International Journal of Engineering & Technology, 7 (4.32) (2018) 53-56

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

The Use of EDM on Predicting the Rate of Moroccan University Dropouts: Sharia university, Fez as A Case Study

MoulayHachemAlaouiHarouni[1*], El-Kaber Hachem[1], Cherif Ziti[1], MustaphaBassiri [2]

[1]Research Team EDP & Scientific Computing, Mathematics & Computer Department,
Faculty of Sciences, Moulay Ismail University, Meknes, Morocco
[2] Laboratory of Physical Chemistry of Materials,
Ben M'sik Faculty of Sciences, Hassan II University of Casablanca, Morocco
[*] Corresponding author: e-mail: harouni.alaoui@gmail.com

Abstract

The present article is going to highlight the concept of educational data mining (EDM) and discuss how it can be mainly used to predict Shariaa University dropouts; in other words, it may demonstrate how to enable academics to help students not to drop out their studies, the collected data were taken from our previous study of dropouts at Shariaa University, Fez-Morocco. Using the Bayes theorem that is stimulated from the statistical learning theory to resolve classification problems, we can determine the strengths and weaknesses of those students who are probably to dropout.

The reasons of conducting this second research on dropping out are to provide the support of continuous university self-evaluation and improvement through using educational data mining, and to help those students who will be expected to leave their university.

Keywords: *EDM, dropouts, Bayes theorem, classification, self-evaluation.*

1. Introduction

Building university-wide capacity for educational data-driven decision-making is a key requirement for supporting the Sharia University, Fes. Data mining encourages educators to figure out different problems that cause students to sop their studies[1]. For these reasons, the concept of University Autonomy comes as a new vision and an essential requirement for achieving better educational results, as well as more efficient and productive university activities[2]. Therefore, university autonomy has been used as the core of education system reform policies in many countries around the world. In the Sahriaa University, nevertheless, the concept of university autonomy will be applied only to discover reasons behind the phenomenon of dropping out university: is it because of ineffective choices? Or is it because financial problems? All these reasons are taken into consideration while we will be using educational data mining.

2. Educational Data Mining

The Educational Data Mining community website, www.educationaldatamining.org, defines educational data mining as follows : "Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in."[3].

Here, educational data mining (EDM) comes as the concept for collecting and organizing information that represents some aspects of Shariaa University; this information consists of any relevant idea about students, their parents, school, environment, teacher, etc. through these information and EDM, we can interpret the students' behaviors, such as the students who will drop out their studies, and may explore some solutions of any problem related to the university operations and activities. Hence, we insist on the fact that the use of EDM is not limited only to students' grades in exams, but also consists of a wide scope of data represented by various sources, coming from internal and external data to the university. For example, we utilize EDM in Shariaa University with the purpose of making appropriate decisions.

3. Data-Driven Decision-Making in University

Data-Driven Decision Making (DDDM) in university has been defined as a continuous cycle of identifying, collecting, combining, analyzing, interpreting and acting upon educational data from different sources in order to report, evaluate and improve the resources, the processes and the outcomes of university. Building university-wide capacity for education data-driven decision-making is a key element required for supporting university autonomy [4][5][6]. University autonomy is internationally discussed as a basic need for achieving better educational results in relation to students' learning processes, as well as sufficient university teaching. For this reason, Arcia et al. [7]base university autonomy on different controls and organizations which enable a particular university to have the power of decision-making over its teaching practices.

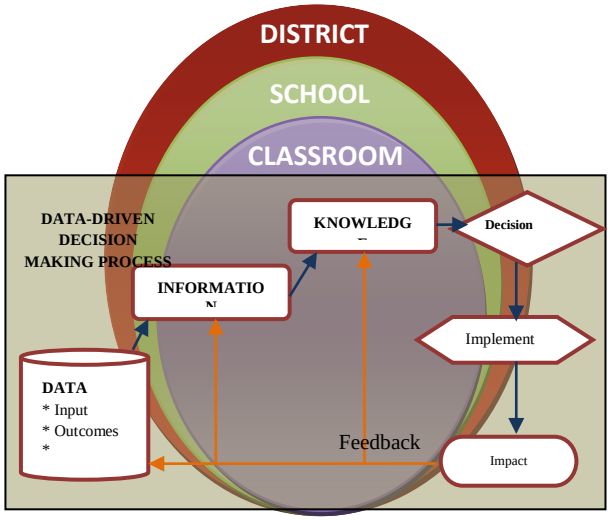

Fig.1: Data-driving decision-making process.

