ITALIAN JOURNAL OF SOCIOLOGY OF EDUCATION

Editor-in-Chief: Silvio Scanagatta | ISSN 2035-4983

Predicting Students' Dropout at University Using Artificial Neural Networks

Anna Siri[*]

Author information:

[*] School of Medicine and UNESCO Chair in "Anthropology of Health", University of Genoa, Italy.

Contact author's email address

[*]anna.siri@unige.it

Article first published online

June 2015

HOW TO CITE

Siri, D. (2015). Predicting Students' Dropout at University Using Artificial Neural Networks. *Italian Journal of Sociology of Education*, *7*(2), 225-247. Retrieved from http://journals.padovauniversitypress.it/ijse/content/predicting-students%E2%80%99-dropout-university-using-artificial-neural-networks

PADOVA UNIVERSITY PRESS

Predicting Students' Dropout at University Using Artificial Neural Networks

Anna Siri[*]

Abstract: This study is part of an ongoing project investigating the first stage of the process of student transition to university. This paper aims to contribute to the continuing debates on the possibilities how to reduce the student failure and improve educational processes with the help of data mining techniques, in particular of the artificial neural networks. The population consists of 810 students enrolled for the first time in a health care professions degree course at the University of Genoa in the academic year 2008-09. The research is based on the analysis of data and information originating from primary sources: administrative data related to the careers of students; statistical data collected during the research through an ad hoc survey; data derived from telephone interviews with students who had not completed the enrolment in the subsequent years. The neural network correctly predicted 84 percent of the cases of group 1,81 percent of the cases of group 2, and 76 percent of the cases belonging to group of dropouts. The application of the artificial neural network model can offer a valid tool to design educational interventions to deliver to those who score high in the level of risk.

Keywords: Artificial Neural Network, Prediction, Student, Dropout

[**]School of Medicine and UNESCO Chair in "Anthropology of Health", University of Genoa, Italy. E-mail: anna.siri@unige.it

Introduction

Decreasing dropout in education is central for employments, productivity and growth. Low levels and low completion rates create a skills bottleneck in the economic sectors and inhibit innovation, productivity, and competitiveness (European Commission/EACEA/Eurydice, 2013).

Student dropout phenomena affect at every level the educational systems of any countries, including the socio-economically most developed ones.

An early recognition and a deepening of the reasons of this phenomenon represent prerequisites for any initiative aimed at reducing the factors affecting the decrease of the rates of the dropout at all levels. Therefore, forecasting students' dropout has become a crucial stated goal.

Recently, predictive modelling programs, like the Artificial Neural Networks, allow educational organisations to construct powerful statistical models that predict the list of students who are going to leave their study. This information can help building an early warning structure, which permits educators to know who is at risk and what is the best intervention program for each specific situation.

This paper aims to contribute to the current debate on how to predict the risk of dropout at university using the Artificial Neural Networks.

A look at the major theoretical models of student departure from higher education

University dropout is a multifaceted phenomenon with significant complex consequences that requires a profound political attention. Over the last fifty years, several theoretical and empirical research studies of student retention have been developed.

Psychological models of educational persistence emphasize the impact of individual abilities and dispositions upon student withdrawal, and try to differentiate stayers and leavers in terms of attributes of personality (Marks, 1967, Rossmann and Kirk, 1970; Ethington, 1990).

The environmental theories of student withdrawal emphasize the impact of social, economic and organizational services on the students behaviour within institutions (Kamens (1971), Pincus (1980), Iwai and Churchill (1982), Stampen and Cabrera (1986, 1988), and Braxton and Brier (1989). These researchers include for example students' social status and race (social), economic hardships (economic), and faculty-student ratios (organizational).

Over the years also other theories and models have been, several of them try to describe the complex phenomenon of student failure as well as to suggest interventions aiming to reduce student dropout. One of the most dominant theories is Tinto's Student Integration Model that emphasizes the importance of academic and social integration of students in the prediction of student retention (Tinto, 1975, 1987, 1993, 2010).

Another rather influential theory has been the Bean's Student Attrition Model that is based on process models of organizational turnover (March & Simon, 1958) and models of attitude-behavior interactions (Bentler & Speckart, 1979, 1981).

All these models highlight that a successful match between students and institutions is particularly important to restrain the student dropout (Cabrera et al. 1992, Hede &Wikander, 1990).

Heublein and colleagues (Heublein, 2003, 2010) recently refined a model inspired to Tinto's theory of student dropout and they constructed it in a European university context.This model includes both pre-university and within-university determinants. The authors point particular attention to externals factors including financial situation, living conditions, counselling services, and future plan.

Researchers have identified numerous variables related to student dropout as well as significant best practices to encourage undergraduate retention (Larsen, 2013; Siri, 2014).

