ПРИЛОЖЕНИЕ НА КЛАСИФИКАЦИОННИ ДЪРВЕТА ЗА ИЗСЛЕДВАНЕ НА СЪСТЕЗАНИЯ ПО ПРЕДПРИЕМАЧЕСТВО

APPLICATION OF CLASSIFICATION TREES FOR ENTREPRENEURSHIP COMPETITIONS ANALYSIS

Kaloyan Haralampiev
Sofia University

Elica Efremova, Iglika Milosheva
JA Bulgaria

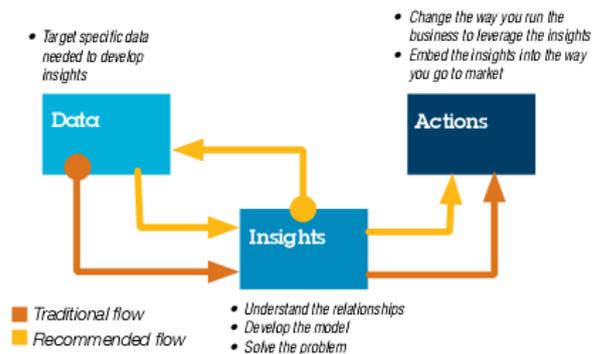**Petko Ruskov**
Sofia University

Abstract

In this paper we present the advantages of the method "Classification Trees" to reveal the predictors for success in Junior Achievement (JA) Entrepreneurship student competitions. JA programs provide high school and post-secondary students the opportunity to experience the entrepreneurship through learning by doing educational methods. The paper reviews the results of analysing a sample of more than 500 high school student companies and 70 university student companies during the last 5 years. As results we identify most important predictors and positive and negative target groups. The classification trees and conclusions are useful for school and university student companies stakeholders.

Keywords: Predictive Statistics, Classification Trees, JA Student Company.

INTRODUCTION

Today's business ecosystem is very dynamic. We study, live and work in a reality of global competition, knowledge economy, successful cooperation between business and education. We reside in a world extremely complete with data. Educational institutions are exclusively positioned to help their students and stakeholders to deal with the explosive nature and complexity of the twenty-first century by generate helpful insight from data and serving as catalysts for innovation. [1,2,3]. In this environment educating young people and tutoring them with entrepreneurial competences, becomes vitally important for their future. Improving and increasing the potential for innovation and entrepreneurial competence will be a major factor to ensure competitiveness on the market for the graduating students realization. [8,9].

Best practices show that "organizations should first pinpoint insights to be leveraged, use available data to test analytic models, then the actions based on those insights can help define the next set of insights and data needed" [12] – Figure 1.

Source: Analytics: The New Path to Value, a joint MIT Sloan Management Review and IBM Institute for Business Value study. Copyright © Massachusetts Institute of Technology 2010.

Figure 1.Lifecycle of data-insight-action [12]

JA programs provide high school and post-secondary students the opportunity to experience the entrepreneurship through learning by doing educational methods.

In this paper we present the advantages of the method "Classification Trees" to reveal the predictors for success in Junior Achievement (JA) entrepreneurship student competitions.

METHODOLOGY AND RESEARCH DESIGN

In the paper we apply a combination of an explicit data base, sophisticated analytical

skills and domain knowledge to uncover hidden trends and patterns in the student entrepreneurship competitions.

Data mining relies on four essential methods: Classification, categorization, estimation and visualization. Classification identifies associations and clusters, and separates subjects under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay." Estimation includes predictive functions or likelihood and deals with continuous outcome variables, such as GPA and salary level. Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. Visualization is used primarily to depict three-dimensional geographic locations of mathematical coordinates.

All level of the education institutions can use four essential methods – classification, categorization, estimation and visualization. Classification could be applied, for example, for a comprehensive analysis of student companies characteristics; estimation is used for predicting the chances of a variety of competitions standings such as creativity, idea identification, business plan preparation and implementation of innovative processes and models for persistence, retention and competition success.

Classification and estimation use either unsupervised or supervised modeling techniques. Unsupervised data mining is used for situations in which particular groupings or patterns are unknown. In student course databases, for example, little is known about which courses are usually taken as a group, or which course types are associated with which student types. Unsupervised data mining is often used first to study patterns and search for previously hidden patterns, in order to understand, classify, typify and code the objects of study before applying theories.

Supervised data mining, however, is used with records that have a known outcome. A graduation database, for example, contains records of students who completed their studies, as well as of those who dropped out. Supervised data mining is used to study the academic behavior of both the graduated and dropouts with the intention of linking behavior patterns to

academic histories and other recorded information.

