BAYESIAN NETWORK MODELING OF ENGINEERING-STUDENT PERCEPTIONS OF SUCCESS IN TERMS OF VOCATION

David Cárdenas-Gonzalo, Marta Rojo-Arce, Hernán Gonzalo-Orden, Ignacio Moreno-Velasco , Mónica Preciado-Calzada

University of Burgos. Avenida de Cantabria s/n 09006 Burgos. España. Telephone: 661316366. Email: dcardenas@ubu.es

Received: 20/Sep/2016--Reviewed: 21/Sep/2016--Accepted: 3/Nov/2016 - DOI: http://dx.doi.org/10.6036/8157

TO CITE THIS ARTICLE:
CARDENAS-GONZALO, David, ROJO-ARCE, Marta, GONZALO-ORDEN, Hernan et al. BAYESIAN NETWORK MODELING OF ENGINEERING-STUDENT PERCEPTIONS OF SUCCESS IN TERMS OF VOCATION . *DYNA*, July 2017, vol. 92, no. 4, p.399-403. DOI: http://dx.doi.org/10.6036/8157

ABSTRACT:

The influence of vocation in success can be studied through various surveys, but a model to validate the results of such studies with a high probability of accurately represent the reality studied is needed. Bayesian networks meet this requirement.
Students who enroll in an engineering degree may have to choose engineering disciplines that differ from their initial choice. In some cases, these differences may impair their professional commitment and leave them with the feeling that his final choice was wrong.
This study is intended to quantify the perceptions of success of engineering students in terms of vocation and remedy deficiencies that may appear in their attitudes.
As for the methodology used have been the Bayesian networks which have given us the choice of a probabilistic model through a series of variables that will validate this model considering parameters such as the ROC (Receiver Operating Characteristic) curve, can calculate probabilities of success in terms of vocation of our engineering students.
As for the main conclusions of this study, it was those relating to professional expectations and learning, having more influence on students whether they are or are not in the specific field of engineering that reflects his vocation. Success rates increase considerably when students perceive variables as "find useful work", "find an easy job" or "enjoy learning".
In short, helping students to not get frustrated, you think they will find employment and that their jobs will help them feel useful in the future. Therefore, it is essential to speak from the university level on professional life from the first year until graduation and the students consider their choice of college.

Keywords: Bayesian Network; Engineering Vocation; Success

1. INTRODUCTION

Student success during undergraduate studies may be interpreted in different ways. The definition of the success rate (an information-system indicator), used at a public University in Spain, is the ratio of the number of credits passed by all students enrolled on the degree course over total obtainable credits for all those students. In many cases, the attrition rate and the issues surrounding dropouts have been studied, rather than the success rate: bad experiences in accommodation, lack of support from teachers and tutors, and excessive workloads, among other reasons [1].Specifically, in the field of engineering, there is increasing concern over attrition and the absence of any links between dropouts and academic results [2-3].

Normally, persistence in the first year and graduation rates are standard measures of both academic quality and institutional effectiveness [4].

In the present research, we wish to delve a little further into the definition of success in terms of credits passed. In this article the success has been analyzed in terms of the feelings of students upon completion of their degree course of being happy with what they have studied and of not having studied the wrong subject.

The three most influential aspects in the perception of success were identified: vocation, integration + learning, and future expectations.

In this research, we have emphasized an observation regarding vocation that is often overlooked in engineering degrees. Students wishing to study engineering often have to commit to a branch of engineering from the beginning of the degree. It will hereafter be assumed unless otherwise specified that student sare following the vocation of their choice when enrolled on the branch. If for any reason (minimum mark required, economic questions, distance from home ...), while wishing to study engineering, they were not following the branch they had chosen as their first option, we will consider that they are not following their vocation. The question remains: is that particular vocation so important? Or with a strong vocation as an engineer, although not in the specific branch they desire, if other aspects are positive, will the students be able to consider their studies a success? There is not much research in this area.

Studies at the University of Oklahoma as, [5] suggested that one sixth of engineering students do actually change their branch of study.

