International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 5 Issue 1, November-December 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

The Impact of Students' Inclinations on the Choice of Study Specialization using Bayesian Networks

Mohammed Ismail Alhussam[1], Yao Hongxing[1], Mohamad Ali Alhussam[2]

[1]School of Finance and Economics, Jiangsu University, Zhenjiang, China
[2]KIIT School of Management, Kalinga Institute of Industrial Technology, Bhubaneswar, Odish, India

ABSTRACT

In recent years, Universities had witnessed multiplicity and diversity in the fields of study specialization, through increasing the number of Specializations' years on one hand and creating faculties of new Specializations on the other hand. Our purpose is to investigate the students' behavior in terms of the choice of their specialization/major during their Bachelor studies. Through this research, we are trying to study the relationship between Students' Inclinations and academic achievements in the chosen department in the school of Economics at Aleppo University using Bayesian Networks. Our findings show that there is no robust relationship between the students' Inclinations and their educational achievements, which concluded that students are not choosing the major based on their knowledge skills. However, the students' scores in accounting subjects, before specialization, played the main role in their decision.

KEYWORD: *Students' Inclinations, Study Specialization, Bayesian Networks*

How to cite this paper: Mohammed Ismail Alhussam | Yao Hongxing | Mohamad Ali Alhussam "The Impact of Students' Inclinations on the Choice of Study Specialization using Bayesian Networks" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-1, December 2020, pp.1044-1049, URL: www.ijtsrd.com/papers/ijtsrd38172.pdf

IJTSRD38172

1. INTRODUCTION

In recent years, Universities had witnessed an increasing number of specialties and diversity of study fields, through increasing the years of specializations on one hand and creating schools of new specializations on the other hand.

The school of Economics is facing unequal distribution of students among the specializations. Most of the students are choosing some specific specializations like Banking, management, and accounting, while other specializations are left ignored, or even closed, because of the lacking and an inadequate number of students. Our aim is to investigate the students' behavior in terms of choice of their specialization during their Bachelor studies in the school of economics at Aleppo University. The importance of this research comes from providing decision-makers with information, to determine the impact of first and second-year scores on student's choice of specializations.

To understand, explain, and predict the behavior of humans and consumers in the workplace, we frequently apply probability theory to model the reality of the work. These models allow managers and analysts to run "what if" scenarios and manipulate variables in order to better utilize resources and influence on human behavior.

Bayesian networks are a widely-used class of probabilistic graphical models that used in various tasks for probabilistic reasoning and causal modeling. A Bayesian network consists of two components: a directed acyclic graph that expresses the conditional independence relations between random variables and conditional probability distributions associated with each variable. Nodes of the DAG correspond to variables, and edges express the dependencies between variables.

The structure of Bayesian network, that is, the DAG, can be easily visualized and may uncover some important characteristics of the domain, especially if the arcs are interpreted to be causal, or in other words, direct cause-effect relations and that depend on the approach of learning Bayesian network.

In the simplest case, Bayesian network is specified by an expert and then it is used to perform inference. In other applications, the task of defining the network is too complex for humans. In this case, the network structure and the parameters should be learned from data. There are two wide classes of algorithms for automatically learning the graph structure of a Bayesian network from data Score-Based Learning and Constraint-Based Learning.

In this study, we used Bayesian Networks to analyze the relationship between students' Inclinations and the Choice of Study specialization, by using the genetic algorithm with Bayesian score as a method to build our network.

2. Literature Review

2.1. Students' Choice of Specializations

The focus of the marketization concept is student choice behavior. Regarding the student choice behavior, there is a growing research interest in how the students, as consumers, make their choices in higher education (Alexander et al., 2011). Extensive research has been conducted on students' decision making regarding their programs and universities (Moogan & Baron, 2003; Vrontis, Thrassou, & Melanthiou, 2007).

