DETERMINATION OF FACTORS EFFECTING ENTREPRENEURSHIP TENDENCY OF STUDENTS OF VOCATIONAL HIGH SCHOOL THROUGH METHODS OF DATA MINING

Ali Erbasi
Selcuk University Vocational School of Social Sciences, Alaeddin Keykubat Campus, Selcuklu, Konya-Turkey.
E-mail: aerbasi@selcuk.edu.tr

Safak Unuvar
Selcuk University Tourism and Hospitality Management College, Alaeddin Keykubat Campus, Selcuklu, Konya-Turkey.
E-mail: safakunuvar@selcuk.edu.tr

Osman Unuvar
Selcuk University Vocational School of Seydisehir, Seydisehir, Konya-Turkey.
E-mail: osmanunuvar@selcuk.edu.tr

Yusuf Uzun
Selcuk University Vocational School of Seydisehir, Seydisehir, Konya-Turkey.
E-mail: yuzun@selcuk.edu.tr

ABSTRACT

The aim of this study is to determine factors that effect entrepreneurship of students of vocational high school, through data mining method. 1510 students of vocational high school have been surveyed with face-to-face interview method and decision tree has been applied to acquired data through mining method algorithm. In conclusion, it has been determined that factors of tolerance of ambiguity, internal control and creativeness do not have any relation with demographic features. In concern with demographic features, factor of relation with human has level four, need for success has level five, need for independence and success has level seven and factor of innovativeness has level nine tree-depth relation. Prominent demographic features in these relations have been explained here.

Keywords: Strategic management, data mining, decision tree, entrepreneurship, tendency, factor.

1. INTRODUCTION

Method of data mining is used to acquire potentially beneficial, applicable and meaningful knowledge from large volumes of data stack to make a decision. It is a tool to acquire necessary knowledge to make predictions, solve current problems and make critical decisions. This targeted knowledge is vague, obscure and undiscovered but useful, meaningful and critical. Data mining has two basic functions: functions of prediction and forecasting (classification, regression etc.), function of recognition (grouping, association analysis, consecutive order analysis etc.). Functions of prediction and forecasting are used to predict a future conclusion, benefiting from data of the past. Functions of recognition aim to find the relations, connections and behaviors between data on the data set. (Arguden, 2008). Decision tree, as one of the data mining, is practical in terms of integration with data bases, construction, interpretation and reliability, thus it is one of the most commonly used methods of forecasting and classification technique. Aim of decision trees is to group data within an order to maximize differences in the dependent variable. (Witten and Frank, 2005).
There are some studies in various fields using data mining. For example, management of customer relations (Berson et al., 1999), examination of customer data base (Ata et al., 2008), finance sector (Kovalerchuk and Vityaev, 2001; Koyuncugil, 2007), health sector (Lyman et al., 2008; Prather et al., 1997; Obenshain, 2004; Chae et al., 2003; Fitzpatrick, 2006; Milley, 2000), quality process (Allen and Seaman, 2006), aviation sector (Bineid and Fielding, 2003), information and communication technologies (Kayri, 2008), traffic safety (Solomon et al., 2006; Tseng et al., 2005), modeling of consumer behaviors (Giudici and Passerone, 2002; Min et al., 2002; Timor and Simsek, 2008), development of industrial standards (Hsu, 2009), modeling of current market prices of second-hand automobiles (Aslikan, 2011) etc.

In literature, there are many studies in which data mining has been used upon students. Baradwaj and Pal (2011) used decision tree of data mining methods to determine factors that affect success of students and make various classifications. As a result of study, authors emphasized measurement of students’ success earlier, not during examination period, and contribution for teachers’ consultancy for students. Ayesh et al. (2010) used methods of data mining to analyze learning behaviors of university students. In this evaluation, they developed some suggestions for development of students’ performances basing degrees of mid-term and final exams. Falakmasir and Habibi (2010) used data mining to underline virtual classes for e-learning used for increasing of students’ success. Chandra and Nandhini (2010) determined environmental factors for absence through data mining. Authors have developed structural suggestions for development of academic success of students. Baepler and Murdoch (2010) developed a system for management of courses in higher education basing methods of data mining. As a result of the study, some data and techniques for increasing success of learning and teaching process were suggested. Yu et al. (2010) used methods of data mining for prediction of students’ failure. Authors argued that those predictions would offer important information for top directors so that the university might not potentially have a negative image. Romero et al. (2008) classified the concerning course notes taken by university students according to various factor using method of data mining. Talavera and Gaudioso (2004) used methods of data mining to classify similar student behaviors in the unconfigured collaboration fields. Luan (2004) also used methods of data mining for classification of students, academic planning, transfer prediction and planning of alumni pledges (predicting alumni pledges). In the study, it was indicated that methods data mining can be used to predict how many student graduate within a term, determine tuition fees and course success. It is possible to increase the number of these examples.