The educational persistence is a product of a complex set of interactions among personal (e.g. background, motivation and study approaches), institutional (e.g. objectives, content, teaching, institutional climate, tutoring, and counselling) and external factors (e.g. financial situation, housing, work and leisure time).

Identifying what the determinants are and how they interact offer information that the university staff can use to help better students for university success.

This paper aims to contribute in an atypical way to the continuing debates on the possibilities how to improve educational processes with the help of data mining techniques, in particular of the Artificial Neural Networks.

Artificial Neural Networks

Artificial Neural Networks (ANNs) are mathematical models that try to imitate the behaviour of the human brain (Kriesel, 2007). ANNs involve nonlinear relationships among different datasets that cannot always be fully identified using conventional analyses. ANNs manage several types of input data to produce a clinically relevant output, for example, the probability of a certain pathology or classification of biomedical objects (Amato, 2013).

ANNs are a potent tool to simulate several non-linear systems. They have been applied to various problems of significant complexity in many fields, such us engineering, agriculture (Rodríguez Galdón et al., 2010), education (Nike et al., 2004; Siri, 2014), medicinal chemistry (Pandini, 2013), diagnostic systems (Liu & Jiang, 2013; Irfan Khan et al., 2013), and pharmaceutical research (Wesolowski & Suchacz, 2012; Patel et al., 2012).

Similarly, to the biological nervous system, ANNs are made up of neurons and each neuron consists of a 'summation block' and an 'activating block'.

The most common network structure is the multilayer perceptron (MLP) where the neurons are grouped into layers (Zhang & Gupta, 2000).

From this perspective, the artificial neuron is similar to regression models. In fact, a bias weight ('constant' in statistical terminology) is used to optimize the final prediction. However, regression models have an important weakness that consists in the inability to reveal intricate relationships in given datasets.

The first and last layers are respectively named 'Input' and 'output', because they represent inputs and outputs of the whole network. The 'Hidden layers' are the remaining layers that are situated inside the network, between the input and the output layers. Usually, an MLP neural network contains an input layer, one or more hidden layers, and an output layer.

The neural network structure that offers the best result for a specific problem needs to be experimentally decided (Suzuki, 2011). The size and the characteristics of the training dataset along with the number of iterations are the other factors affecting the generalisation capabilities of a neural network. Generally, the network size influences its complexity and 'learning time' but, most importantly, it affects the generalisation capabilities of the network. The power of an ANN depends on how well it generalises from the training data.

ANNs are best suited for cases where the particular nature of the connection between the variable inputs and outputs is not clear. ANNs are helpful when the link between several different variables needs an accurate mathematical model that still has not been developed. ANNs have the added value of being able to extract patterns and trends from datasets that are too complex to build with conventional computers. Another advantage of ANNs is their ability to include uncertainty in a given dataset. Furthermore, ANNs make no assumptions regarding the statistical nature of the data and can integrate nominal and ordinal data.

Research questions

This study asked two research questions:
1. How accurately do pre-entry students' characteristics predict the risk of dropout?
2. Which characteristics weigh most in predicting the risk of dropout?

Method

The identification of an appropriate topology (i.e., the number of hidden layers and the number of nodes per hidden layer) is one of the most crucial aspects of the application of neural networks. In this section, participants, data collection and analysis of this study were described in a detailed way under four subsections.

The identification of the objective of the prediction

The objective of this research is the construction of a forecasting model of the university students' career development in terms of success, irregular path and abandonment.

The construction of the data files on which to implement the neural network learning process

The research is based on the analysis of data and information originating from primary sources (relative to the universe of reference):

- Administrative data related to the careers of students enrolled in one of the degree courses on health care professions at the University of Genoa, collected by the appropriate offices.
- Statistical data collected during the research through an ad hoc survey, which covered all students enrolled in the degree courses on health care professions starting from the academic year 2008-09 (the data were recorded by proposing to the students a paper questionnaire within two months after the start of classes).
- Data derived from telephone interviews (made during the period July - September 2011) with students (enrolled for the first time in the first year in the academic year 2008-09) who had not completed the enrolment in the subsequent years.

In particular, with regard to the information collected directly during the research:

- The data from administrative sources of the University of Genoa have been used to plot the profiles of the three main reference categories of the analysis (regular students, irregular students or students who had failed to get a degree within a prescribed time, abandonments, including the changing of courses) with respect to some variables such as gender, age at registration, the type of secondary school diploma and the grade obtained, the curricular results at the date of October 24, 2011 (36 months after the beginning of the first year).
- The data obtained from administrative sources associated with those obtained by the research on students enrolled in the first year have been used to analyze the relationships between the phenomena of delay/abandonment and variables such as the channels activated for the choice of the university, the reasons for the choice of the course of study, the events occurred during the period between the graduation at the secondary school and the enrolment at the university, the attitude with which the student approaches the academic world and some characteristics of the family of origin.

