International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.6 (2015), pp.9-16
http://dx.doi.org/10.14257/ijmue.2015.10.6.02

Prediction Approach for Improving Students Orientation in University: Case of Sidi Mohammed Ben Abdelah University (SMBAU)

Satauri Imane[1] and El Beqqali Omar[2]

Department of computer science,
Faculty of sciences
Fez, Morocco
[1]imane.satouri@gmail.com, [2]omarelbeqqali@gmail.com

Abstract

At school there are always first and second degree of education, higher education and training. The question is: the transition from one level to the other should they exist regulatory mechanisms and a prediction crowned by tool for decision support in school and university?

National decision-making approach of orientation to accompany the student in the development of his identity and his personal path and vocational is required.

We can resume our work by implementing a predictive model making integrated educational project facility through the impact of certain variables of different types of academic success. This is the case of the university sidi mohamed ben abdellah (SMBAU).

Keywords: *Prediction; university orientation; datamining; decision tree; CART algorithm*

1. Introduction

In the first year of faculty, the percentage of exam failure is always higher, a large number of students in first year of faculty give up or fail (statistics of SMBAU 2012). Assuredly, beginning studies at faculty is a delicate period which student must deal with new and complex situations.

The orientation of a new student after taking his bachelor depends on diverse contextual factors (socio-economic, socio-cultural, socio-demographic) and personal (aptitudes cognitive, personality) and other environmental variables. In addition to the effect of these classical history, today we are interested by possible impact of process 'transactional' (process of evaluation and adjustment developed in response to adversity) on academic results [1].

Statistical and artificial intelligence methods are used to provide counselors and universities with a tool for monitoring student pathways. We target to improve policy choices mainly we are working on real data from the University data base: APOGEE

Indeed, considering the improvement of orientation and selection process of the new graduates at the entrance to the university to avoid stalling or failure. We are thinking that if some failure situations may be inevitable (bad adaptation to the proposed training *etc.*), much of them can be prevented by early detection that would result in cropping and careful orientation and customized.

The rest of the paper is structured as follows: Section 2 introduces state of the art. Section 3 discusses the data collection (source and methodology), choice of the method and proposed architecture. Experiment results are discussed in Section 4. Finally Section 5 summarizes this paper.

ISSN: 1975-0080 IJMUE

2. State of the Art

One of the biggest current challenges of higher education is to predict the academic course of each student and alumnus. Institutions would like, for instance, to know who are the students wishing to enroll in a specific curriculum and who are those who will be in need of help to graduate? In addition, traditional issues such as managing students' enrollments and delays in graduation have always been subject to discussion in that institutions of higher education are seeking better solutions for such issues.

Several explanatory factors of academic failure in the first year have been highlighted. For Coulon [2], "entry and success in higher education come from learning, from acculturation and those who cannot join, fail." It is in the first year that the affiliation process is taking place, allowing students to fit into academia, even if they are often confronted with "big integration difficulties" [3-4]).

Works, investigating the determining factors of success at the University, have all agreed on the importance of prior education and social origin. Thus, the type of baccalaureate degree, the age at which it was obtained, gender and social background have all a decisive influence on access and success in higher education [5], holders of science baccalaureates (BAC "S") are generally more successful than others, however, students who have already failed in high school and obtained later their baccalaureate degrees, be it technological or professional, are at greater risk of failure than others [6].

Taking into account other factors, more contextual in nature, may also explain the failure at the end of first year as explained by Duru-Bellat [7].

Also the impact of the university context in which the student develops must be carefully considered: the university campus, teaching practices, curricula, support devices are all factors to be taken into account to explain differences in success at the end of the first year [8-9].

Since 2007, new features have emerged in the context of the establishment of the university Success Plan. But research has shown that methodological support or teaching devices, most often optional, rarely touch those who need it most: those who attend these methodological and pedagogical meetings are indeed those whose chances of success are already high and are already partly familiar with the codes of academic work [10].

In order to promote the success of a large number of students at the end of the first year, various policies have been implemented in different countries (pre-selection on the basis of aptitude; modulation of enrollment fees based on students' performance, initiating an optional course map allowing greater freedom of choice for students). All these measures have as a common goal maximizing the degree of matching between the motivations and abilities of students and characteristics of academic institutions [11].