In Turkey, studies using methods of data mining upon students are rare. As an example of these studies, Bozkir et. al. (2008) used methods of data mining to determine factors that affect internet use of university students for educational aims. They applied three data mining algorithms concerning topic to 380 students and determined frequent patterns in their thoughts and behaviors for educational internet use. According to results of analysis, According to results of analysis, some factors that effect technically educational use of internet were defined. In consequence of rarefaction, linear relation between frequency of students’ internet use and computer use were defined. Knowledge of downloading a file affects many processes. Besides, results showed that chatting of class, method of learning use of internet for the first time, internet use for homework or research do not have intense interaction with other qualities. Bozkir et. al. (2009) also determined factor effecting success of students in the Student Selection Examination (OSS) of 2008 in Turkey through methods of data mining. According to findings of the study, five prominent factors effecting this type of score are age, type of school, interest in art courses, time spent for homework for math and usage rage of science lab if achievement score of secondary school is accepted as success criteria of all the students in the OSS. Results showed that age has a great effect on other types of scores. However, technical opportunities of schools have an effect on success in the OSS. Especially, it was emphasized that social, cultural and economic opportunities of students have considerable contribution to success in the OSS. Results also indicated that interest in art courses has a remarkable effect on achievement score of secondary school and effect of rate of having training is lower than expected. Koc and Karabatak (2011) showed effects of social networks upon students through association rule of methods.
of data mining. In the study, frequency of use of these networks and their effects upon student were analyzed.

Sample data examined in a study were previously evaluated through SPSS method to determine entrepreneurship of students and some demographic features effecting this tendency. (Erbasi, Unuvar and Unuvar, 2011). According to findings of this study, entrepreneurship average of students surveyed is 2.01. Furthermore, hypotheses measuring relation between tendency of entrepreneurship and some variables such as students’ gender, class, age, income level of their family were not subject to the stud, because there are not considerable differences between tendency of entrepreneurship and these variables. However, considerable differences between tendency of entrepreneurship and residence area, living in rural or urban area, frequency of following publications or any course-taking about entrepreneurship were observed. In this study, methods of data mining were also applied for same data.

2. MATERIAL AND METHOD

Research population is associate degree students. Samples of research consist of students of eight vocational high school of Selçuk University, Turkey’s biggest university. In those schools, some questionnaires were made with students who were at the school on the dates of 15-16 May 2010 and wanted to have questionnaire, in accordance with random sampling method. 1690 students filled out those questionnaire and questionnaires of 180 students were deemed as invalid for various reasons, thus questionnaires of 1510 students were subject to analysis.

Questionnaire consists of 2 sections. In the first section, 12 questions were asked to determine demographic feature and 40 questions were asked to determine tendency of entrepreneurship in the second section. Choosing questions of the second section, a questionnaire developed by Avsar (2007) benefiting from questions of the study of Summers (1998) and Brice (2002) was based on. Avsar (2007: 11-19) who investigates these factors in his study thoroughly, presents 8 main factors effecting the entrepreneurship. These factors: risk-taking, internal control, creativeness, innovativeness, tolerance of ambiguity, independence demand and relationship with people. In the sense of being easily understandable, different terms and concepts were used in some questions, but essence of factors representing each entrepreneurship element was not changed. Especially in groups in which calculated Cronbach Alpha values are low, questions were changed. A questionnaire with 52 questions in total is prepared adding questions relative to demographic attributes to the last group. Cronbach Alpha value of re-prepared questionnaire is calculated as 0.873. This result is considerably sufficient for the validity of analysis.

Each five questions representing one factor and questions about eight dimensions determining tendency of entrepreneurship in the second section of the questionnaire were asked according to five point likert scale. Acquired data was collected as raw data in the SPSS 18.0 application. This raw data was used through SPSS Clementine application in the methods of data mining. Decision tree classification analysis of methods of data mining was applied. In this application, decision tree algorithm was preferred due to its feature of producing graphic of dependency net showing effects of qualities upon each other. C5 decision tree algorithm was used for this analysis.

3. FINDINGS AND ASSESSMENTS

Some demographic features belonging survey participants are shown in Table 1.

The accuracy rates, general tendencies and tree-depth of eight factors constituting students entrepreneurship tendencies based on the answers given in survey are shown in Table 2.

**Factor 1. Tolerance of Ambiguity:** The accuracy rate, made with decision tree for tolerance of ambiguity factor is 74.79% and the general tendency is “I agree” with 83.70%
rate. But the tree-depth of this factor is zero. So factor for tolerance of ambiguity has no relation with demographic features.

