Bulletin of the *Transilvania* University of Braşov
Series V: Economic Sciences • Vol. 10 (59) No. 2 - 2017

Assessing the impact of attendance in students' final success using the Decision-Making Tree

Sadri ALIJA[1] Benjamin APAM[2]

Abstract: *In this paper, we use the decision-making tree to explain the impact attendance has on students' final success. The paper analyses the results of 56 students in 3 subjects during the academic year 2016/2017 (first, second and third-year students of Business Mathematics, Statistics and Managerial Economics at the SEE University in Tetovo). The results show that attendance is the most important of the 5 attributes in this study, placing it at the root of the tree. In constructing the Decision-making Tree, we have used the ID3 Algorithm within the Weka software package.*

Key-words: *Student's attendance, Decision-Making Tree, ID3 Algorithm, Weka Package*

1. Introduction

In present -day conditions, when the information and communication technology is developing vigorously, and electronic devices and services take, on a daily basis, a great amount of our time, student's attendance remains in the spotlight to many Universities.

At the South East European University, student-centered teaching has an important place: there has been given great importance to teaching and learning interactively.

On the other hand, knowledge of mathematics is crucial both during studies in enabling an individual to study other subjects in her/his curricula, and also afterward in his/her career by creating better chances to get a job, having higher productivity at her/his workplace and of course earning a higher salary and other benefits.

The question why students miss the classes is raised by many researchers. They find different reasons and explanations that point to why students are missing classes (Gump 2006; Nicholl and Timins 2005; Hughes 2005; Timmins and Kaliszer 2002; Hunter and Tentley 1999; Longhurst 1999). Studies show that some of the

[1] Faculty of Business and Economics, South East European University, Tetovo, Republic of Macedonia, s.aliji@seeu.edu.mk
[2] Department of Statistics, Bolgatanga Polytechnic

reasons are valuable and occur as consequences of the circumstances of their daily lives. Some studies indicate that student attendance is connected with their good health and this effectively results in higher academic achievements.

Through this study we try to shed light on the attendance issue in a way that will incite students to think carefully when they decide to miss the lectures and exercises in the teaching and learning process. To promote a better understanding of the educational process undertaken in these circumstances raises the problem of making an optimal decision. Thus, we mention the methods of decision-making on student attendance at lectures and exercises and on their final success/failure.

The Decision-making Algorithm is the ID3 Algorithm built by Quinlan J. Ross in 1986. ID3 constructs a Decision-making Tree from a set of fixed, usually discrete, data. The Leaf of the tree contains the name of the class attribute, and if the vertex is not a leaf then it is a decision vertex, which serves as a test attribute for new branches of the tree.

The ID3 Algorithm uses the Shannon entropy (Shannon, 1948), where the entropy in the theory of information measures how certain or uncertain the value of a variable by coincidence is. A smaller value implies less uncertainty, whereas the bigger value implies more uncertainty.

If we have a set S with n-attributes, which contain different values, then the entropy is defined as:

$$Entropy\ (S) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i \tag{1}$$

where's p_i is the proportion that S belongs to class *i*.

In building a Decision-making tree, the ID3 algorithm also uses a statistical feature called Information Gain (IG), which measures the effective changes of entropy after a decision is made based on the values of an attribute. In the context of building a Decision-making tree, we are interested in knowing how much information is needed about the outflow attribute, which can be gained by knowing the value of an attribute.

The formula for calculating the IG is:

$$IG(S,A) = Entropy\ (S) - \sum_{j=1}^{n} \left[p_j \cdot Entropy\ (p_j) \right] \tag{2}$$

where's p_j is the set of all possible values for attribute A.

2. Research methodology

This research is focused on the success of students from the Faculty of Business and Economics, based on their attendance at lectures and exercises. The data was collected from 56 students studying at the Faculty of Business and Economics, SEE University in Tetovo. We have concentrated on three subjects: Business Mathematics, Statistics and Managerial Economics, studied by first, second and third-year bachelor students. Five variables were collected from the student's file: student's gender, year of study - subject respectively, if the student was working part-time, attendance and the final success of the student.

Gender	Male; Female
Year of study	First year; second year; third year
Employed	Yes, No
Attendance	Att>75%; 50%<Att<75%; Att<50%
Success	Passed; Failed

To illustrate the interpretation of data, the conclusions drawn and the statistical decision-making based on data, we built, in addition to descriptive statistics, the Decision-making tree using the ID3 Algorithm and the Weka software package.