The second variable, integration + learning, represents what is built up during the period of study at university. Although very important, this study has not focused on issues linked to student integration as much as with vocation-related issues. These variables of integration have also been important in the model created. So variables as "get good grades" or "enjoy learning" are variables directly related to the integration of the student in the University that have been linked as can be seen in the chapter of results with the "success" of the student.

The other factor, learning, has to do with student perceptions of the subject matter learnt and how difficult it is to learn. In fact, it would have something to do with the concept of self-efficacy defined as the strength of one's belief that a specified level of attainment based on ability can be achieved [6] The concept of self-efficacy in relation to vocation and choice of study has been extensively studied [7, 8, 9].Let us say that if the vocation is successfully chosen and the student sare convinced of their choice, then the commitment of students towards their studies is easier. On factors that affect self-efficacy that have been linked to success [10-11] it was found that subject comprehension and motivation were of great importance (motivation was slightly more important for women than for men). This motivation could be linked to vocational and professional expectations of a not-too-distant future.

The third variable, future expectations, plays a major role in the fact that engineering studies sometimes involve great efforts for the students. If they had the perception that their professional life was going to reap benefits, the "sacrifices" made may be worthwhile.

A survey has been conducted at a public University in Spain (School of Industrial Engineering) during the years 2012-2015, with various questions (Authors, 2015) to study all these fundamental aspects. Subsequently, it was decided to evaluate the responses according to the Bayesian Networks model. The questionnaire was complete though short (one sheet) and answers were rated from 1 to 5 with spaces available for open answers.
Certain questions are listed below to validate this model, as representative of those aspects. These questions and other factors listed below are referred to as variables:

- Those which belong to the group of future expectations: "make money", "to be useful to society, "easily find work," "find an interesting job" and "enjoy while learning".
- Those forming part of the group of learning: "enjoy learning" and "get good grades"
- Vocation is a variable that identifies the student.
- Sex which will allow us to conduct comparative studies between men and women (through the survey of first-year students). Over the past few years, significant samples of women have been unavailable. All of this, because the percentage of women compared to men in the survey was below 25%, respectively, 24% and 76%.
- Finally, the fundamental variable of our study will be "success". From this point of view has been analyzed success as the feeling you have students studying a degree, that he is happy with what he has studied and has not been wrong. Bayesian Networks are an increasing popular method in the field of data mining (statistical automatic learning techniques) that permit the extraction of "knowledge" from existing data in a specific

Publicaciones DYNA SL -- c) Mazarredo nº69 - 4º -- 48009-BILBAO (SPAIN)
Tel +34 944 237 566 – www.revistadyna.com - email: dyna@revistadyna.com

problem like databases, surveys, transactions, etc.[12]. In this article, the method of building a probabilistic model was chosen and therefore the relationship between the above-mentioned variables and the success of our students is also analyzed.

Bearing this in mind, the initial probability of each variable ("enjoy learning", "make money", "good grades", "getting an interesting job", "feeling useful to society" , “sex”, "enjoy while learning", “vocation”, "get good grades" and "succeeding") may be defined, taken on this occasion from the responses to the survey administered to students. In other words, there will be five different probabilities for each variable, corresponding to the five states or groups of each one of them.

However, there are relationships of dependence and independence between the different variables that make it possible to factorize this joint-probability distribution (JPD), reducing the number of parameters. For example, an expert can give the following local dependences attending to fundamental concepts in what are called intelligent systems to later create probabilistic networks. These concepts are:
• Knowledge Base (relevant information on the perception of success, by different authors). That is, the bibliography used for finding the variables that will define the model.
• Module reasoning gets conclusions on the level of perception of student success. For example, determine the probability that the student be successful in their studies.
In response to the knowledge base they have been obtained variables such as:

a) The variables "enjoy learning" and "good grades" are independent of each other.
b) The variables "finding interesting employment" and "being useful to society" influence the variable success, and at the same time the variable "finding interesting employment" affects "being useful".
c) Finally, the variable "finding interesting employment" is affected both by "enjoy learning" and "good grades".
d) Given paragraph and paragraph a, this implies that with some knowledge of "finding interesting employment", the variables "enjoy learning" and "having good grades" become dependent (conditional dependence).