4. Educational Data Analytics Technologies

Data analytics has been defined as tools for analysing large group of different types of data, collected from various sources with the purpose of supporting and improving decision-making process. Data analytics is mature technology currently applied in real-life financial, business and health systems. The data analytic technology helps policymakers to have the power of decision and act up on educational systems. Decision taken from collected data has an impact at various levels of university activities and operations to satisfy the needs of each student so as not to make them drop out their studies. Most of our collected data are based on the following categories of educational data, which Lai and Schildkamp[8]have extended from Ikemoto and Marsh's [9] to input data, process data, context data and outcome data. Each category is presented with indicative examples of educational data:

Table 1: Categories of mining educational data

Input data	Student characteristics, such as demographics, prior academic performance, transfer records, native language. Teacher characteristics, such as teacher competences, academic qualifications or professional experience.
Data Process	Data generated during the teaching, learning and assessment processes, both within and beyond the physical classroom premises, such as lesson plans, methods of assessments, classroom management.
Context data	The Curriculum such as subject syllabus (including learning outcomes) and additional educational programs. School Human Resources, Infrastructure and Financial Plans, including educational and non-educational personnel, buildings, hardware/software, and expenditure. School Culture such as school climate, student / parent / teacher/ community relations.
Outcome data	Students' Achievements in classroom-based formative assessments, homework, standardized tests, (inter-) national exams. Students' Wellbeing, and Social and Emotional Development such as safety, support, respect for diversity and special needs. Graduate Data on employment after graduation or further academic studies.

Furthermore, data analytics, according to Johnson et al. [10], has been used in educational field as technologies that can support teaching and learning processes. Educational data analytics can be classified into three main types:

- ***Teaching Analytics*** is thought to be as methods that enable educators and administrators to analyze their instructional designs (ex. the lesson plan) in order to better reflect on them with the aim of improving learning conditions for their targeted individual learners or groups of learners. Typically, the analysis of educational designs is combined with insights from their implementation (for example through learning analytics).
- ***Learning Analytics*** is considered as tools that monitor the learning process to collect, measure, analyze and report on learners' educational data and the learning context so as to improve the learning. It is related to teaching analytics, which provides the means to analyze the learning context and the instructional design.
- ***Teaching and Learning Analytics*** joins both teaching analytics and learning analytics together to make teachers systematically reflect on their teaching design using evidence from the delivery to the students in the classroom.

To have all these types of educational data successfully work; we should use "multiple types of data, including input data, such as school expenditures or the demographics of the student population; process data, such as data on financial Table 1operations or the quality of instruction; outcome data, such as dropout rates or student test scores; and satisfaction data, such as opinions from teachers, students, parents, or the community"[9].

5. The Predictive Model by Exploring Data

In the field of education, data analysis is becoming more common to predict optimal results, artificial intelligence provides software solutions; among them is Weka[11].

1.1.Exploring the rate of Shariaa University dropout

This part describes the data we use in exploring the rate of Shariaa University dropout. We obtained the diffusion data from our previous study of the prediction of Shariaa university dropout, Fes-Morocco between 2012 and 2016, which has not been published yet.

This is a statistical study tracked approximately 3000 students from 2012-2016. The frequency of university dropouts, as shown in the following table, is about 30%.

Table2: Number of dropouts during two different periods

University drop-out	Cases
Indicated before the end of the first year	909
Indicated after the first year	315

We constructed a large matrix of about 3000 students describing by 86 variables: a mixture of binary functions, numeric functions, categorical characteristics, and a large amount of missing values in the data. We prepared these data to feed it to our platform not for easily predicting the rate of dropouts from their first year but also for finding out the suitable solution. We managed the complexity of data by organizing functionalities into different groups. The first one is DMG, a demographic group, which includes characteristics such as student age, marital status, family support, income - an indicator of the student's socio-economic status, gender, etc. The second group is PSH that stands for Previous Study History. We also have information on the status of modules, SM, which represents both exam and catch-up one for each semester (S1, S2, S3, S4, S5 and S6). We also have had all the information concerning the types and the grades of the baccalaureate and the region where it was taken (IB).

All in all, the following calendar in Fig.5 illustrates the rate of Shariaa University dropouts from 2012 to 2016 with different characteristics such as DMG, PSH, SM, and IB:

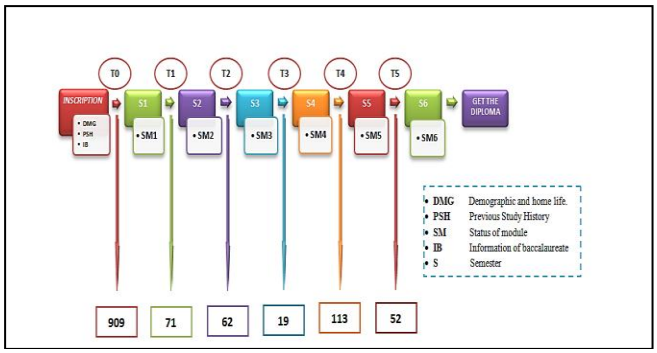

Fig.2: Calendar of the student characteristics

Table 3: Number of University dropouts into different periods

Period	Number of dropping out
T0	909
T1	71
T2	62
T3	19
T4	113
T5	52

With the help of educational data mining, these collected data were used to feed the machine learning, for example our website, to predict automatically their characteristics of dropouts and the rate of dropouts. That is to say, in each semester students will be asked to fill in different questions, including their information about their educational background, baccalaureate, their family status, interests of choosing major, and so on.