In conclusion, most of the studies cited above have denied the importance of the repercussions of social support on academic success, they have been little studied. Scientifically speaking, our approach is to provide in addition to social support a set of explanatory indicators of academic orientation. The identification of the weight of each set of factors in the success can then provide useful information on the success chances (or failure risk) of students based on their personal characteristics in a new academic year. More broadly, this type of analysis is targeted to better help and provide support, especially at the first year, for the sake of clarifying the referral procedures of students in the different sectors.

3. Methodology

3.1. Data Collection

Our study was based on 1200 students from two different schools. The inclusion criteria were: having obtained his bachelor in 2012 and be registered for the first time at a university in "2012-2013". We have chose students in 1st year because they represent the

most homogeneous population in terms of statistical analysis. But it is also possible to interview all new entrants to an institution, and secondly because our research focuses on predicting pre-university orientation.

The sample including 1200 students (76.4% of men and 23.6% for women). Students in our sample were enrolled in four courses of two university institutions: 452 in the Faculty of Human sciences (French), 584 in the Faculty of Human sciences (English), 84 from multidisciplinary faculty of TAZA (SMA: Science Applied Mathematics), 80 from multidisciplinary faculty of TAZA (SMI: Mathematics and Computer Science).

3.2. Procedure

The SMBAU has an information system (APOGEE), which manages the careers of students from different sites (9 schools) and identifies shifts in institutions. Our work is divided into two main parts:

- A series of questionnaires (demographic, socioeconomic and socio-cultural), will allow to analyze the school-university transition with more qualitative data than databases of the university, the collection took place at the time of registration administrative between July and October 2012.
- Retrieving results of each student in the sample from APOGEE and converting these into categories (July 2013)

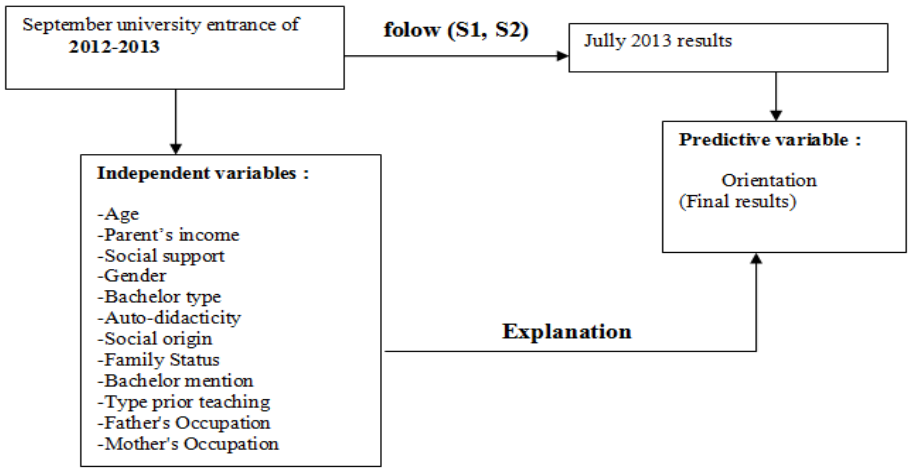


Figure. 1. Study Conducted on 1200 Students of SMBAU

3.3. Choice of the Method

Our approach is based on a predictive modeling to predict the results of the first academic year. This type of modeling is one of the two main branches of data modeling (the other being descriptive modeling) She objects to identify strong links between variables (columns) of the data table. Such a link may lead, for example, by equation (approximate) connecting a variable y, called the "dependent variable" to a group of n other variables {xi}, called "variables" or "predictors".

In our study, we are interested in analyzing the effect of the variables mentioned above on the choice of pre-university orientation. To identify these effects, our model was applied to four courses from different institutions and takes the following form:

Success in the first year
=
F (socio-demographic characteristics, socioeconomic characteristics, educational background, sociological factors, courses)

Among the forms of representation of predictive models, we find:

- Classes which are groups of individuals with common properties.
- Associations, that is to say rules to know what condition triggers which accordingly.
- Decision trees.