**Factor 2. Need for Independence:** The accuracy rate of the analysis, made with decision tree for the need for independence factor is 71.78% and the general tendency is “I agree” with 77.15% rate. Tree-depth of the need for independence is seven. Decision tree of the need for independence factor that has level seven tree-depth is shown in Figure 1.

Important impressions of especially working of mother have been determined in the need for independence factor. Students whose mothers are not working accept to a large extent the need for independence factor. So the need for independence of those students, whose mothers are not working, is higher when compared with the ones whose mothers are working. Besides when the need for independence factor is taken as basis, significant relations determined among the ones who has working mothers in public sector and their siblings who cannot establish his/her business and the ones whose mothers are not working and who wants to establish a business in future.

**Factor 3. Risk Taking:** The accuracy rate of risk factor made with decision tree is 72.25% and the general tendency is “I agree” with 58.03% rate. Tree-depth for risk taking is level nine. Decision tree of risk taking that has level nine tree-depth is shown in Figure 2.

Important impressions of especially working of mother have been determined in the risk taking factor. Students whose mothers are not working accept to a large extent the risk taking factor. So the risk taking level of those students, whose mothers are not working, is higher when compared with the ones whose mothers are working. Besides when the risk taking factor is taken as basis, significant relations determined among the ones whose mothers are not working and the ones who want to establish their own businesses in future and among the ones, whose mothers are working in private sector ,and the ones who would try to establish their own businesses if they have 50.000 Turkish Liras. In addition to results of decision tree analysis relating to this factor, it is determined that the answers, given to questions of students of Graphic Designers department, about their father's field and working conditions, are inconsistent.

**Factor 4. Innovativeness:** The accuracy rate of innovativeness factor made with decision tree is 65.33% and the general tendency is “I agree” with 70.53% rate. Tree-depth for innovativeness factor is level five. Decision tree of innovativeness that has level five tree-depth is shown in Figure 3.

Important impressions have been determined in this factor about the thought establishing their own businesses in future. Students, who think about establishing their own businesses accepts to a large extent the innovativeness factor. So the harmony of students, who think about establishing their own businesses in future, to innovativeness factor is higher when compared the ones who do not.

**Factor 5. Need for Success:** The accuracy rate of analysis, made with decision tree for the need success factor, is 66.63% and the general tendency is “I agree” with 71.06% rate. Tree-depth for need for success factor, is level seven. Decision tree of need for success that has level seven tree-depth is shown in Figure 4.

Important impressions especially about working of mother have been determined in the need for success factor. Students whose mothers are not working accept to a large extent the need for success factor. It was observed that students, whose mothers are not working, have more need for success when compared with the ones whose mothers are working.

**Factor 6. Human Relations:** The accuracy rate of analysis, made with decision tree for human relation factor, is 70.89% and the general tendency is “I agree” with 77.88% rate. Tree-depth for human relations factor, is level four. Decision tree of human relations that has level four tree-depth is shown in Figure 5.
Important impressions about the living areas of students before they came for school, has been determined in human relations factor. It was determined, that students living especially in Mediterranean, Aegean, Southeastern Anatolia, Marmara region and Central Anatolia, accept to a large extent human relations factor. Entrepreneurship tendencies of students, living in these regions, concerning human relations, differentiate when compared the ones who lives in other regions.

**Factor 7. Internal Control:** The accuracy rate of analysis, made with decision tree for internal control factor, is 66.80% and the general tendency is “I agree” with 74.76% rate. But tree-depth level of this factor is zero. So no relation has been determined between internal control factor and any demographic feature.

**Factor 8. Creativeness:** The accuracy rate of analysis, made with decision tree for creativeness factor, is 70.36% and the general tendency is “I agree” with 78.74% rate. But tree-depth level of this factor is zero. So creativeness factor has no relation with any demographic feature.

4. **CONCLUSION**

In the study, demographic features, affecting entrepreneurship tendencies of students having education on associate degree in universities, has been determined and the ability extracting, potentially useful, applicable and meaningful information, is used for taking decisions from large volumes of data piles of data mining. A study, carried out on survey data with data mining to make these determinations, and unforeseen relations between entrepreneurship tendencies and demographic features tried to be revealed, through evaluating 1510 surveys.