Thus, no further direct dependencies are encoded from local dependencies, such as for example, the relationship between the variable "enjoy learning" and "success". Although they are independent variables, when the variable "find interesting employment" is known, they become dependent when the variable "being useful" is known. In this way, the probability of success for each of the model variables may be calculated depending on its status.

As [13] have affirmed, probabilistic networks combine graphs and probability to define models that comply with a certain set of dependency relationships. The dependency relationships between variables are expressed in a graph, a visual and intuitive representation; the nodes of the graph represent the variables and the edges of the graph, the relations between them. The graph provides a criterion to factorize the joint probability distribution (JPD).

It is important to note that these variables have been linked in several studies, such as those mentioned above, although none of them have quantified and validated the value of those variables.

The variables under consideration are key to a successful career. This observation has been empirically demonstrated using Bayesian networks and validated using the ROC curve. With possible endless combinations, the first step was to validate good variable combinations, so as then to perform sensitivity studies. Bayesian networks were used here to present interesting results on the most significant combinations that students at this University have expressed in pursuit of success.

The main objective of this study is to analyze the influence of the three variables, vocation, efficiency, and the future expectations of students at this University on the probability of their success. The study was conducted in four parts:

1. A data-mining selection was carried out taking into account the relevance of the data to our study, in order to work with a reasonable source of data. The variable "Success", chosen a priori on the basis of previous literature, serves as the output attribute.

Publicaciones DYNA SL -- c) Mazarredo nº69 - 4º -- 48009-BILBAO (SPAIN)
Tel +34 944 237 566 – www.revistadyna.com - email: dyna@revistadyna.com

2. A ranking of the most important predictors affecting the principle variable of the study ("Success") was established, by means of strong variable evaluators together with data mining and "Matlab" programming language tools. This program will be used to calculate the joint probability function (FPC) of each of the combinations of the variables that will influence the primary variable of our study will be the "perception of success" The joint probability function you get with "Matlab" is the type:

$$p(x) = \Pi_i \; p(x_i | \; \pi_i)$$

Where Π_i is the set of parent variables of the node xi of the graph.

3. Through Bayesian networks, the relations of conditional dependence will be found between some variables and others.
4. After this, probability calculations of the fundamental variable of the study, success, will be obtained, with respect to the other variables involved in the creation of the model. That is, the variable "Success" will be the reference variable or output variable relative to the combination of the other variables in the model.
An example is exposed with five variables of our study model: enjoy learning, have good grades, get an interesting job, be useful to society and succeed, as can be seen in Figure 1, with 5, 5, 5, and 5 different states of each variable respectively, there will be a joint probability function (PCF) of all with approximately 5x5x5x5x5 = 3125 parameters to define it.

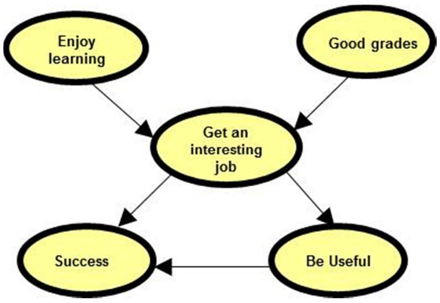

Fig. 1. Example of Bayesian network with variables of the model.

This set of relationships of dependence and independence encoded in the graph in Figure 1, make the joint probability function is to factor as follows (calculating the joint probability function):
P(A,N,C,S,E)=P(A) P(N) P(C|A,N) P(S|C) P(E|C,S)
Being "A" variable "enjoy learning", "N" "have good grades," "C" "Getting interesting work", "S" "be useful to society" and "E" the "success".
The number of parameters of this joint probability function is now 185 parameters (5 + 5 + 5 * 5 * 5 + 5 * 5 + 5 * 5 * 5 = 185).
Thus, from local agencies indirect dependencies the relationship between the variable "enjoy learning" and "success" are also encoded, for example. While independent variables known variable "get an interesting job", they remain dependent variable known "useful".
So, this way, you can calculate the odds of success for each of the model variables depending on their status or evidence.