In the literature persist six goals for application of Classification trees [4,11]:

1. "**Segmentation.** Identify persons who are likely to be members of a particular group." [11] In our case we will use this property for identification of a positive target and a negative target groups. We define positive target groups as groups of student companies with high percentage of national competition winners. The negative target groups are, conversely, the groups with low percentage of winners. The identification of the positive target groups could help us to understand the pattern of success. The identification of the negative target groups will show us where we have to make efforts. Thus we could offer special policies for the specific groups.

2. "**Stratification.** Assign cases into one of several categories…" [11] This is an individual level of the analysis which is not a topic of interest in this paper.

3. "**Prediction.** Create rules and use them to predict future events…" [11] If we succeed to identify stable positive target groups we could prescribe future behavior which could provide the winning of the national competitions.

4. "**Data reduction and variable screening.** Select a useful subset of predictors from a large set of variables for use in building a formal parametric model." [11] The method will help us to identify only these predictors which really affect the dependent variable. This is very important because the high schools database contains 19 predictors and the universities database contains 48 predictors. In the second place, the method ranges the predictors by their importance.

5. "**Interaction identification.** Identify relationships that pertain only to specific subgroups and specify this in a formal parametric model." [11] Building of formal parametric models is future task for us so that for the moment we will not use this property.

6. "**Category merging and discretizing continues variables.** Recode group predictor categories and continues variables with minimal loss of information." [11] If several categories of the predictor have almost the

same winning results they are merged. This is part of the description of the positive and negative target groups.

STUDENT EDUCATIONAL DATA, PROCESSES AND EXPERIENCE

Having in mind the specifics of the different educational levels and the characteristics of student companies the data descriptions are divided in two groups – high schools and universities.

High schools

JA Bulgaria entrepreneurship programs have been implemented in the Bulgarian high school educational system since 1997. So far, more than 200 000 young people have been educated in business and economic competence [5]. The educational method "learning by doing" is most successfully implemented through JA "Company Program" – Table 1.

Table 1. High school student companies in Bulgaria

Year	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011
Number of schools	28	36	36	49	51
Number of students	932	1120	1080	2106	1726
Number of student companies	75	80	75	144	114

The data we have collected for the high school students has been gathered for a five year period (2006-2011). The tool for collecting the data is an application form, which has been translated and adapted into Bulgarian language. Also, we added information from JA student company competitions. The additional data includes specific information, such as participating companies, special awards and the major category (Company of the Year) top three award-winning places. The format of that tool has not been changed for those five years, so after processing the data we have distinguished the following five groups of indicators:
1. Schools profile;
2. Teachers profile;
3. Business consultants profile;
4. Students profile;
5. Student company profile.

The sub categories of each target group include the following variables:
1. School characteristics include: type of school, according to the National classification of the Bulgarian Ministry of Education, Youth and Science, place of school – rural area, central city or municipality city etc;
2. Teachers characteristics include: gender, number of teachers per student company etc;
3. Business consultants are characterized by only one factor – gender.
4. Students characteristics are categorized by the following factors: gender, class in school/year of education, overall number of students in company etc;
5. Student company characteristics: specification of the scope of the company economic activity according to the Statistical Classification of Economic Activities in the European Community (NACE), participation in the national competition for student companies, participation in competitions abroad, if the company has won any special category award or has been ranked among the top three Company of the Year winners. .

The reason that we have indentified only 19 factors for Company Program high school students is that the application form did not allow any changes up to now.

Universities

JA student company program for university students is now officially called Start Up Programme [7]. The program was developed in UK in 1998. Since then it has been implemented in 15 European countries and more than 20 000 university students have participated in the program. In 2007 the program was introduced publicly in Bulgaria. For the purpose of the current study, we processed and analyzed the data collected for the university student companies for the period 2007-2011. Comparable to the high school company program, the Start Up Programme curriculum is based on the "learning-by-doing" educational method. For a period of one term to one academic year students found and operate mini startup companies. They form teams and under the guidance of a university tutor and voluntary business consultant/s generate business idea for a product/service, write a business plan, create prototypes, keep financial

records and have the option to make sales. [10,13] Through their work in the program the university students not only gain knowledge in economics, business and management, but also develop entrepreneurial attitude, as well as, a set of specific functional managerial skills and a variety of soft skills. Because of the educational purpose of the program, student companies are not legally registered as real businesses, but the students have the option to register a real enterprise and still participate in the program. In fact, after passing the program they are even encouraged to transform their educational projects into real businesses. The study period ends with a national competition for the student companies, where a jury of entrepreneurs and business professionals evaluate the students work. After the competition the student companies are liquidated. As described above the direct program stakeholders are: Junior Achievement as a program administrator, the participating universities, tutors, students, business consultants, and the student companies themselves [10].

The number of students that have undergone the study and the number of companies established by them in different universities in Bulgaria are shown in Table 2. The interest in the subject by universities in different fields of study and the progress achieved are impressive.