In addition to the primary sources, the research used a series of secondary ones, represented mainly by studies conducted at national and international level by the major institutions of reference (ISTAT, MIUR, CNVSU, Banca d'Italia, Eurostat, OCSE, and UNESCO).

In the analysis phase, the data sets from the different sources have been crossed and joined together, in order to optimize all available information.

Data collection

The population consists of 810 students enrolled for the first time in a health care professions degree course activated by the University of Genoa in the a.y. 2008-09 (30.5 percent males, 69.5 percent females, mean age 26.35 with a minimum of 19 years and a maximum of 61 years; standard deviation = ± 6.29). 93 percent of the population has Italian citizenship, only a citizen of the European Union is enrolled in a course in the academic year of survey, while 56 students (6.9 percent) have their citizenships in countries outside the UE. Most of the students reside in the Liguria region (78 percent).

The secondary school from which comes the highest percentage of students is the high school on sciences (35.3 percent), followed by the technical institute (15.3 percent), the high school of humanities (12.2 percent) and the classical lyceum (9.3 percent).

61.9 percent of students attend the degree program in Genoa; in Imperia and Spezia, where as many as three courses are activated (Nursing; Physiotherapy; Technical Radiology, Imaging and Radiotherapy), respectively 9.5 percent and 8.5 percent of students play all academic activities both in the classroom and of professionalizing training. Two courses (Nursing and Physiotherapy) are also activated in Chiavari and Pietra Ligure. Only students enrolled in the Nursing Course can follow in Savona the entirely degree course.

Considering the variable student/student worker, 53.4 percent declared to be full-time students, 30.5 percent declared to carry out some work from time to time and 16.2 percent declared to be student-workers.

The construction of the archive of data on which to enable the neural network learning process

The database consists of 49 variables related to personal data and data on the educational and academic careers of 810 students enrolled in the first year of one of the courses of study on health care professions in the a. y. 2008-09. This research considered the significant aspects in determining dropouts, delays and changes of course. It also took into account and deepened other important variables, such as the school curricula prior to the enrolment at the university, the outcomes of the trials of access to the university courses on health care professions and the reasons for the choice of the degree course.

The input and output variables selection

The main input variables considered in this study are shown in the following Table 1. In order to highlight any inconvenience caused by attending a course of study in a place far away from the place of residence/domicile, an index has been created, called ”index of logistic discomfort”. A value = 0 indicates the absence of discomfort, as the place of residence coincides with the municipality of the training institution, while a value =1 means the existence of a moderate discomfort, as the municipality of residence is located less than 30 km from the training place. Finally, a value = 2 indicates the maximum discomfort (municipality of residence at more than 30 km from the training place). On the selected data, a correlation analysis has been carried out in order to identify any dependencies between the different variables considered in order to purify the database from redundant information. The analysis performed showed that there are no strong correlations between the indicators identified, except for the father and mother's place of birth with respect to citizenship. In consideration of what has emerged, all data were used in the experiment.

Table 1. Input variables

Input variables	**Description**	**Scale**
Laurea Degree Program	Course of study in which the student is enrolled	nominal
Gender	Gender composition (M/F)	nominal
Age	Age bracket	ordinal

Input variables	**Description**	**Scale**
Index of logistic discomfort	Coefficient based on the distance from the residence address to the university location adress	ordinal
Citizenship	Belonging legally to a country (Italy, Europe Union, other countries outside UE)	nominal
Grade of high school diplomas (based 100)	Grade of high school diplomas (based 100)	interval
Educational qualification	The secondary school from which students come	nominal
Location in Liguria region	Location in Liguria region where students attend the degree program	nominal
Degree programs preference	Enrolling for the admission test, candidates must provide, in order of preference (max 3), the degree programs that would like to enrol in.	ordinal
Grants and scholarships	Situation of student grants and scholarships reserved for students enrolled on university	nominal
Students previously enrolled at university	Student who was previously enrolled in an academic program at the same university or at other university (yes/no)	nominal
University name of previous enrollment	University in which student was previously enrolled in.	nominal
Previous healthcare degree program	Students previously enrolled in another healthcare degree program (yes/no)	nominal
Admission scores	Students level of performance in the admission tests	interval
Correct answers on culture and logical reasoning (percentage)	Percentage of the correct answers on the questions covered the culture and logical reasoning subject included in the admission test	ratio
Incorrect answers on culture and logical reasoning (percentage)	Percentage of the incorrect answers on the questions covered the culture and logical reasoning subject included in the admission test	ratio
Left unanswered questions on culture and logical reasoning (percentage)	Percentage of the left unanswered questions on the culture and logical reasoning subject included in the admission test	ratio