2. TOOLS AND METHODS

The survey was conducted on 243 freshmen Engineering during the years 2012 and 2015 to different populations, collecting a variety of questions (55 questions) properly analyzed by teachers and primarily aimed at analyzing the needs and concerns in the university sector . Overall, among others, questions such as are collected:
• What were your options when you chose your university career?
• Why did you choose the career?
• What do you expect to find when you get the degree?
All the questions have their weighting of 1 to 5.

The numbers of questions and the fields that are covered have been enlarged with each successive survey, maintaining a number of key questions, to permit the analysis of trends in the university sector.
Survey preparation included the revision of the statistical process and the design of a strict quality-information system using the best available methodology. A study of [17] specifies specifically through various rigorous theoretical procedures abandonment of higher education students provides very similar situations that are discussed in this article.

The information quality-control system guided the development of the survey. This quality control system was developed by university staff.

2.1 STATISCAL METHODS

In this section, the probabilistic methodology of Bayesian Networks will be developed to construct the model of first-year students.

Relations between the model variables are represented intuitively through directed graphs, encoding both the marginal and the conditional dependencies of the different variables. This permits the visual exploration of relationships that may exist in the data set. In the model generated through Bayesian networks, as seen in Figure 2, "success" is shown to depend mainly on variables related directly with the vocation and sex and indirectly with the efficiency and future expectations. In this research, our principal focus is on obtaining the most representative and the strongest relations that are modeled, which will allow us to draw conclusions to improve the chances of success of the students at the University which has been analyzed. The variables that directly influence a given variable are called the "parent variables".