3. Results

In this research, we kept the ratio between genders equal, and as for the distribution between generations, we had 28.57 % first-year students, 42.86 % second-year students and 28.57% third-year students. Regarding the other attribute about working part-time or not, 42.86 % of the participants declared that they were working part-time during the academic year. In terms of attendance at classes, 32.14% stated they went to classes regularly with over 75% attendance, 35.71% went regularly between 50%-75% of the time and 32.14% did not attend classes regularly, with less than 50% attendance.

In terms of success, 71.43% passed the subjects and 28.57% did not pass them. Table 1 and Figure 1 illustrate the descriptive analysis of the data.

Variable	**Characteristic**	**Percent**
Gender	Male	50.00
	Female	50.00
Year of study	First Year	28.57

	Second Year	42.86
	Third Year	28.57
Working Relationship	Yes	42.86
	No	57.14
Attendances	Att>75%	32.14
	50%<Att<75%	35.71
	Att<50%	32.14
Success	Pass	71.43
	Failed	28.57

Table 1. *Descriptive analysis of the sample*

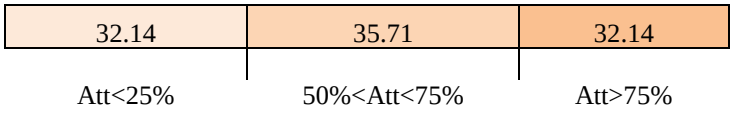

Fig. 1. *Proportion of lectures and exercise sessions attended?*

We calculate the Information Gained to sort the attributes and to construct the Decision-making tree, where every vertex localizes the attribute that has the highest value of the Information Gained, compared to other attributes that are not being considered across the path from the vertex, classifying step by step in every new subdivision.

According to this information, we are going to choose the attribute that has the highest value by naming it the vertex of our tree. After we have found the vertex, this attribute moves from the set. For the next level, data is divided according to the values of this attribute. The entropy of students' success is given below like the entropy of a whole system:

$$Entropy\ (S) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i = -\frac{40}{56}\log_2\frac{40}{56} - \frac{16}{56}\log_2\frac{16}{56} = 0.863$$

- **We calculate the entropy for attendance**
  - $Entropy\ (S_{Att.>75\%}) = -\frac{18}{18}\log_2\frac{18}{18} = 0$
  - $Entropy\ (S_{50\%<Att.<75\%}) = -\frac{12}{20}\log_2\frac{12}{20} - \frac{8}{20}\log_2\frac{8}{20} = 0.971$

- $Entropy\ (S_{Att.<50\%}) = -\frac{10}{18}\log_2\frac{10}{18} - \frac{8}{18}\log_2\frac{8}{18} = 0.991$

Now let us calculate the information gained of S regarding "Attendance". We have 18 values where "Att>75%", 20 values "50%<Att.<75%", 18 values "Att.<50%", from a total of 56 answers.

$$IG(S, Attendence) = Entropy\ (S) - \sum_{j=1}^{3}\left[p_j \cdot Entropy\ (p_j)\right] =$$

$$= Entropy\ (S) - \frac{18}{56}Entropy\ (S_{Att.>75\%}) - \frac{20}{56}Entropy\ (S_{50\%<Att.<75\%}) - \frac{18}{56}Entropy\ (S_{Att.<50\%})$$

$$= 0.863 - 0.321 \cdot 0. - 0.357 \cdot 0.971 - 0.321 \cdot 0.991 = 0.198$$

- **"Year of study" attribute**

  - $Entropy\ (S_{Sfirst\text{-}year}) = -\frac{10}{16}\log_2\frac{10}{16} - \frac{6}{16}\log_2\frac{6}{16} = 0.954$

  - $Entropy\ (S_{5second\text{-}year}) = -\frac{18}{24}\log_2\frac{18}{24} - \frac{6}{24}\log_2\frac{6}{24} = 0.811$