In our approach we have chosen to work with decision trees to build our decision model based on the variables mentioned above, algorithms for creating the most famous trees are ID3 and C4.5 decision, they can discover model classification of the data.

We were forced to make a comparative study of induction methods in terms (reliability of the model, number of rules, execution time) and go through the following steps:

- Import given in the software.
- Select the variable to predict (the "class" attribute) and descriptors.
- Select the induction of decision trees method.
- Start learning and view the tree.
- Use four re-sampling techniques to assess the quality of the induced model namely: Cross validation, leave one out validation, bootstrap validation, partition “learning and test “.

The theoretical error rate of a supervised induction method is defined as the probability of misclassifying an individual in the population. However, it is impossible to calculate directly because it is not possible to reach all of the population. We have produced estimated based on the selected starting sample.

Table 1. Comparative Study of Decision Trees

Classifier	Error rate	Number of rules	Execution time	cross-validation	leave one out	bootstrap
CART Learning	0,1808	7	31 ms	-	-	
CART test	0,19	-	0ms	0,1876(1170 ms)	0,1850(31215ms)	0,1824(1966)ms
C4.5 Learning	0,1982	5	31ms			
C4.5 test	0,3394	-	0ms	0,2462(1014ms)	0,2321(30576ms)	0,2414(2558ms)
Rnd Tree Learning	0,1013	10	16 ms			
Rnd Tree Test	0,3818	-	0ms	0,3026(827ms)	0 ,2653(29266ms)	0,2748(1935ms)
Decision List Learning	0,2291	4	50ms			
Decision List Test	0,2485		0ms	0,1897(795ms)	0,2066(26348ms)	0,2117(1701ms)
CS-CRT	**Results of CART**	-	-	-	-	-

- Error learning is usually too optimistic,
- Error test penalizes the estimation of the random error seen cutting (learning data / test data).

Table 2. Results of Sampling Techniques

Error	CART	Discard	C4.5	Discard	RndTree	Discard	Decision List	Discard	CS-CRT	Discard
Resubstitution	0,1808	-0,0092	0,1982	-0,1412	0,1013	-0,2805	0,2291	-0,0194	Res CART	
True « test »	0,19	*	0,3394	*	0,3818	*	0,2485	*		
Cross validation	0,1876	-0,0024	0,2462	-0,0932	0,3026	-0,0792	0,1897	-0,0588		
Leave one out	0,1850	-0,005	0,2321	-0,1073	0 ,2653	-0,1165	0,2066	-0,0419		
bootstrap	0,1824	-0,0076	0,2414	-0,098	0,2748	-0,107	0,2117	-0,0368		

Cross-validation is usually the best in terms of variability around the real value of the error and differences between the techniques of re-sampling remains minimal in our case.

We chose to apply the CART supervised learning method not only because it is the best in terms of accuracy but also because of the nature of the explanatory variables and the dependent variable:

- Explanatory variables: a mixture of quantitative and qualitative variables.
- Variable to explain: qualitative variable (coding quantitative variable "result" in a pass / fail or abundant).

3.4. Proposed Architecture

In our study we chose a combination of a data warehouse and predictive data mining model, this coupling process allows a multidimensional analysis of student data and analyzing the successes, failures and policy changes. The additional operations such as aggregation and data mining may be performed.

The proposed architecture combines data from two heterogeneous data sources: student information (predictive indicators in our case) and the oracle database final results. The data thus obtained are filtered and homogenized in the data warehouse and then loaded and used for the purpose of analysis and prediction.

The Data Warehouse is used to provide reliable statistics on the final results of students to be able to detect several levels of granularity and then use machine learning methods to treat these statistics more thorough.

At first we used data partitioning methods for detecting students of similar behavior groups in order to offer them suitable courses. In a second step, the automatic classification of students will understand what makes the difference between those who succeed and those who fail. Finally, the regression will predict the final results and indirectly policy choices.