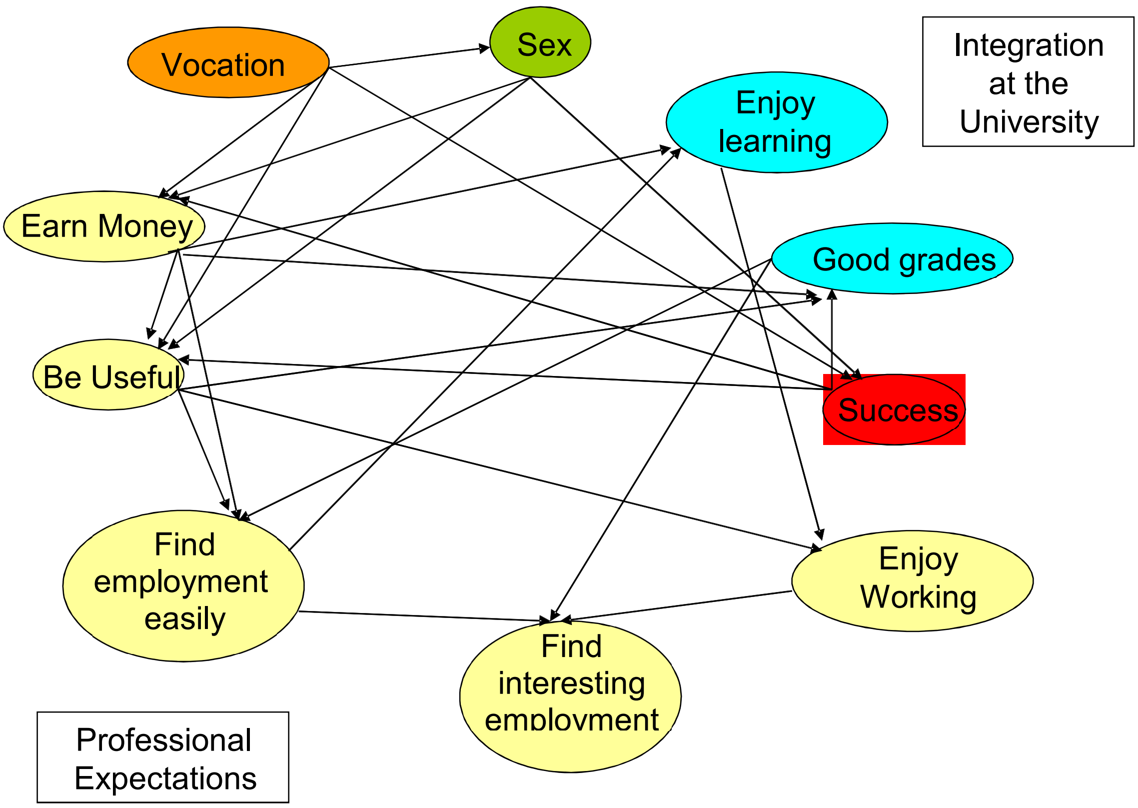

Fig. 2 .Model: "Professional expectations-Integration at the UBU- Vocation–Success" for 1st-year students

The relations that determine the effect of the variable "Success" on the set of the other variables is shown in the above figures. Generally, the Bayesian networks can be said to permit the construction of global probabilistic models from a set of variables

X = (X1 ...... Xn) on the basis of a particular data set.

This model explicitly represents knowledge of the given problem, in probabilistic terms, through a joint probability distribution of the variables (Bayes theorem):

p(x) =p(x1,....,xn).

A joint probability is not easily defined in the proposed model, as there are many degrees of freedom that exist. The problem has been efficiently solved, limiting the number of degrees of freedom on the basis of data-related dependencies and independencies shown in the resulting graph. Thus, the joint probability distribution has been defined through a factorization of local probability distributions, based on the conditioned probability of each variable with regard to its parents:

$$p(x) = \Pi_i \; p(x_i | \pi_i)$$

Where Π_i is the set of parent variables of the node xi of the graph.

Having defined the probabilistic model p(x), Bayesian networks calculate the initial or the marginal probabilities p(Xi = xi), for each state of each variable xi of x. These are called 'a priori' or initial probabilities and correspond to the initial state of the variables in the data set (the frequencies of the different states).

Having obtained the probabilities "a priori", the networks permit the calculation of conditioned probabilities $p\left(\frac{x_i}{e}\right)$ for each variable $X_i \notin E$, given certain evidence "e" and (for example, given the value of some variables of the model: vocation=1; e=1). These new probabilities reflected the effect of the evidence on the remaining variables of the model. The difference between the marginal and the conditioned probabilities allow us to analyze which responses have the greatest influence on whether the level of stress is high, thereby permitting the results of the survey to be explored and quantified through sensitivity studies that will be presented in the chapter on results.

The most complex process in the use of Bayesian Networks is training the model with the data. To do so, different algorithms were used based on statistical tests of dependency and on automated searches of optimal models, which represent the set of given data [12].

2.2 . VALIDATION OF GLOBAL MODEL

The results of the Bayesian models have been validated using the ROC curve (Receiver Operating Characteristic) method, which has had different, uses over the years;

In recent years, authors such as [14] and later on [15] stated that they can become very useful to demonstrate different evaluation techniques in machine learning and "data mining".

This ROC validation technique can be defined as a graph in which two variables are identified: sensitivity and specificity for a binary classifier system, as the discrimination limit varies. Another possible definition of the graph is the identification of true positives (TP) against the definition of false positives (FP), also against changes in the discrimination limit, on which basis it may be decided whether a case in any one set is positive.