Many attributes play a role in the student decision making but some of them are more important than others. The variables that influence the student choices were divided into two categories (Moogan & Baron, 2003): The first category is about the universities` and programs` characteristics. The second category groups variables regarding the influencers of choice such as the sources of information and the influences of the students' decision making(Moogan & Baron, 2003). Some of the most important attributes that influence the students` decision making seem to be: the personal interest in the program, the labor market and the location(Van Deuren & Santema, 2012). It seems that the most important characteristic that is taken into account in the decision-making is the personal interest in the program (Owen & Jensen, 2004). Also, another issue reported by the students as being very important in the selection process is the personal interest in the subject taught(Lapan, Shaughnessy, & Boggs, 1996; Maringe, 2006; Noble Calkins & Welki, 2006). In the second place, considering importance, seem to be the variables related to the labor market: expected earnings, employability, career opportunities(Van Deuren & Santema, 2012), variables that were found as being important in many consumer behavior researches(Malgwi, Howe, & Burnaby, 2005; Maringe, 2006; Noble Calkins & Welki, 2006). Another important aspect seems to be the location of the university (Van Deuren & Santema, 2012). Other important attributes for the students decision making regarding a major or a bachelor are the following: reputation (Maringe, 2006; Moogan & Baron, 2003; Moogan, Baron, & Bainbridge, 2001; Van Deuren & Santema, 2012), educational characteristics like the study materials, practical assignments(Owen & Jensen, 2004). According to(Worthington & Higgs, 2004), students' decisions are based on these two main factors: the location of an institution and the reputation of a course.

In this paper, we have taken the students' scores for all the subjects in the first and the second academic year as groups, according to the fields of study then we studied their effect on the choice of specializations. Moreover, we have tried to discover the other hidden factors that affect the students' decision.

2.2. Bayesian networks

The Network is a diagram, consisting of a set of nodes or nodes vertices, and a set of edges that bind the nodes. If all the edges in the chart are oriented, we get a vector chart Directed Graph (Greenland & Pearl, 2014).

The name of Bayesian networks driven from the conditional probability rule, which is known as Bayes rule:

$$p(h|e) = \frac{pr(e|h) * pr(h)}{pr(e)} \quad (1)$$

In another word:

$$Posterior = \frac{Likelihood * Prior}{p(evidence)} \quad (2)$$

Bayesian networks are defined as a directed acyclic graph or DAG, where each node is conditionally independent of its non-descendants, given its parents (Markov condition) (Bojduj, 2009).

In general, the Bayesian network consists of the following components:

- A directed acyclic graph G = G (V, E) where V represents the set of nodes and E is the set of edges in the graph G.
- (Ω, p) where Ω represents the sample space and p the probabilities associated with each event.
- $V = \{V_1, V_2, \ldots, V_n\}$ A set of random variables, discrete or continuous and observable random variables associated with the graph nodes.

The random variables distributed in probabilities space (Ω, p) where:

$$\boldsymbol{p}(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} p\big(V_i \mid \mathrm{pa}(V_i)\big) \quad (3)$$

Where: $\mathrm{pa}(V_i)$is the parents of the node V_i.

Bayesian networks differ from Markov networks in that Markov networks have undirected edges. In Markov networks, we cannot distinguish between the cause and the effect, unlike the Bayesian networks in which the directed edges determine the affecting and the resulted nodes, so sometimes it is called as "Causal Bayesian Networks". The network can be represented as follows: (*V, G, P(Vi |Pa(Vi))*)(Pearl, 2009)

The basic property of Bayesian Networks is that each Edge Pa(Vi)→Vi represent causal relation inanition represent Conditional probability values: $\boldsymbol{p}(V_i \mid \mathrm{pa}(V_i))$ (Lauritzen, 2001).

In Bayesian networks, the edge shows the direct effect of the father variable on the son variable. Thus, the model assumes that there are no intermediate variables, between the father variable and the son variable, which mediate this effect.

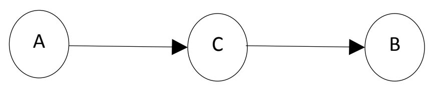
The Conditional Independence:

If we havethree nodes A, B, and C, we can say that node A and B are independent given C, for C if and only if our knowledge about A does not change when our knowledge about B changed and the value of C is known. That mean: P(A|C,B)=P(A|C). In addition, we can compute the joint probability as follows:

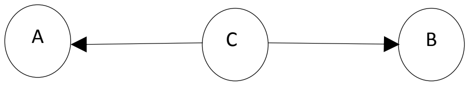
P(A,C,B)=P(C) P(B|C) P(A|B,C) = P(C) P(B|C) P(A|C) (4)

Here we have three cases:

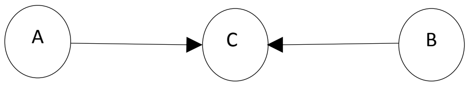
1. A is conditionally independent of B given C: P(B|C,A)=P(B|C)



2. A is conditionally independent of B given C
P(B|C,A)=P(B|C), P(A|C,B)=P(A|C)



3. A is conditionally dependent of B given C
P(B|C,A) ≠P(B|C)



3. Data and Variables

In this study, we investigate the students' behavior in terms of choice of their specialization during their Bachelor studies in the Faculty of Economics at Aleppo University. When the students passed the first and second years of study, they have to choose one specialization to continue the study in it. Our dataset contains 1740 record about the students in the Faculty of Economics at Aleppo University

Variables:
The variables represent student score in every subject of the first and the second academic year, and we have classified these subjects into groups, as follows:
1. Economics.
2. Accounting.
3. Management.
4. Management Information Systems MIS.
5. Statistics.
6. English language.

The summarize of the variables was as “Table1”.

Table 1 the variables of the study

Variable	Meaning	Values of variables
Economics	The scores that student achieved in every group of subjects in the first and second year before they choose the specialization	Not passed(Fail) Acceptable Good Very good Excellent Honors
Accounting		
Management		
MIS		
Statistics		
English		
graduation	Student's graduation average in the department that he chose	
Department	The department that is chosen	

We note that most variables represent the main subjects' scores in the first and second year, however, "graduation" represents the average score of the student through the third and fourth years for all courses.

With regard to the Department variable, we can see summarize of Department numbers in “Table2.”.

Table 2 Summarize of Department

Department No	Department Name	Graduated	Not Graduated	Total
0	Economics	9	2	11
1	Accounting	308	107	415
2	Management	303	98	401
3	Banking	578	157	735
4	Marketing	46	9	55
5	Statistics and Information Systems	104	30	134
Sum		1348	403	1751

4. Methodology

Learning the Structure Bayesian Networks

It is meant to discover the causal relationships between the variables, i.e., to identify the edges between the nodes, and to direct them. The general principle is that we draw an arc between the nodes that have causal relations. Generally, there are two main ways to teach network structure (Margaritis, 2003):

1. A score and search approach through the space of Bayesian network structures.
For the data set D and network B and P (B) the prior distribution of network B, the posterior probability of network B is calculated as follows (Cooper & Herskovits, 1992; Yang & Chang, 2002):

$$P(B|D) = \frac{P(B,D)}{\sum_{B'} P(B',D)} = \frac{P(B)P(D|B)}{\sum_{B'} P(B')P(D|B')} \qquad (5)$$

2. A constraint-based approach that uses conditional independencies identified in the data.
When we test the independence between two variables, it is known that the $\chi 2$value is calculated as follows(Spirtes, Glymour, & Scheines, 2000):