Input variables	**Description**	**Scale**
Correct answers on biology (percentage)	Percentage of the correct answers on the questions covered the biology subject included in the admission test	ratio
Incorrect answers on biology (percentage)	Percentage of the incorrect answers on the questions covered the biology subject included in the admission test	ratio
Left unanswered questions on biology (percentage)	Percentage of the left unanswered questions on the chemistry subject included in the admission test	ratio
Correct answers on chemistry (percentage)	Percentage of the correct answers on the questions covered the chemistry subject included in the admission test	ratio
Incorrect answers on chemistry (percentage)	Percentage of the incorrect answers on the questions covered the chemistry subject included in the admission test	ratio
Left unanswered questions on chemistry (percent)	Percentage of the left unanswered questions on the chemistry subject included in the admission test	ratio
Correct answers on physics and mathematics (percentage)	Percentage of the correct answers on the questions covered the physics and mathematics subject included in the admission test	ratio
Incorrect answers on physics and mathematics (percentage)	Percentage of the incorrect answers on the questions covered the physics and mathematics subject included in the admission test	ratio
Left unanswered questions on physics and mathematics (percentage)	Percentage of the left unanswered questions on the physics and mathematics subject included in the admission test	ratio
Clinical training performance	First year clinical training examination passed (yes/no)	nominal
Transition school to university	Student perception of the difficulties in transitioning from High School to University (yes/no)	nominal
School preparation	Student opinion about school preparation for the University	ordinal
Working situation	Working situation (full-time workers, workers from time to time; workers)	nominal
Students' motivation	Students' motivation to study healthcare professions	nominal
Entry program orientation	Student opinion about university entry program orientation	nominal

Input variables	Description	Scale
Parents in healthcare	Having parents who work in healthcare (yes/no)	nominal
Parents attitude	Parents' attitude towards studies and career choice	nominal
Factors potential dropout	Factors affecting the probability of students' dropout choice (students opinions)	nominal
Worker in the healthcare field	Already worker in the field of Healthcare system	nominal
Internal Locus	Internal locus of control	ordinal
External Locus	External locus of control	ordinal
Father age	Father's age bracket	ordinal
Mother age	Mother's age bracket	ordinal
Father's place of birth	Father's place of birth	nominal
Mother's place of birth	Mother's place of birth	nominal
Dominant language spoken	Dominant language spoken at home	nominal
Fathers' educational level	Fathers' educational level	ordinal
Mothers' educational level	Mothers' educational level	ordinal
Fathers' occupational status	Fathers' occupational status	nominal
Mothers' occupational status	Mothers' occupational status	nominal
Computer	To have a computer (yes/no)	nominal
Internet	Possibility of using internet sources (yes/no)	nominal

Source: own elaboration

The output variable

For the purposes of the research, three distinct students' categories have been identified, based on academic achievements. The identification of the three groups was based on the ministerial definition of 'regular' students (namely students enrolled since a number of years not exceeding the normal duration of the course) and, consequently, of irregular students. In the specific case of healthcare professions it was also checked (in October 2011 namely 36 months after the start), the number of credits earned by students about to graduate in November 2011 and March 2012. They had achieved more than 70 percent of the credits expected. In detail, the groups are composed as follows:

Regular

Category of students regularly enrolled in subsequent years, who never changed course and by October 2011 (or 36 months after the starting of the training process) have obtained more than 70 percent of the CFU required by the course of study, equal to 165 (i.e. 180 CFU curtailed by the credits related to the elective educational activities, including those relating to the graduation thesis, altogether equal to 15, for a total of 375 students).

Irregular (latecomer)

Category which includes the students who, in addition to being regularly enrolled in the years following the first one, by October 2011 got less than 70 percent of the CFU required by the course of study, equal to 165 (i.e. 180 CFU curtailed by the credits related to the elective educational activities, including those relating to the graduation thesis, altogether equal to 15, for a total of 283 students).

Dropout

Category to which belong the students, enrolled for the first time in the a.y. 2008/2009 in degree courses on health care professions, who by April 2010 (i.e. by the end of the reference a.y.) formalized the termination of their studies or did not renew their registration within two years. In this category were also included the changes of courses and consequently the students, enrolled for the first time in the a.y. 2008-09, who formalized a change of the course of study, for a total of 14 students. Overall, there have been 138 abandonments, equal to 17 percent of the 810 students enrolled in the a.y. 2008/2009 in the health care professions degree courses, and 14 changes of courses, equal to 1.7 percent.

The last two groups include the students who are in difficulties and consequently those who substantiate the so-called phenomenon of the 'university dispersion', while the first group is the benchmark for the 'academic success'.