Publicaciones DYNA SL -- c) Mazarredo nº69 - 4º -- 48009-BILBAO (SPAIN)
Tel +34 944 237 566 – www.revistadyna.com - email: dyna@revistadyna.com

To analyze whether the results have been good or not, there are several identifiers. The most widely used is the AUC (area under the ROC curve). This is identifier that has been used in the article.
It is calculated by the following relationship according to [14]:

- G+1=2AUC

Where G es el coefficient of Gini

$$G_1 = 1 - \sum_{k=1}^{n} (X_k - X_{k-1})(Y_k + Y_{k-1})$$

X: is the cumulative probability of the study variable (in our case the variable success).
Y: it is the cumulative probability of other variables in the model

All results for combinations of study have been above 0.5 which is what we recommend the Roc space, according to studies [14] and [15]. Thus the Bayesian Network is validated.

3. RESULTS

The following results were obtained for 1st-year students.

First, it should be said that all the variables were analyzed in their state number 3 of the 5 states , except for vocation, which was analyzed first of all in a "good" state for men (Table 1 and 2)

VARIABLES		GROUP	% PROB. OF SUCCESS
SEX		Man	51.23%
VOCATION		Good	
ENJOY LEARNING	Integration at the University	3	
GOOD GRADES		3	
FIND EMPLOYMENT EASILY	Professional expectations	3	
EARN MONEY		3	
FIND INTERESTING EMPLOYMENT		3	
ENJOY WORKING		3	
BE USEFUL TO SOCIETY		4	
INITIAL PROBABILITY OF VARIABLE "SUCCESS"		2	16.59%

Table I. Probability of success for men with a vocation in a good state and being useful to society.

VARIABLES		GROUP	% PROB. OF SUCCESS
SEX		Man	41.94%
VOCATION		Bad	
ENJOY LEARNING	Integration at the University	3	
GOOD GRADES		3	
FIND EMPLOYMENT EASILY	Professional expectations	3	
EARN MONEY		3	
FIND INTERESTING EMPLOYMENT		3	
ENJOY WORKING		3	
BE USEFUL TO SOCIETY		4	
INITIAL PROBABILITY OF VARIABLE "SUCCESS"		2	16.59%

Publicaciones DYNA SL -- c) Mazarredo nº69 - 4º -- 48009-BILBAO (SPAIN)
Tel +34 944 237 566 – www.revistadyna.com - email: dyna@revistadyna.com

Table II.Probability of success for men with vocation in a bad state and being useful to society.

VARIABLES		GROUP	% PROB. OF SUCCESS
SEX		Man	64.29%
VOCATION		Bad	
ENJOY LEARNING	Integration at the University	3	
GOOD GRADES		3	
FIND EMPLOYMENT EASILY	Professional expectations	3	
EARN MONEY		4	
FIND INTERESTING EMPLOYMENT		3	
ENJOY WORKING		4	
BE USEFUL TO SOCIETY		4	
INITIAL PROBABILITY OF VARIABLE "SUCCESS"		2	16.59%

.

Table III. Probability of success for men with vocation in a bad state being useful to society, earning money, and enjoy working.

The results for women are presented in the following tables: First, the results with good vocation were considered (Table IV).Secondly, table V was obtained with the chances of succeeding when vocation is in a poor state.

VARIABLES		GROUP	% PROB. OF SUCCESS
SEX		Woman	26.53%
VOCATION		Good	
ENJOY LEARNING	Integration at the University	3	
GOOD GRADES		3	
FIND EMPLOYMENT EASILY	Professional expectations	3	
EARN MONEY		3	
FIND INTERESTING EMPLOYMENT		3	
ENJOY WORKING		3	
BE USEFUL TO SOCIETY		4	
INITIAL PROBABILITY OF VARIABLE "SUCCESS"		2	16.59%

Table IV. Probability of success for women with vocation in a good state and being useful to society.

Publicaciones DYNA SL -- c) Mazarredo nº69 - 4º -- 48009-BILBAO (SPAIN)
Tel +34 944 237 566 – www.revistadyna.com - email: dyna@revistadyna.com

VARIABLES		GROUP	% PROB. OF SUCCESS
SEX		Woman	64.29%
VOCATION		Bad	
ENJOY LEARNING	Integration at the University	3	
GOOD GRADES		3	
EASILY GET WORK	Professional expectations	3	
EARN MONEY		4	
GET AN INTERSTING JOB		3	
ENJOY WORKING		4	
BE USEFUL TO SOCIETY		4	
INITIAL PROBABILITY OF VARIABLE "SUCCESS"			16.59%

Table V .Probability of success for women with vocation in a bad state, being useful to society, earning money and enjoy working.