  - $Entropy\ (S_{third\text{-}year}) = -\frac{12}{16}\log_2\frac{12}{16} - -\frac{4}{16}\log_2\frac{4}{16} = 0.811$

$$IG(S, \text{Year\_of\_study}) = Entropy\ (S) - \sum_{j=1}^{3}\left[p_j \cdot Entropy\ (p_j)\right] =$$

$$= Entropy\ (S) - \frac{16}{56}Entropy\ (S_{Sfirst\text{-}year}) - \frac{24}{56}Entropy\ (SS_{5second\text{-}year}) - \frac{16}{56}Entropy\ (S_{third\text{-}year}) =$$

$$= 0.863 - 0.286 \cdot 0.954 - 0.429 \cdot 0.811 - 0.286 \cdot 0.11 = 0.011$$

- **"Gender" attribute**

  - $Entropy\ (S_M) = -\frac{22}{28}\log_2\frac{22}{28} - \frac{6}{28}\log_2\frac{6}{28} = 0.750$

  - $Entropy\ (S_F) = -\frac{18}{28}\log_2\frac{18}{28} - \frac{10}{28}\log_2\frac{10}{28} = 0.940$

$$IG(Sex) = Entropy\ (S) - \frac{28}{56} Entropy\ (S_M) - \frac{28}{56} Entropy\ (S_F)$$
$$= 0.863 - 0.5 \cdot 0.750 - 0.5 \cdot 0.940 = 0.018$$

- **"Employment" entropy**

  - $Entropy\ (S_{Yes}) = -\frac{14}{24}\log_2\frac{14}{24} - \frac{10}{24}\log_2\frac{10}{24} = 0.980$

  - $Entropy\ (S_{No}) = -\frac{26}{32}\log_2\frac{26}{32} - \frac{16}{32}\log_2\frac{16}{32} = 0.696$

$$IG(\text{Employed}) = Entropy\ (S) - \frac{24}{56} Entropy\ (S_{Yes}) - \frac{32}{56} Entropy\ (S_{No})$$
$$= 0.863 - 0.429 \cdot 0.980 - 0.571 \cdot 0.696 = 0.045$$

The information with the highest value will serve as a vertex of the tree, i.e. the "Attendance" attribute. Since "Attendance" attribute has three qualitative variables, three branches emerge from this vertex that form the first level of the hierarchy. For these vertexes we calculate the Information Gained by building three new branches that belong to the attributes in this way:

- For the quality "Att>75%" as the first branch of the tree, IG will be:

$$IG(S, \text{Year\_of \_study}) = 0;\ \ IG(S, \text{Gender}) = 0;\ \ IG(S, \text{Employed}) = 0$$

All the attributes are the same, so the answer is "Pass". In this case, this is a decision leaf.

For the quality "Att<50 " as the second branch of the tree, IG is:

$$IG(S, \text{Year\_of \_study}) = 0;\quad IG(S, \text{Gender}) = 0.561;\quad IG(S, \text{Employed}) = 0.02$$

After we calculate the entropy for this new system and the Information Gained for every attribute as above, we find out that the highest value has the attribute "Gender" which now serves as a new vertex of decision.

For the quality "50%<Att<75%" as the third branch of the tree, IG will be:

$$IG(S, \text{Year\_of \_study}) = 0.019;\ \ IG(S, \text{Gender}) = 0.046;\ \ IG(S, \text{Employed}) = 0.224$$

So the winning attribute in this third branch is "Employment". By continuing this way we obtain this tree, where in fact the attribute "Year of Study" has minimal values and we do not take it into consideration.