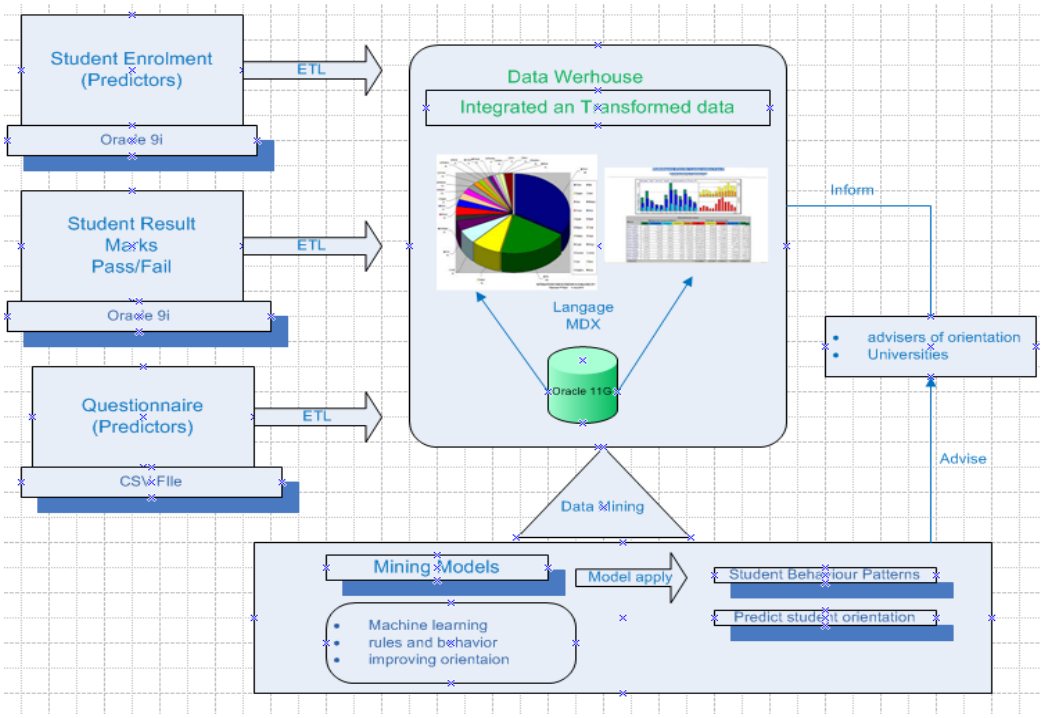


Figure. 2. Proposed Architecture

4. Results

The independent variables listed above (Figure. 1) were selected as input variables in creating the diagram CART and predictor "Orientation" as output variable. (See Figure 3).