4. DISCUSSION

Given the different levels of maturity and the different views of the 1ststudents in this study, the results of each group are separately discussed. However, as a general point, it may be remarked that among the questions to identify important variables for "Integration and university life", the only two that could be validated were to "enjoy learning" and "good academic results." These two just have to do with the concept of self-efficacy, which, as already mentioned in the introduction, has repeatedly shown itself to be one of the most important.

If we combine the variable vocation for the (male) gender with all the other variables in their intermediate state (3),the likelihood of success stands at 12% in men and 10% in women. In other words, this means that we would have to increase the level of some of the variables to over the initial 16.59%.

We conducted the second study, where it may be seen that an increase in attachment to the variable "Being Useful" up to level 4 is accompanied by a rise in the probability of success to 51.23%for men and to 26.53% for women. This result contradicts those reported by [2], where being useful to society appeared to be much more important for women than for men. In fact, in men with vocation in either a good or a bad state, the expectation of being useful to society increases the probability of success to 51.23% and 41.94%, respectively, while the feeling of success with vocation in a good and a bad state in women barely reaches 26.53% and 33.33%,respectively. However, this result shows that the probability of success does not depend so heavily on vocation, because the increase was significant with a bad vocation and feeling useful to society.

Nevertheless, even when there is bad vocation, if the student feels useful to society, has money-earning expectations (state 4) and enjoys working (state 4), then the probability of success rises to up to 64.29% in men and women. It should be noted that such expectations may not represent the same in men and women, as according to [16],and [17], men expect to earn more than women when studying at university. In their study, [18] distinguished between per sisters and non-persisters and tried to see what made the per sisters stay. The motives of two women with no clear engineering vocation, for not abandoning their course was to find a good job in the future, and salary was explicitly mentioned in the interviews reported in that study.

Publicaciones DYNA SL -- c) Mazarredo nº69 - 4º -- 48009-BILBAO (SPAIN)
Tel +34 944 237 566 – www.revistadyna.com - email: dyna@revistadyna.com

In other words, the variables under consideration are very important to manage to achieve success in our career. Above all, we have empirically demonstrated the most important aspects by using Bayesian networks (statistical techniques of automated learning) that are validated using the ROC curve. (It is important to validate good combinations). The combinations are infinite. Hence, what has to be done first of all is to validate the good combination of variables, so as then to conduct sensitivity studies, as has been done.

5. CONCLUSIONS

First I must say that not all survey questions have been chosen to create the model variables although in principle could be interesting. Variables as "the influence of acquaintances or friends when choosing your career" and others like "carry out my ideas or projects" have not been enough to integrate them into the model consistency as AUC values obtained in the ROC curve established below 0.5.

It is necessary to analyze whether all the surveys are consistent and discard those that do not meet certain criteria. Bayesian networks are a widely used method in various fields for the validation of certain results. They are these and not others that a priori might look like the best (for example those that respond affirmatively or negatively to the highest degree) from which we can draw conclusions.

The specific vocation of students is certainly an important factor to consider, although our students express greater concern (when they finish and even when they are in their first year) over their future employment. Expectations with regard to satisfactory employment can raise doubts over whether the choice of course was the right one.

The study had to focus on 1st-year students (with the highest attrition rate) and the classes they followed. This can be done through laboratory practical, close to reality, giving real examples that make them see that what they are studying has real-world applications. So, conferences and, of course, practicums and end-of-course projects in companies can bring them into the real world and help them in their future search for employment. If students feel that they are on the right track to achieve success, i.e. they were not wrong about their career choice; their motivation will increase, even if their initial vocation was not that particular branch of engineering.