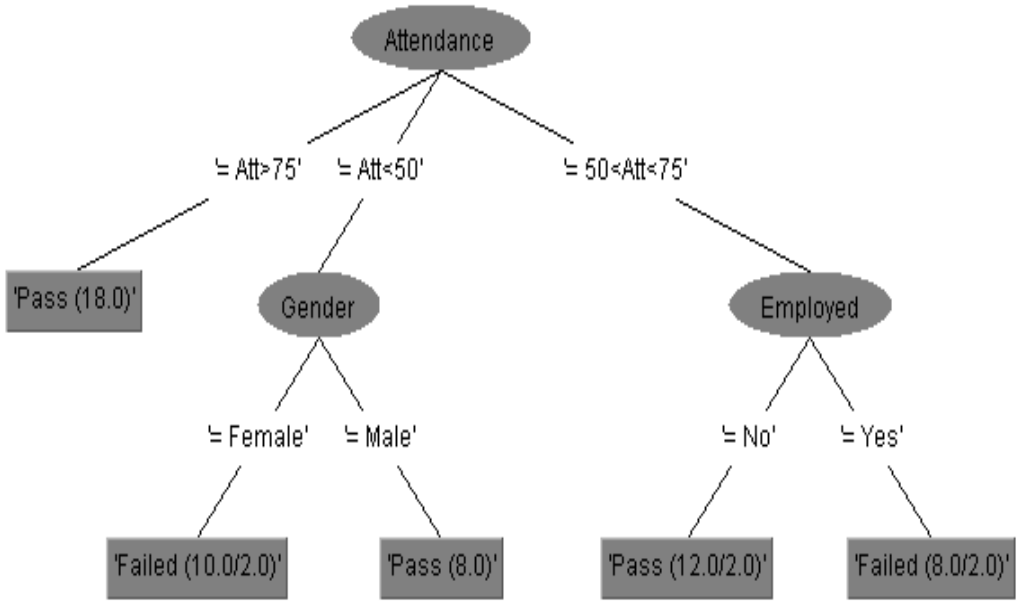

Fig.2. *The decision making tree with Weka*

4. Conclusions and recommendations

Using the data classification of this paper in the form of a Decision tree with the help of the ID3 algorithm, applied to the software WEKA, we found the qualitative values-attributes that weigh more than other values.

The decision leaves were the key points out of which we chose the best alternative. We found that the root of the tree, the attribute "attendance" had greater importance on students' passing rate and that it was followed by the other attributes. The first threshold was the number of students who attended over 75% of the classes regularly, *which resulted in all students passing the exam?*

We think that through this study, using the Decision-making tree, we have provided an additional reasoning on the importance of students' attendance at lectures and exercises on their final success.

Based on the results of this research, the Faculty of Business and Economics ,SEE University, must, in collaboration with the students, analyze the results and

take the necessary measures to encourage student participation in two directions: in the aspect of quality of teaching and in the moral aspect.

5. References

Agresti A., 2002. *An Introduction to Categorical Data Analysis*. New York: Wiley.

Ahmad, A. and Sahak, R., 2009. Teacher-Student Attachment and Teachers'Attitudes Toward Work. *Jurnal Pendidik dan Pendidikan, Jil.* 24, 55–72.

Cader, J., Stevens, D. and Brown, R., 2003. Business Student's Attendance At Lectures. *ANZAC 2003 Conference Proceedings Adelaide*

Cleary-Holdforth, Joanne, 2007. Student non-attendance in higher education. A phenomenon of student apathy or poor pedagogy? *Level3* – June 2007 – Issue 5.

Fox, J., 1997. *Applied Regression Analysis, Linear Models and Related Methods*. Thousand Oaks, CA: Sage Publications.

Kottasz, R., 2005. Reasons for Student Non-Attendance at Lectures and Tutorials: an analysis. Investigations in university teaching and learning, Vol. 2.

Mallik, Girijasankar, [s.a.]. *Lecture and Tutorial Attendance and Student Performance in the First Year Economics Course: A Quantile Regression Approach*. School of Economics and Finance, University of Ëestern Sydney.

MedCalc, Easy-to-use statistical software, 2005. MedCalc Software, Broekstraat 52, 9030 Mariakerke, Belgium.

Mirtcheva, D., 2009. *Attendance & GPA: Health as a Deciding Factor*. The College of New Jersey.

Q.Zhang et al., 2011. Application of ID3 Algorithm in Exercise Prescription. In: *Proccedings of the International Conference on Electric and Electronics*, 2011, pp. 669-675.

Qendraj (Halidini) D., Mitre. Th. and Halidini. E., 2015. An application of decision tree in technology". In: *ISTI (Information Systems and Technology Innovations Inducting Modern Business Solutions)*, 5-6 june, 2015 (Proceeding online)

Qendraj (Halidini) D. and Xhafaj. E., 2015. Evaluating Risk Factors of Being Obese, By Using Id3 Algorithm in Weka Software. *European Scientific Journal*, vol.11, No.24 ISSN: 1857 – 7881 (Print) e - ISSN 1857- 7431, August 2015.

Quinlan, J.R., 1986. Induction of decision trees. *Machine Learning*, 4, 81-106.

Quinlan, J.R., 2009. *C4.5 programms for machine learning.* Online google.

Shannon, C.E., 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27: 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x