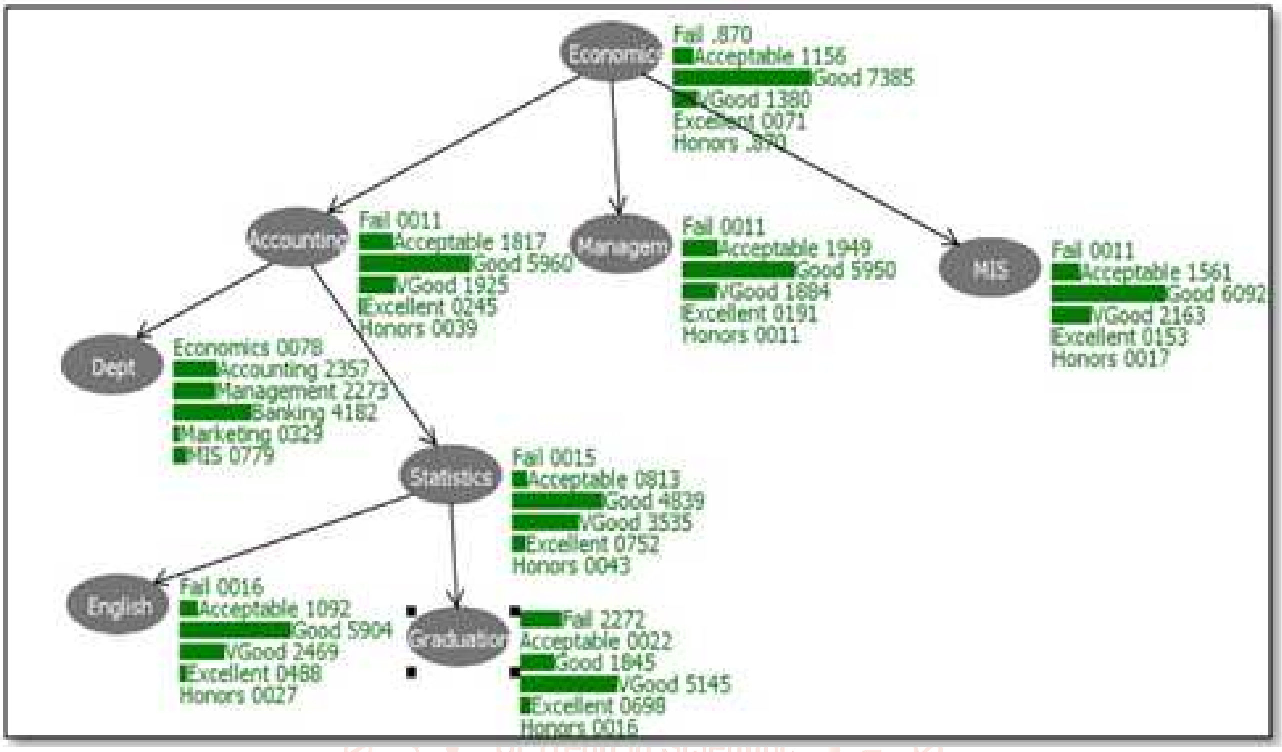
$$\chi^2 = \sum_i^{r_Y} \sum_j^{r_X} \frac{\left((O)_{ij}-E_{ij}\right)^2}{E_{ij}} \qquad (6)$$

Because the searching for the best structure of a network is a complex process of class N-P Hard(Robinson, 1977). We often use a genetic algorithm to search for a network structure (Perry, 2003).

5. Results and Decision

We learned a Bayesian Network structure by using our data and Weka software. We chose a genetic algorithm with Bayesian score as a method to build our networks, and then we get the network as "Fig.1".

Fig 1 Student scores and Specializations network



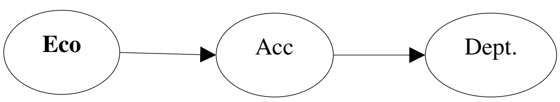
In addition, we can getmore information about the Strength of relations between Bayesian Networks nodes as "Table3".

Table 3 Strength of relations between Bayesian Networks nodes

X	Y	Mutual Information	Symmetric	For X Given Y	For Y Given X	Entropy X	Entropy Y
Accounting	Department	0.0258	0.0211	0.0245	0.0185	1.0405	1.3818
Statistics	graduation	0.0411	0.0356	0.0363	0.0349	1.1324	1.1789
Accounting	Statistics	0.0383	0.0352	0.0368	0.0338	1.0405	1.1324
Statistics	English	0.0278	0.0255	0.0245	0.0265	1.1324	1.0502
Economics	Accounting	0.0274	0.0301	0.0351	0.0263	0.7808	1.0405
Economics	Management	0.0459	0.0511	0.0588	0.0452	0.7808	1.0152
Economics	MIS	0.0279	0.0315	0.0357	0.0282	0.7808	0.9882

From the "Fig.1" and "Table3", we can read the results as follows:

First, the most important finding is that students' scores in Accounting before specialization played the main role in students' decision of choosing their department with strength relation.



That means if we know the value of Accounting score for a student, then just this value will share in determining the value of Department:

P (Dept. |Acc,Eco)= P (Dept. |Acc).

Second, students' score in Math and Statistics before specialization (x7) can help us to predict the student's average in the department that he/she chose (x12). Third, students' score in Economics x1 is the most independent variable from the other variables.