The learning process and the choice of the architecture with respect to its significant parameters

The networks used in the experimentation are supervised feed forward neural networks, trained to solve a given problem repeatedly subjecting the network to a set of examples, called training set, consisting of pairs (V, T), where V is a possible input and T represents the corresponding correct output. For every V, the network produces an output O. The error (T - O) is used to change the behavior of the network, varying the weights of the connections between the nodes, called synaptic weights, and to minimize future errors. The networks, with architecture 49-X-1 (where X is the number of hidden neurons), have been instructed by means of the learning algorithm known as backward propagation. The software used in this research is STATISTICA (vers. 10). The objective of the first phase of testing has been to verify the learning ability of the network, by calculating the standard deviation (mean square error or MSE) as a measure of the error. The network was provided with 49 input data. The learning algorithm used, Back Propagation, modifies the synaptic weights in order to minimize the global error made by the network, evaluated by means of the mean square error (MSE) between the network output and correct results.

After fixing the number of neurons in the input (49) and output (1) layers, several networks have been tested, varying the number of neurons of the hidden layer, the parameters a (learning rate) and *b* (momentum), the error tolerance, the methodology for the weights updating and the starting maximum value of the weights (Interconnection Weight Matrix: IWM).

Very good performances, in any case in line with the first evidences, have been obtained by using a feature of the system that autonomously changes the parameters of the network (structure, coefficients, number of hidden layers, etc.) in subsequent processing iterations. Starting from a simple architecture the software modifies the parameters depending on the results achieved in terms of MSE calculated on the training set. The network was trained on 80 percent of the population taken into consideration; 20 percent of the population has been used as cross-validation group: this group is essential to the neural network to avoid that it stores the result instead of finding the right weight for each variable and thus generalizing the result obtained. In order to determine the optimal sequence of variables in predicting the expected result among those used in the input, the analysis of the network sensitivity was performed. This analysis provides the respective importance of input variables in determining the output, with the zero value representing a variable that has no validity in predicting and the 1 value representing a variable that does have the highest efficacy in predicting.

The results of this first phase of experimentation, in terms of measurement of the standard deviation of the actual output compared to the expected one both in the training process and in the cross validation test, seem to demonstrate a high level of rapid learning.

The network considered to be more predictive is the MLP 49-34-1, constructed with an input layer formed by 49 neurons, a hidden layer formed by 34 neurons and an output layer to represent the curricular outcome (1= regular student; 2= irregular student; 3=student at risk of abandonment).

The generalization of the output for the prediction.

In the second phase of the testing, the training of network has been iterated 5,000 epochs (i.e.all records under examination have been processed by the network 5,000 times), training the network with average values for both the initial weights (IWM), the learning rate and the momentum (respectively 0.5 and 0.7) and has been detected the epoch in which the sum of the error on the training set and the test set resulted smaller. This is defined freezing epoch, or the epoch in which the processing is terminated and the weights are frozen. In accordance with the model used at this point, the network has reached the optimal degree of training in relation to the ability to 'understand' the phenomenon in analysis.

Results and discussion

The network has taken into consideration 552 cases, selecting those that had a number of variables sufficient for a successful processing by neural networks. The missing values were not replaced by the usual elaborations that are adopted in these cases (e.g. replacement by the mean), as they relate to subjective information not adaptable to any other value.

The training was performed on 442 cases, randomly selected (80 percent of cases) and the testing on the remaining 20 percent or 110 cases. In the sample considered, 274 are regular students while 198 are irregular students (or 'fuori corso', namely students who have failed to get a degree within the prescribed time). The remaining 80 are students who have dropped out presenting an official communication to the university or have not confirmed their enrolment for the following year.

The neural network correctly predicted 68 percent of the cases of regular students, 66 percent of the irregular ones and 75 percent of the students who dropped out, considering reliable the predictions having a delta with respect to the expected values lower than 0.5.

Considering a delta higher than 0.6, the network correctly predicted 84 percent of the cases of group 1, 81 percent of the cases of group 2, and 76 percent of the cases belonging to group 3.

A particular attention has been devoted in this work to the analysis of group 3 (dropouts), giving rise to very interesting results.

It was decided to investigate more deeply some cases that the network apparently did not predict correctly (Table 2).