According to the results of the study, tree-depth of tolerance of ambiguity, internal control and creativeness factors, among eight factors affecting entrepreneurship tendency, is zero. No relation has been determined between these factors and any demographic features of students. Decision tree of human relations factor, having level 4 tree-depth, has been formed. In this context, a significant relation has been determined between human relation factor and the regions students living before coming for school. Entrepreneurship tendency concerning human relations factor, is higher in students living in Mediterranean, Aegean, Southeastern Anatolia, Marmara Region and Central Anatolia when compared with students living in other regions. Decision tree of need for success factor, having level five tree-depth, has been formed. This factor is affected especially by field of mother. It was observed that students, whose mothers are not working, have more need for success when compared with the ones whose mothers are working. Decision tree of need for independence factor, having level seven tree-depth, has been formed. Important impressions of the working of mother have been determined in this factor. The need for independence of students, whose mothers are not working, is higher when compared with the ones having working mothers. Besides when the need for independence factor is taken as basis, significant relations determined among the ones who has working mothers in public sector and their siblings who cannot establish his/her business and the ones whose mothers are not working and who wants to establish a business in future. Decision tree of need for success factor, having level seven tree-depth, has been formed. This factor is affected especially by field of mother. It was observed that students, whose mothers are not working, have more need for success when compared with the ones whose mothers are working. Decision tree of innovativeness factor, having level nine tree-depth, has been formed. Important impressions have been determined in this factor about the thought establishing their own businesses in future. The harmony of students, thinking about establishing their own businesses in future, to innovativeness factor is higher when compared the ones who do not. Findings of study, provides important information to academic staff in courses related especially business, management and entrepreneurship.
REFERENCES


Luan, J. (2004). Data Mining Applications in Higher Education. SPSS Inc.


### Tables

<table>
<thead>
<tr>
<th>Variables</th>
<th>N=1510</th>
<th>%</th>
<th>Geographical Region</th>
<th>N=1510</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>968</td>
<td>64.1</td>
<td>Mediterranean</td>
<td>164</td>
<td>10.9</td>
</tr>
<tr>
<td>Female</td>
<td>542</td>
<td>35.9</td>
<td>Black Sea</td>
<td>80</td>
<td>5.3</td>
</tr>
<tr>
<td>Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>666</td>
<td>44.1</td>
<td>Marmara</td>
<td>183</td>
<td>12.1</td>
</tr>
<tr>
<td>1. Grade</td>
<td>752</td>
<td>49.8</td>
<td>Central Anatolia</td>
<td>880</td>
<td>58.3</td>
</tr>
<tr>
<td>Beklemeli</td>
<td>92</td>
<td>6.1</td>
<td>Eastern Anatolia</td>
<td>23</td>
<td>1.5</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-20</td>
<td>689</td>
<td>45.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-25</td>
<td>757</td>
<td>50.1</td>
<td>Yes</td>
<td>272</td>
<td>18.0</td>
</tr>
<tr>
<td>26 and Above</td>
<td>64</td>
<td>4.3</td>
<td>No</td>
<td>1238</td>
<td>82.0</td>
</tr>
</tbody>
</table>

| Number of Siblings | Which child of the family, in order |
Table 1. Information Related to Demographic Features of Students Attending the Survey

<table>
<thead>
<tr>
<th>Factor</th>
<th>Only Child</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8 and above</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>87</td>
<td>528</td>
<td>431</td>
<td>239</td>
<td>109</td>
<td>66</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>Accuracy Rate (%)</td>
<td>5.8</td>
<td>35.0</td>
<td>28.5</td>
<td>15.8</td>
<td>7.2</td>
<td>4.4</td>
<td>1.5</td>
<td>1.8</td>
</tr>
<tr>
<td>General Tendency (%)</td>
<td>572</td>
<td>495</td>
<td>207</td>
<td>106</td>
<td>74</td>
<td>33</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Tree-depth</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 2. Decision Tree Data of Entrepreneurship Tendency Factors

<table>
<thead>
<tr>
<th>Factor No</th>
<th>Factor</th>
<th>Accuracy Rate (%)</th>
<th>General Tendency (%)</th>
<th>Tree-depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tolerance of Ambiguity</td>
<td>74.79</td>
<td>83.70</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Need For Independence</td>
<td>71.78</td>
<td>77.15</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>Risk Taking</td>
<td>72.25</td>
<td>58.03</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Innovativeness</td>
<td>65.33</td>
<td>70.53</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>Need for Success</td>
<td>66.63</td>
<td>71.06</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>Human Relations</td>
<td>70.89</td>
<td>77.88</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Internal Control</td>
<td>66.80</td>
<td>74.76</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Creativeness</td>
<td>70.36</td>
<td>78.74</td>
<td>0</td>
</tr>
</tbody>
</table>
Figures

Figure 1. Decision Tree of the Need for Independence that has Level Seven Depth

Figure 2. Decision Tree of Risk Taking Factor that has Level Nine Tree-Depth

Figure 3. Decision Tree of Innovativeness Factor that has Level Five Tree-Depth
Figure 4. Decision Tree of Need for Success Factor that has Level Seven Tree-Depth

Figure 5: Decision Tree of Human Relations Factor that has Level Four Tree-Depth