Table 2. University student companies in Bulgaria

Year	2005	2006	2007	2008	2009	2010	2011
Number of Universities	1	1	1	5	9	13	15
Number of students	25	39	45	51	108	305	500
Number of student companies	1	2	2	6	11	28	35

The tools used for collecting the data are the preliminary intention-to-participate registration form, the actual company registration form, and the company liquidation form. The forms were translated, adapted and developed following the original program materials [14]. The 48 indicators identified from the collected data are divided into 5 major groups:
1. University profile;
2. University tutors profile;
3. Business consultants profile;
4. Students profile;
5. Student company profile.
The indicators within the groups include:
Group 1: University name, Location, Postal code, Field of education etc;
Group 2: Gender, Faculty, Previous program experience etc;
Group 3: Gender, Occupation/Job position, Type of organization, Scope of economic activity according to the NACE, Previous program experience etc;
Group 4: Year of study, Field of education, Gender, Previous program experience etc;
Group 5: Number of participants, Gender, Scope of economic activity, according to the NACE etc.

The identified indicators are selected based on empirical experience and the collector intention to have the most comprehensive profile of the program participants, their activities, the factors influencing the decision to join the program and success factors.

RESULTS, DISCUSSIONS AND FUTURE WORK

The classification trees (see Figures 2 and 3) shows that in all cases the most important predictor is only one. Thus we could identify simultaneously the most important predictors and the positive and negative target groups.
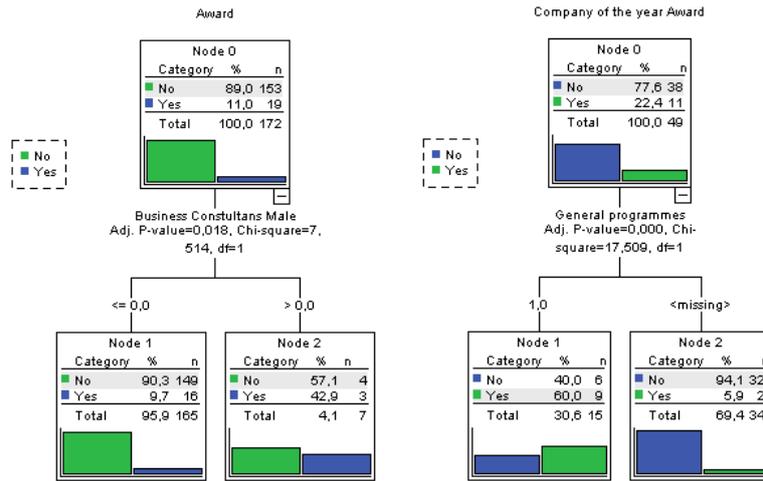
a) high schools   *b) universities*
Figure 2. Winners in the national competitions.
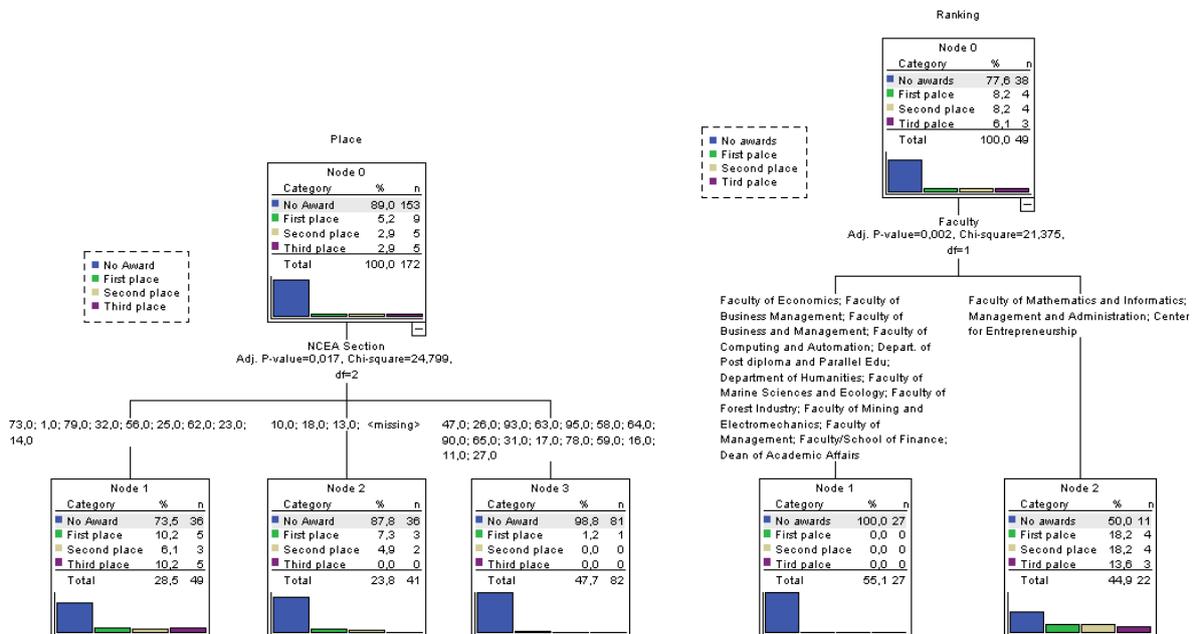
a) high schools   *b) universities*
Figure 3. Place standing in the national competitions.