The use of educational data mining shows different variables that cause to dropout from university, but some variables can change over time (family situation, economy ...), which lead to easily influence the results during studying at university. It can also be seen that in T4 and T5, the number of student decreases remarkably, this big fall can have an explanation that some students who got their DEUG diploma could either have applied for other jobs or have enrolled in other universities or institutions.

1.2.The predictive model

In a learning environment envisioned, the student will take the time to answer some questions, something essential to determine his/her appropriate learning style, even give them a personal profile where their strengths and weaknesses in a specific discipline/ course will be taken into account. The style agent of ASTEMOI system has a predictive model that tries to achieve this goal [11][12]; for this fact an essential step must be followed:

- Exploitation of Alumni data to form the core of the model using Naive Bayes classifier to determine students that most likely to dropout.
- Estimation of the strengths and weaknesses of new students based on the analysis of old data.
- Student redirection to a suitable profile that responds best to their inspirations and provides additional links course proposed by tutor agent[13]. These courses are presented in several forms (video, audio, text) depending on the learner's style (Reflection, Reasoning, Sensory, and Progression).
- Comparison between the final results of new students and those predicted by the model, allowing updating and improving the predictive model.

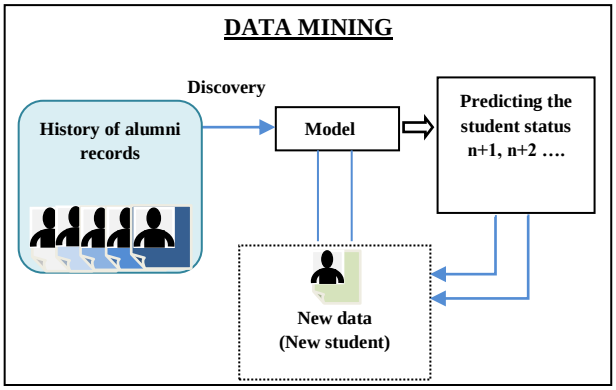

Fig.3: Predictive model in the system ASTEMOI

1.3.Students dropping out in an appropriate profile

To estimate the appropriate learner profile for students who are likely to quit and to determine their strengths and weaknesses, several classification algorithms are developed.

Our work will be limited to using the naive fuck that leads to more realistic results compared to other algorithms (see the previous study [11].

We have as inputs information of students from Shariaa Faculty of Fez including: the type of baccalaureate, the notes of the B.A subjects, diploma score, regions, The classification for each criterion leads to separating the students into groups. A representative profile model will be assigned to each group.

For example our experience will be based on the following attributes:

- Baccalaureate major: Literature, Sciences, Islamic Studies.
- Age : [19,24] ; [24,30] ; > 30
- Mark : <10 ; [10,12] ; >12

However, we just want to know the students who found difficulty in each module relying on their "yes" or "no" answers; and to assign appropriate profiles to different groups of students.

1.4.Bayes theorem

Bayes' theorem is used in statistical inference to update or modify the estimation of a probability or any parameter from observations and laws of probability of these observations [15].

$$\left(\frac{w_i}{x}\right) = \frac{p(w_i).p\left(\frac{x}{w_i}\right)}{p(x)} \quad (2)$$

Such as

$\left(\frac{w_i}{x}\right)$: Post

$p(w_i)$: Prior

$p\left(\frac{x}{w_i}\right)$: Likelihood

$p(x)$: Evidence

1.5.Naive Bayes classifier

The Naïve Bayes classifier is a type of simple probabilistic Bayesian classification based on Bayes' theorem with a strong independence of the hypotheses. It implements a naive Bayesian classifier[16].

We express our problem as a probability, in which the variables $x_1, x_2 \ldots x_n$ are considered independent, so:

$$p(x_1, x_2 \ldots x_n) = p(x_1) * p(x_2) * \ldots p(x_n) \quad (3)$$

And the naive Bayes becomes [17]:

$$p\left(\frac{w_i}{x_1, x_2 \ldots x_n}\right) = \frac{p(w_i) * p\left(\frac{x_1}{w_i}\right) * \ldots * p\left(\frac{x_n}{w_i}\right)}{p(x_1) * p(x_2) * \ldots * p(x_n)} \quad (4)$$

Whether the students have difficulty w_1= (yes) or not w_2= (no). We have a new object (student) to classify: x_1, x_2, x_3 the attributes respectively (type bac, age, mark). We want to know if this object (student) belongs to $w_1$or w_2. The student belongs to the class that maximizes this probability. We are not obliged to calculate the evidence because it is a constant and we want to find the maximum of these probabilities.