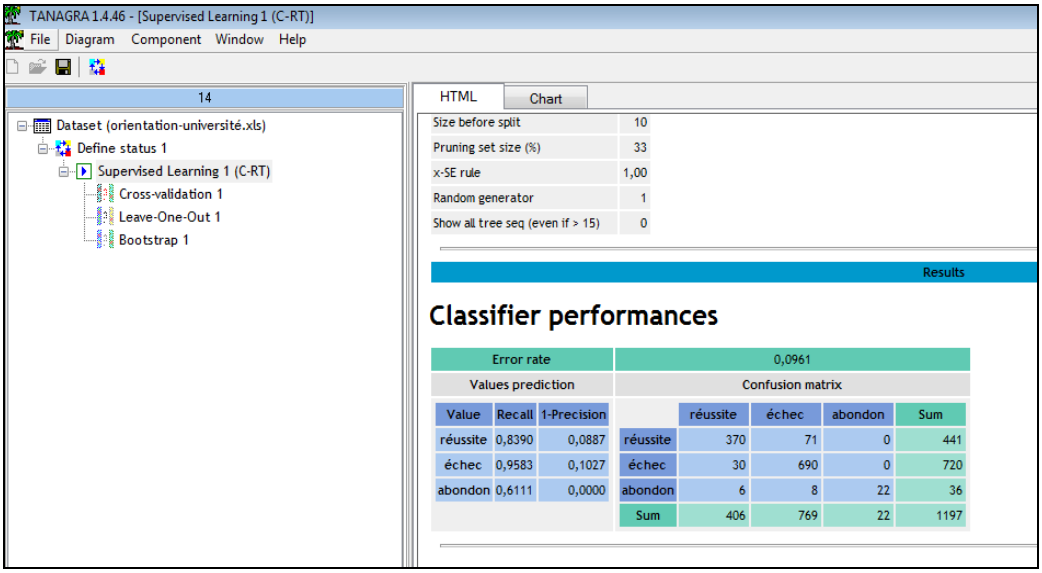
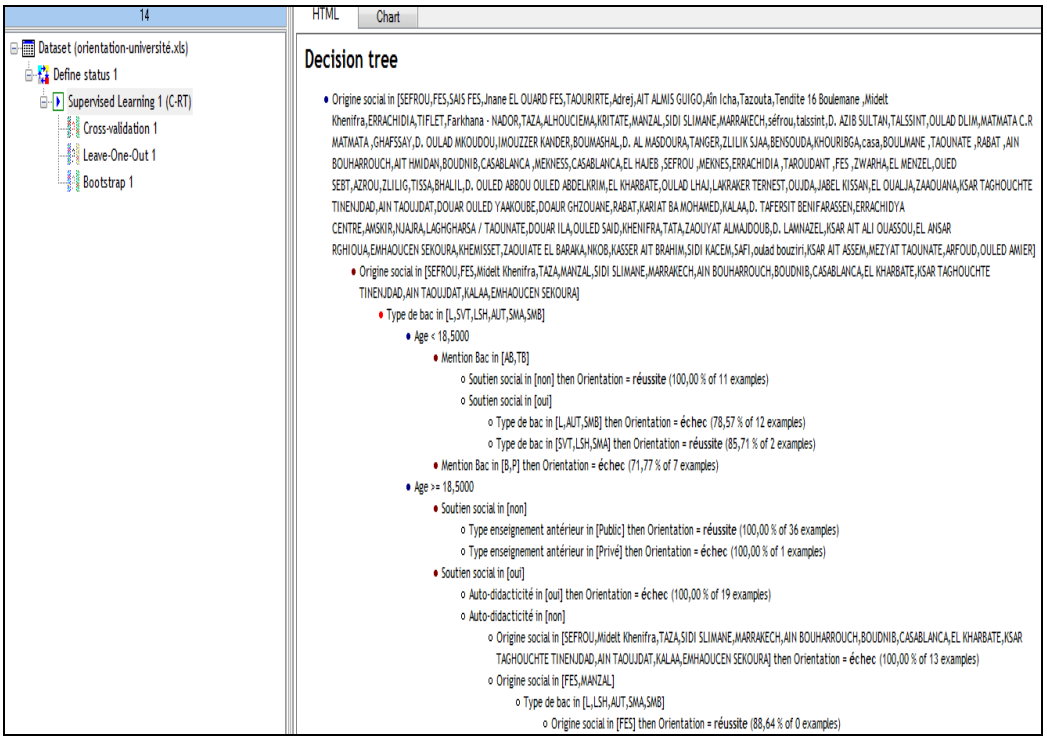


Figure. 3. Confusion Matrix on the Training Sample



Decision Tree:

Figure. 4. Resulting Predictive Model of Learning.

In order to ensure the effectiveness of learning and the relevance of the product model, we conducted a data slicing in learning test because we don't have access to all data of the university.