Table 2. Cases that the network apparently did not predict correctly (GROUP 3=ABANDONMENTS)

ID	TARGET	ANN MLP 49-34-1	Difference between output and neural network prediction	Absolute value of the difference between output and neural network prediction (Delta)
504	3	2.572336	- 0.43	0.43
487	3	2.508620	- 0.49	0.49
499	3	2.477173	- 0.52	0.52
416	3	2.391733	- 0.61	0.61
388	3	1.930442	- 1.07	1.07
399	3	1.854196	- 1.15	1.15
397	3	1.842426	- 1.16	1.16
394	3	1.768959	- 1.23	1.23
386	3	1.747465	- 1.25	1.25
395	3	1.718680	- 1.28	1.28
389	3	1.666289	- 1.33	1.33
482	3	1.649344	- 1.35	1.35
383	3	1.590455	- 1.41	1.41
505	3	1.577393	- 1.42	1.42
398	3	1.571801	- 1.43	1.43

ID	TARGET	ANN MLP 49-34-1	Difference between output and neural network prediction	Absolute value of the difference between output and neural network prediction (Delta)
381	3	1.570592	- 1.43	1.43
498	3	1.543178	- 1.46	1.46
798	3	1.493186	- 1.51	1.51
380	3	1.373303	- 1.63	1.63
401	3	1.356025	- 1.64	1.64
400	3	1.252400	- 1.75	1.75
402	3	1.237824	- 1.76	1.76

Source: own elaboration

It is a matter of a total of 22 cases. In detail: 5 are closer to the output 1 (regular), 13 are closer to the output 2 (irregular), 4 cases (ID = 504, 487, 499 and 416) stand apart with a value of delta placed between 0.43 e 0.61.

The access to the university administrative data set has allowed verifying the updated curricular position of the students on which this close examination has been defined.

1° subgroup

As regards eight students (ID = 380, 383, 389, 402, 416, 482, 504, 798), the network had predicted good academic performances. Six students of this subgroup changed their course of study in the subsequent a.y., gaining access to very selective courses such as Medicine and Physiotherapy and now they are actually having good results.

As regards the other two students (ID 383 and 482), it should be noted first of all that they were assigned to a seat that was more than 30 km away from the place of residence (maximum logistic discomfort). One requested the transfer to the same course of study activated at the University of Bologna, while the other changed his course of study choosing a path in the law studies, in which, however, he did not have a good performance.

Concluding the analysis of the first subgroup, it can be said that the outcome of the curricula of at least six (ID=380, 389, 402, 416, 504, 798) of the eight students tested was correctly predicted by the network.

2° subgroup

The network also predicted as regular or slightly behind six students (ID= 388, 397, 401, 498, 499, 505), who actually abandoned. They are students belonging to higher age groups (students so called "nontraditional", which need to balance work, family and studying. This private condition is difficult to reconcile with the enrollment in courses of study on the health professions, which are very demanding for the students, being not allowed the part-time enrollment, given the mandatory nature of the frequency organization, alternating between theoretical and practical training.

3° subgroup

As for the remaining eight cases a close examination showed that it is a case of young students with scholarships (n=4) and consequently a low income of their families and logistic difficulty (n=2), due to the allocation to a seat far away from the place of residence. They have not signed up to the subsequent academic years, but for some of them it could be a stop out, such as for example the case ID = 400. This hypothesis is suggested by the fact that some of them declared, in the telephone survey started in 2011, to have encountered many difficulties in taking the exams of the first semester integrated courses.

Concluding, with reference to the deepening of investigation made on the third group, relative to abandonments (80 cases), a revision of the calculation of the percentage of predictivity of the network equal to 75 percent with a delta equal to 0.5 (see above) can be proposed. Therefore it is correct to consider among the well-predicted cases also the six students (ID=380, 389, 402, 416, 504, 798), who, changing their course of study, obtained good performances in the new course in which they are now enrolled. Therefore, the network, as regards the third subgroup (abandonments), was able to correctly identify the 92 percent of the cases, thus proving to be a valuable and early investigative tool backing the institutional policies, as it is able to predict with good approximation the phenomenon of the abandonment of educational pathways. Finally, an analysis of the weights awarded to the variables is useful to try to identify which factors have more influence in the process of network convergence. As the weights are expressed in positive and negative values, their interpretation is difficult since the sign of a weight does not automatically determine its compensation by a value of opposite sign. In order to assess the relevance of the predictors, some measurements have been calculated, obtaining them by the square root of the sum of squares of the weights of the connections that connect each predictor with each unit of the hidden's layer. This measurement provides information on the major or minor contribution of each variable to the prediction of the output: the higher the square root the more the variable contributes (Table 3).