1. High schools

For the winning of the national competitions the one important predictor is the number of male consultants (Fig. 1a). The positive target group contains the students companies with male consultants and the negative target group contains the student companies without male consultants.

For the place standing in the national competitions the one important predictor is the NACE code, i.e. the nature of the activity of the student companies (Fig. 2a). The positive and negative target groups contain the students companies which activities are listed in the corresponding node.

2. Universities

For the winning of the national competitions the one important predictor is the presence of the general programs in the

corresponding universities (Fig. 1b). The positive target group contains the student companies from the universities with general programs and the negative target group contains the student companies from the universities without general programs.

For the place standing in the national competitions the one important predictor is the faculty of the tutors (Fig. 2b). The positive target group contains the student companies with tutors from the Faculty of Mathematics and Informatics, from Faculty of Management and Administration and from Center for Entrepreneurship. The negative target group contains the student companies with tutors from the all other faculties.

These results present only partially the patterns for success of student companies. To reveal the big picture we have to answer the following challenging questions:
- What types of companies will attract more students?
- What types of companies are more likely to make sales?
- What types of companies are more likely to establish a real business?

For this purpose we are developing new on-line questionnaires and guides for in-depth interviews.

CONCLUSION

Possibly the most useful insight to emerge from this study, however, is not what makes schools and universities the same, but what makes them and their companies different. Based on the presented results and the feedback provided by the different stakeholders we can reach the following conclusions:
- Classification trees are useful method to reveal successful patterns to gain insights into high school and university student competitions;
- The obtained results show that we need to gather more additional data and to expand the number of the indicators.

To improve the efficiency of decision making we plan to continue iteratively our further work following the methodology of Steve LaValle, Michael Hopkins, Eric Lesser, Rebecca Shockley and Nina Kruschwitz.

REFERENCE

[1] Adrian J. Slywotzky, Profit Patterns: A Strategic Shortcut, Harvard Business School Publishing Corporation, 1999.

[2] Clayton M. Christensen, Scott D. Anthony, and Erik A. Roth, Seeing What's Next: Using the Theories of Innovation to Predict Industry Change, Harvard Business Review Press, 2004.

[3] Gary Hamel and C.K. Prahalad, Competing for the Future, Harvard Business Review Press, 2004.

[4] IBM Corp., Predictive Modeling with IBM SPSS Modeler, Student Guide, Course Code: 0A032, ERC 1.0, 2010

[5] JA Worldwide Company Program Kit, 2003

[6] JA-YE Europe, Competitions, http://www.ja-ye.org/pls/apex31mb/f?p=17000:1002:2037617819926485:::1002:P1002_HID_ID:6787

[7] JA-YE Europe, Graduate Programme, http://ja-yecoreprogrammes.org/graduate_programme/

[8] Jing Luan, Data mining applications in higher education, © Copyright IBM Corporation 2010.

[9] Petko Ruskov, Milena Stoycheva, Caroline Jenner, (2011) JA-YE Graduate Student Company Strategy, Proceedings of the International Conference for Entrepreneurship, Innovation and Regional Development ICEIRD 2011, Ohrid, Macedonia 4-7 May 2011.

[10] Ruskov P., Stoycheva M., Todorova Y., (2009) Pattern for Graduate Student Company Life Cycle, 14th European Conference on Pattern Languages of Programs, July 8-12, 2009, Bavaria, Germany.

[11] SPSS Inc. SPSS Classification Trees™ 16.0. 2007

[12] Steve LaValle, Michael Hopkins, Eric Lesser, Rebecca Shockley and Nina Kruschwitz, Analytics: The new path to value, How the smartest organizations are embedding analytics to transform insights into action, IBM Institute for Business Value, MIT Sloan Management Review, 2010.

[13] Stoycheva M., Ruskov P., (2008) Graduate Student Company, How we develop technology entrepreneurship competency in Bulgaria, Stanford Accel REE Europe 2008, September 3-5, 2008, Queens University Belfast, Ireland (invited presentation), http://ree.stanford.edu/europe_2009/programschedule.html

[14] Young Enterprise Graduate Programme Student Kit, Young Enterprise, 2005