. Sun & R. Barba. Integration of Virtual Labs into Science E-learning, *Procedia Computer Science*, Volume 75, 2015. Pages 95-102, ISSN 1877-0509, <http://dx.doi.org/10.1016/j.procs.2015.12.224>
- [20] D. Liu, L. Barba-Guamán, P. Valdiviezo-Díaz & G. Riofrio. Intelligent tutoring module for a 3Dgame-based science e-learning platform. *Inteligencia Artificial*. 2017, 20(60): 1-19. Doi: 10.4114/intartif.vol20iss60pp1-19. In press.
- [21] J. Aguilar, M. Sánchez, J. Cordero, P. Valdiviezo-Díaz, L. Barba-Guamán and L. Chamba-Eras. Learning analytics tasks as services in smart classrooms. *Universal Access in the Information Society*, pp. 9-10 ,2017.