The fact that less importance is attached to the specific vocation, when defining successful career choices, raises the vocation of engineer more generally, so it would also be necessary for students to be very clear over different career paths before enrolling on a particular engineering course.

A further variable of importance throughout the study has been to "enjoy learning", which of course can always be improved, if teachers endeavor to make their subjects attractive to students. In fact, the following conclusion may also be added to the above: if students realize that study is of interest for their future, receive information with pleasure and are satisfied with the degree that they decided to study.

REFERENCES

[1] Bandura, A. 1986, Social foundations of thought and action, Englewood Cliffs, NJ, Prentice Hall.
[2] Betz, NE. & Hackett G., 1981, The relationship of career-related self efficacy expectations to perceived career options in college women and men, J. of Counselling Psychology, 28, 399-410.
[3] Castillo, E., J. M. Gutiérrez and A. S. Hadi (1997). Expert systems and probabilistic network models, Springer Verlag.
[4] Chung, Y B 2002, Career decision-making self-efficacy and career commitment: gender and ethnic differences among college students. J. of Career Development, vol. 28, issue 4, 277-284.
[5] Gutiérrez, J. M., R. Cano, A. S. Cofiño and C. M. SORDO (2004). Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas. Madrid, Ministerio de Medio Ambiente, Universidad de Cantabria e Instituto Nacional de Meteorología
[6] Haag, S., Hubele, N., García A., McBeath, K., 2007, Engineering undergraduate attrition and contributing factors, Int. J. Engng. Ed., vol. 23, pp 929-940.
[7] Hackett, G., Betz, N. E., Casas, J. M., and Rocha-Singh, I. A., Gender, ethnicity, and social cognitive factors predicting the academic achievement of students in engineering, J. of Counseling Psychology, Vol. 39, No. 4, 1992, pp. 527-538.

[8] Hamshire, C., Willgoss, T. G., Wibberley, C., What are reasonable expectations? Healthcare student perceptions of their programmes in the North West of England, Nurse Education Today, 33 (2013), pp. 173-179.
[9] Hand, D. J. and R. J. Till (2001). “A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems.” Machine Learning 45(2): 171-186.
[10] Hartman, H., Hartman, M., Leaving engineering: lessons from Rowan University’s College of Engineering, J. of Engineering Education, 95, 1, 2006, pp. 49-61.
[11] Hogue, M., Yoder, J.D. and Singleton, S.B. (2007), “The gender wage gap: an explanation of men’s elevated wage entitlement”, Sex Roles, Vol. 56 Nos 9-10, pp. 573-579.
[12] Huang, G., Peng, S. S., Entry and persistence of women and minorities in College Science and engineering education, National center for education statistics, research and development report, 2000.
[13] Lent, R. W., Brown, S. D., Schmidt, J., Brenner, B., Lyons, H., and Treistman, D., Relation of contextual supports and barriers to choice behavior in engineering majors: test of alternative social cognitive models, J. of Counseling Psychology, Vol. 50, No. 4, 2003, pp. 458-465.
[14] Matusovich H. M., Streveller, R. A., Miller R. L., 2010, Why do students choose engineering? A qualitative, longitudinal investigation of students’ motivational values, J. of eng. Education, October, 289-303.
[15] Pajares, F., Self-efficacy beliefs in academic settings, Review of educational research, vol. 66, n 4, 1996, pp 543-578.
[16] Seymor, E., and Hewitt, N. M., 1997, Talking about leaving: Why undergraduates leave the sciences, Oxford: Westview Press
[17] Schweitzer S, M Lyons S., Kuron L. K. J., Ng E. S. W., 2014, The gender gap in pre-career salary expectations: a test of five explanations. Career Development International, vol. 19, 4, pp. 404-425.
[18] Tinto, V. “Dropout from Higher Education: A Theoretical Synthesis of Recent Research.” Review of Educational Research, 1975, 43 (1), 89-125.
[19] Zou, K. H., A. J. O’Malley and L. Mauri (2007). “Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models.”Circulation 115(5): 654-657.