Table 3. Measurement of the predictors' importance

Connections	Square root of the sum of the squared weight
Left unanswered questions on chemistry (percent)	1.0228
Left unanswered questions on culture and logical reasoning (percentage)	0.5523
Index of logistic discomfort	0.5164
Left unanswered questions on physics and mathematics (percentage)	0.4514
Incorrect answers on chemistry (percentage)	0.4398
Entry program orientation	0.4215
Educational qualification	0.4116
Parents attitude	0.3625
Internal Locus	0.3391
Clinical training performance	0.3164
Working situation	0.3092
Mother_age	0.2850
Transition school to university	0.2802
Mothers' educational level	0.2770
Internet	0.2769
Father age	0.2656
School preparation	0.2623
Dominant language spoken	0.2594
Admission scores	0.2512
Location in Liguria region	0.2445
Students' motivation	0.2362

Connections	**Square root of the sum of the squared weight**
Incorrect answers on biology (percentage)	0.2230
Grade of high school diplomas (based 100)	0.2179
Factors potential dropout	0.2057
Degree programs preference	0.1940
Correct answers on physics and mathematics (percentage)	0.1934
Mothers' occupational status	0.1927
Internal locus	0.1819
Age	0.1811
Computer	0.1738
Correct answers on chemistry (percentage)	0.1664
Citizenship	0.1661
Mother's place of birth	0.1645
Fathers' occupational status	0.1639
Correct answers on biology (percentage)	0.1626
Incorrect answers on culture and logical reasoning (percentage)	0.1588
Worker in the healthcare field	0.1568
Incorrect answers on physics and mathematics (percentage)	0.1529
Previous healthcare degree program	0.1463
University name of previous enrollment	0.1442
Father's place of birth	0.1429
Laurea Degree Program	0.1422
Fathers' occupational status	0.1332
Parents in healthcare	0.1212
Students previously enrolled at university	0.1204
Correct answers on culture and logical reasoning (percentage)	0.1156
Grants and scholarships	0.1021
Gender	0.0982
Left unanswered questions on biology (percentage)	0.0977

Source: own elaboration.

The interpretation of the connection coefficients in fact is not as immediate as in the case of the regression coefficients in linear models, for which reason a substantive interpretation of them will not be given.
From the estimation of the model presented, among the factors, which are most significant in defining the probability of abandoning and the velocity of acquisition of formative credits, emerge:

- Family background: parents' limited educational qualification.
- Formative background before enrollment: provenance from secondary schools different from the lycees, low degree marks.
- Non-participation in experiences of pre-university guidance and/or ineffectiveness of them.
- Choice of enrollment for stereotyped and/or inadequate reasons: relevance of the diploma, possibility of studying with friends, degree course believed prestigious or presumed offering easier job and career opportunities.
- Choice of the course as a makeshift after the exclusion from another limited number course.
- Little interest in what is being studied and has been chosen without a proper motivation.
- Lack of satisfaction with the results achieved, which engages a circuit of demotivation with respect to the work still to be done.

Having a predictive model of career development of university students, in terms of success, irregular paths and abandonment, allows to know precociously the difficulties and thus to initiate targeted policies of guidance and support.

Conclusion

Forecasting student dropout is a crucial and challenging task for schools, universities, policymakers, and educators. It allows the decision-makers to identify programs and strategies to support the persistency and the success in educational systems, but also permit the families to recognize the problems in which their intervention, accompanying or not the institution, is decisive. In fact, in order to innovate the local policies, the only ones able to appeal to the real causes of abandonments, it is necessary to involve on an equal footing all the social actors.

The analysis model is proposed as a simple and flexible pilot instrument to support the activities of monitoring and evaluation of all the educational systems.

Future research should consider the extension of the study to other groups of students not only enrolled in the healthcare area because it could reveal the importance of including additional variables not considered in this research.

In this sociologically renewed view, which therefore does not limit itself to relegate the phenomenon of abandonment in a segment of people weak by definition and conditioned by structural mechanisms of exclusion, emerges a shared perspective of taking charge of the problem and testing corrective and preventive interventions based on the key actors involved in the phenomenon. The fundamental choice to put at the center not the institutions, but all the persons interested is an implicit claim of bias towards the realization of an equitable educational policy, able to bring the system to greater levels of effectiveness and efficiency.

References

Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. *Journal Applied Biomedicine*, 11, 47-58. doi:10.2478/v10136-012-0031-x.

Bentler, P. M., & Speckart, G. (1979). Models of attitude-behavior relations. *Educational Research Journal,* 22, 35–64.

Braxton, J., & Brier, E. (1989). Melding organizational and interactional theories of student attrition: A path analytic study. *The Review of Higher Education*, 13, 47–61.

Cabrera, A. F., Castaneda, M. B., Nora, A., & Hengstler, D. (1992). The Convergence Between Two Theories of College Persistence. *Journal of Higher Education 63*(2), 143–164.

Ethington, C. A. (1990). A psychological model of student persistence. *Research in Higher Education*, 31, 279–293.

European Commission/EACEA/Eurydice, (2013). *Education and Training in Europe 2020: Responses from the EU Member States. Eurydice Report*. Brussels: Eurydice.