For example: Student = {Literature, >28, <10}, belongs to which class?

$$p\left(\frac{yes}{\text{Literature}, >28, <10}\right) \propto p(yes) * p\left(\frac{\text{Literature}}{yes}\right) * p\left(\frac{>28}{yes}\right) * p\left(\frac{<10}{yes}\right)$$

And of the same for the class «no »

$$p\left(\frac{no}{\text{Literature}, >28, <10}\right) \propto p(no) * p\left(\frac{\text{Literature}}{no}\right) * p\left(\frac{>28}{no}\right) * p\left(\frac{<10}{no}\right)$$

So the student belongs to the class that has more probability than the other.

6. Conclusion

The present article elaborates a new way of finding out not only the rate of students who will leave university from the very beginning of each semester but also the causes of dropping-out. To sum up, our study relies on educational data mining to survive the learning process of each student in particular and educational systems in general. All collected data will be useful for policymakers to make an appropriate decision for helping those who may stop studying at university.

This study opens an opportunity for new researchers to look for the main use of EDM first to motivate students while they are studying at university, and to satisfy their desires through finding out their interests.

References

[1] Kalina, Y., S.Ryan and J.d. Baker. 2009. The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1(1).

[2] Kemal, G. 2011. University Autonomy and Academic Freedom: A Historical Perspective. International Higher Education (63), pp. 13-14, DOI: https://doi.org/10.6017/ihe.2011.63.8549

[3] Witten, I.H. and E. Frank. 1999. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Francisco, CA.

[4] Mandinach, E. 2012. A Perfect Time for Data Use: Using Data driven Decision Making to Inform Practice. Educational Psychologist, 47(2), pp. 71-85.

[5] Marsh, J.A. and C.C. Farrell. 2014. How Leaders Can Support Teachers With Data-Driven Decision Making A Framework For Understanding Capacity Building. Educational Management Administration & Leadership, pp. 1-21.

[6] Schildkamp, K. and W. Kuiper. 2010. Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. Teaching and Teacher Education (26), pp.482-496.

[7] Arcia, G., K. Macdonald, H. Patrinos, and E. Porta. 2011. School autonomy and accountability. System Assessment and Benchmarking For Education Results (SABER).

[8] Lai, M. K. and K. Schildkamp. 2013. Data-based Decision Making: An Overview. In K. Schildkamp, M.K. Lai & L. Earl (Eds.). Data-based decision making in education: Challenges and opportunities. Dordrecht: Springer.

[9] Ikemoto, G. S. and J. A. Marsh. 2007. Cutting through the Data-Driven Mantra: Different Conceptions of Data-Driven Decision Making. RAND Corporation.

[10] Johnson, L., R. Smith, H. Willis, A. Levine and K. Haywood. 2011. The 2011 Horizon Report. Austin, Texas: The New Media Consortium.

[11] El Emary, I. and A. Brzozowska (Eds.). 2017. Shaping the Future of ICT: Trends in Information Technology, Communications Engineering, and Management. Boca Raton: CRC Press. Chapter 6: Artificial Intelligence in E-Learning eBook ISBN : 9781498781190.

[12] AlaouiHarouni, H., E. Hachem and C. Ziti. 2018. Modern Probabilistic: Model for Massive Data in E-Learning. J FundamAppl Sci., 10(4), 456-459.

[13] AlaouiHarouni, H., E. Hachem and C. Ziti. 2016. Data Mining For the Service of Intelligent Tutoring System. Int.J.Mult.disc.scie., 1(1): 61 -65.

[14] Lee, P.M. 2012. Chapter 1. Bayesian Statistics. Wiley. ISBN 978-1-1183-3257-3.

[15] Efron, B. June 2013. Bayes' Theorem in the 21st Century, Science, vol. 340, no 6137,june 2013, p. 1177-1178.

[16] Dimitoglou, G., A.A. James and M. Carol. 2012. Comparison of the C4.5 and a Naïve Bayes Classifier for the Prediction of Lung Cancer Survivability Journal of Computing, 4(8).DOI: http://dx.doi.org/10.4314/jfas.v10i4s.192

[17] Layachi, B. 2007. Data mining for scientists bishop's university. Retrieved 02 April 2018 from http://aqualonne.free.fr/Teaching/csc/DM.pdf