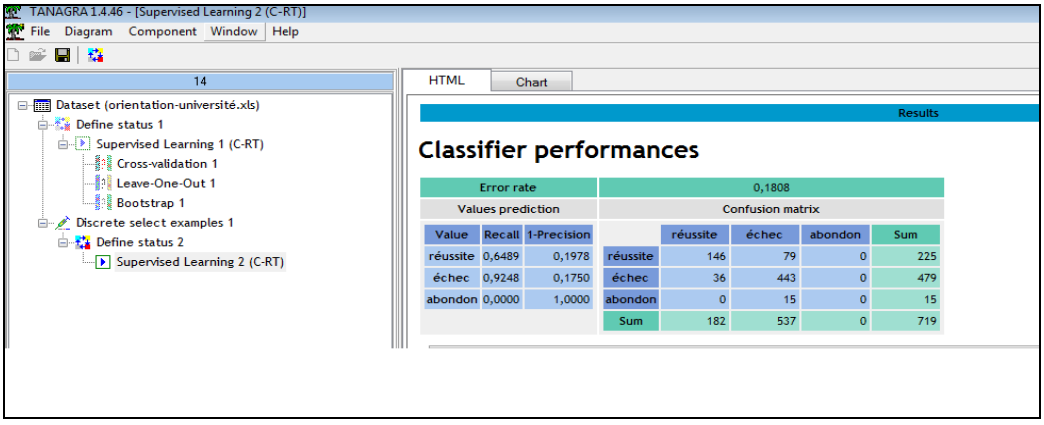


Figure. 5. Confusion Matrix After Cutting the Sample

The re-substitution error is 18.08%. Judging this figure, when we classify a new observation from the model provided by C-RT, 18.08% is the chance of achieving a misallocation.

To verify the model obtained, we will measure the performance of the model on the remaining observations (test data).

The "real" rate of error (the test sample) of our prediction model is 19%, there is still a gap compared to the previously evaluated re-substitution error (18.08%).

5. Discussion

The resulting model strongly meets reality relative to rates of successes and failures. However this model does not include all the independent variables introduced at the outset. One might therefore consider possible improvements using other supervised learning methods that could introduce other explanatory variables.

In against part, we achieved significant error rates that challenge the adoption of this model and push us to test other predictive methods to find the best predictive model of policy choices. We discussed how to use data mining to improve student orientation. According to us, the information incorporated into the data warehouse is the historical data of previous students and the features associated with the current and future potential students. We use this information to build the model of machine learning of student results. The use of these results helps us to extract rules and behaviours to improve student orientation.

We are also working to improve the system. For example, we are investigating whether there is a pattern in each individual's reasoning in the decision to continue, and how we should implement an intervention programme to upgrade student orientation.

6. Conclusion

Our study aimed to describe the characteristics differentiating career choices and paths of first year of a sample from SMBAU (demographics, personality, social support, socioeconomic characteristics) and among them, those who predict significantly final results and subsequently the choice of orientation.

We strive to improve these results through the implementation of other predictive methods of supervised learning and the integration of other explanatory variables.

References

[1] Z.K. Ü nlüa, N. Ö ztürk, R. Demir and E. Benli, "Turkish elementary school students' performance on integrated science process skills", 3rd World Conference on Educational Sciences, vol. 15, **(2011)**, pp. 3469–3475

[2] A.Coulon, Le Métier d'étudiant. L'entrée dans la vie universitaire," Paris Presses universitaires de France**, (2005)**, pp. 219.

[3] N.BEAUPERE, L'abandon des études supérieures, Rapport réalisé pour l'Observatoire de la Vie Etudiante, la documentation française, **(2007)**, pp. 162.

[4] F.DUBET, Dimensions et figures de l'expérience étudiante dans l'université de masse, revue française de sociologie, vol. 35, no.4, **(1994)**, pp. 511-532.

[5] N. BEAUPERE and G. BOUDESSEUL, Sortir sans diplôme de l'Université. Comprendre les parcours d'étudiants "décrocheurs", La Documentation Française, coll. « Etudes & recherches », **(2009)**, pp. 221.

[6] D. PROUTEAU, Parcours et réussite en licence des inscrits en L1 en 2004, Note d'information de la DPD, no. 23, **(2009)**, pp. 1-6.

[7] M. DURU-BELLAT, Des tentatives de prédiction aux écueils de la prévention en matière d'échec scolaire en première année d'université, Savoir Education Formation, no. 3, **(1995)**, pp 399-416,

[8] M. ROMAINVILLE, L'échec dans l'université de masse, Paris L'Harmattan, **(2000)**, pp. 28.

[9] C.MICHAUT, L'efficacité des dispositifs d'aide aux étudiants dans les universités, Recherche et Formation, no. 43, **(2002)**, pp 101-113.

[10] R. SHANKLAND, Pédagogies nouvelles et compétences psychosociales, Paris L'Harmattan, **(2009)**, pp. 217.

[11] P. ADNOT and J. DUPONT, Autonomie budgétaire et financière des universités et nouveau système d'allocation des moyens (SYMPA), 9 janvier **(2012)**.