Hede, G., & Wikander, L. (1990). *Interruptions and Study Delays in Legal Educations*. A Follow-Up of the Admissions round Autumn Term 1983 at Uppsala University.

Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D., & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor und in herkömmlichen Studiengängen*. HIS: Forum Hochschule, 2.

Heublein, U., Spangenberg, H., & Sommer, D. (2003). *Ursachen des Studienabbruchs. Analyse 2002.* Hochschulplanung Band 163. HIS GmbH Hannover.

Irfan Khan, Y., Zope, P. H, & Suralkar, S. R. (2013). Importance of Artificial Neural Network. In: Medical Diagnosis disease like acute nephritis disease and heart disease. *International Journal of Engineering Science and Innovative Technology* (IJESIT), 2(2):210-217.

Iwai, S.I. & Churchill, W.D. (1982). College attrition and the financial support systems of students. *Research in Higher Education*, 17, 105–113.

Kamens, D. (1971). The college "charter" and college size: Effects on occupational choice and college attrition. *Sociology of Education*, 44, 270–296.

Kriesel, D. (2007). *A Brief Introduction to Neural Networks*. Retrieved from http://www.dkriesel.com.

Larsen, M. S., Kornbeck, K. P., Kristensen, R. M., Larsen, M. R., & Sommersel, H. B. (2013). *Dropout Phenomena at Universities: What is Dropout? Why does Dropout Occur? What can be done by the universities to prevent or reduce it? A systematic review.* Copenhagen: Danish Clearinghouse for Educational Research, Department of Education, Aarhus University.

Liu, B., & Jiang, Y., (2013). A multi-target training method for artificial neural network with application to computer-aided diagnosis. *Medical Physics.* doi:org/10.1118/1.4772021.

Marks, E. (1967). Student perceptions of college persistence and their intellective, personality and performance correlates. *Journal of Educational Psychology*, 58, 210-211.

Naik, B., & Ragothaman S., (2004). Using neural network to predict MBA student success. *College Student Journal*, 38(1), 143-149, 2004.

Pandini, A., Fraccalvieri, D., & Bonati, L. (2013). Artificial neural networks for efficient clustering of conformational ensembles and their potential for medicinal chemistry. *Current Topics in Medicinal Chemistry, 13*(5), 642-51.

Patel, A., Mehta, T., Patel, M., Patel, K., & Patel, N. (2012). Design Porosity Osmotic Tablet for Delivering Low and pH-Dependent Soluble Drug Using An Artificial Neural Network. *Current Drug Deliv*ery, 9: 459-467.

Pincus, F. (1980). The false promise of community colleges: Class conflict and vocational education. *Harvard Educational Review*, 50, 332–361.

Rodriguez Galdon, B., Penna-Mendez, E., Havel, J., Rodriguez, E., & Diaz Romero, C. (2010). Cluster Analysis and Artificial Neural Networks Multivariate Classification of Onion Varieties. *Journal of Agricultural and Food Chemistry*, 11435-11440.

Rossmann, J.E., & Kirk, B. A. (1970). Factors related to persistence and withdrawal among university students. *Journal of Counseling Psychology*, 17, 55–62.

Rossmann, J.E., & Kirk, B.A. (1970). Factors related to persistence and withdrawal among university students. *Journal of Counseling Psychology*, 17, 55–62.

Siri, A. (2014). *Predicting Students' Academic Dropout Using Artificial Neural Network.* New York: Nova Science Publishers.

Stampen, J. O., & Cabrera, A. F. (1988). The targeting and packaging of student financial aid and its effects on attrition. *Economics and Education Review*, 7, 29–46.
Suzuki, K. (2011). *Artificial Neural Networks – Methodological Advances and Biomedical Applications*. InTech.
Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45, 80-125.
Tinto, V. (1987). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago, IL: The University of Chicago Press.
Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition (2nd ed.)*. Chicago, IL: The University of Chicago Press.
Tinto, V. (2010). From Theory to Action: Exploring the Institutional Conditions for Student Retention. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (pp. 51–89).
Waterman, A. S., & Waterman. C. K. (1972). Relationship between freshman egoidentity status and subsequent academic behavior: A test of the predictive validity of Marcia's categorization system of identity status. *Developmental Psychology*, 6, 179.
Wesolowski, M., & Suchacz, B. (2012). Artificial neural networks: theoretical background and pharmaceutical applications: a review. *J AOAC Inter*., 95: 652-68.
Woods, K., & Bower, K.W. (1997). Generating ROC Curves for Artificial Neural Networks. *IEEE Trans Med Imag*., 16: 329-337.
Zhang Q-J., & Gupta, K.C. (2000). *Neural Networks for RF and Microwave Design*, 3, 61-103. Artech House Inc.