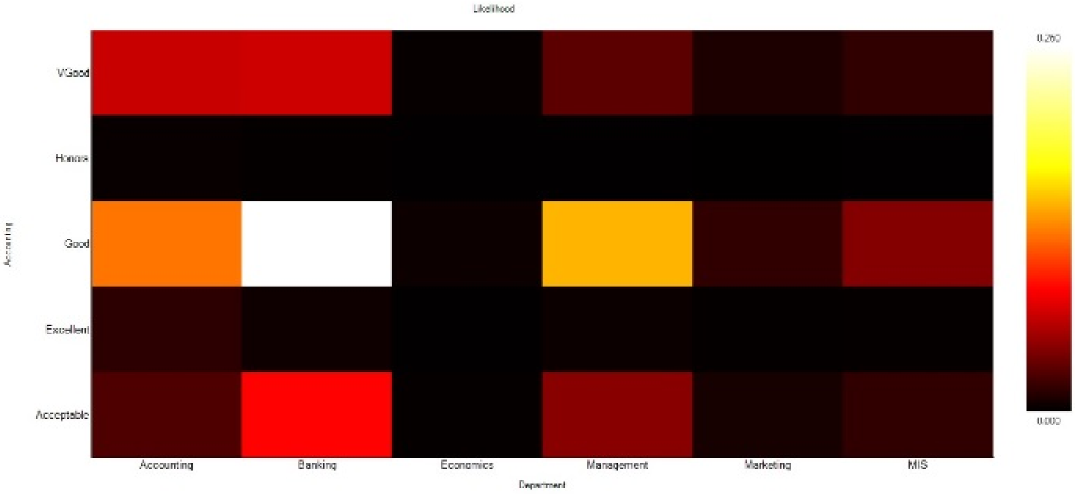


Fig 2 Mesh query[1] of Accounting variable and department variable

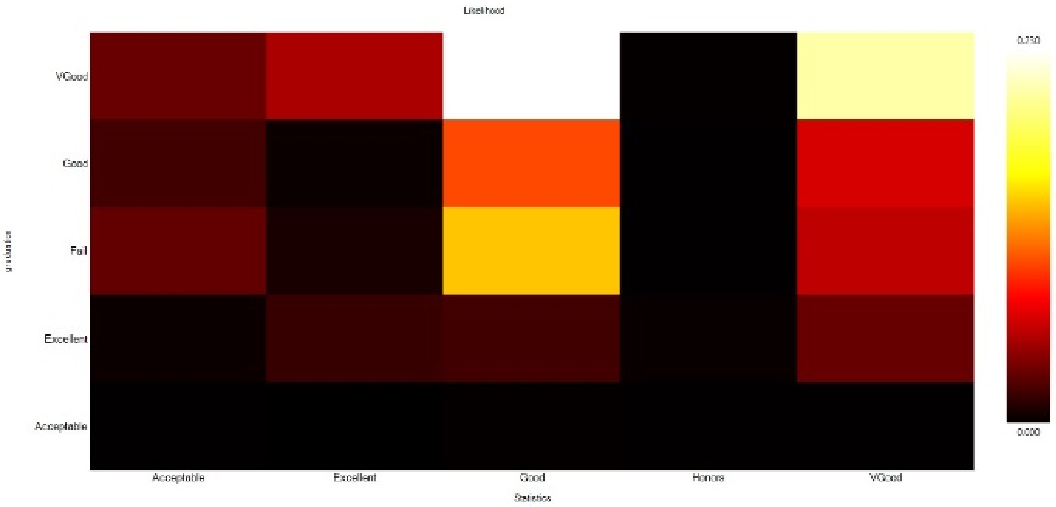


Fig 3 Mesh query of Statistics variable and graduation variable

In addition, if we run the inference by using software we can see:

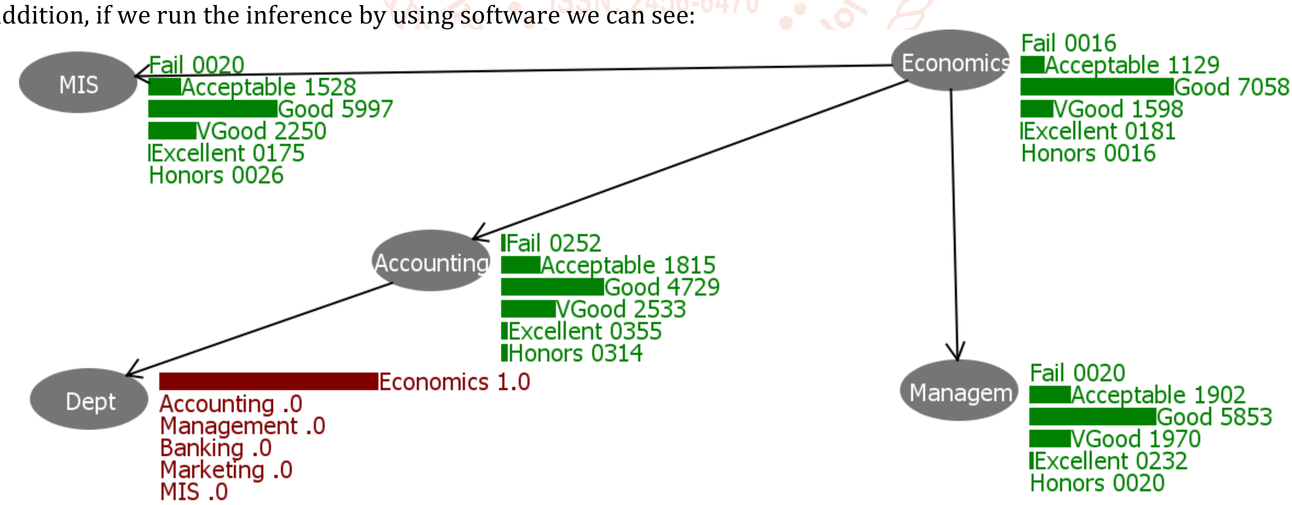


Fig 4 An example of inference

We can read "Fig.4" as the following: the students who chose Economics as department, 15.9, 25.3, 19.7 and 22.5 of them achieved a Very Good score in Economics, Accounting, management and MIS subjects respectively. That shows that students did not choose the economy based on their skills.

In this research, we try to focus on how the students' skills can release the innovation power. For example, in "Fig.5" we can see that when the students chose the Economics department depending on their skills then the percent of the excellent average of graduation will increase from 7.3% to 14.4% and the percent of failing will decrease from 22% to 17.5%. The Bayesian Network could show that, although of exist many latent variables unobserved.

---

[1] Mesh query generates a two dimensional graph, by repeatedly querying a network, while varying the values of two variables.

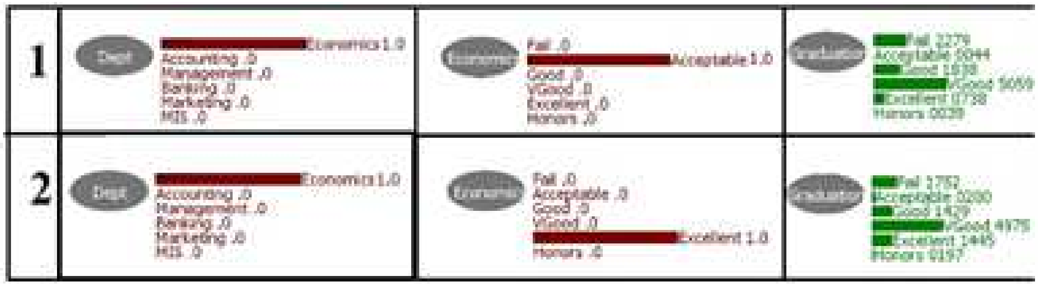


Fig 5 an example of inference

6. Conclusions

We note from the results, that students' decisions did not depend on their orientation, but the greater role was for their accounting skills in selecting their majors. when we know that the job opportunities for accounting students are more than other students then we will know that these results indicate that market requirements significantly affect students' decisions.

Therefore, we recommend that the decision-maker at the faculty of economics organize workshops in order to draw the students' attention to the importance of following their interests, developing their knowledge skills and improving their talents to reach to more creativity and innovation.

References

[1] Alexander, P. M., Holmner, M., Lotriet, H. H., Matthee, M. C., Pieterse, H., Naidoo, S., . . . Jordaan, D. (2011). Factors affecting career choice: Comparison between students from computer and other disciplines. *Journal of Science Education and Technology, 20*(3), 300-315.

[2] Bojduj, B. N. (2009). Extraction of causal-association networks from unstructured text data. *Master's Theses and Project Reports*, 138.

[3] Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine learning, 9*(4), 309-347.

[4] Greenland, S., & Pearl, J. (2014). Causal diagrams. *Wiley StatsRef: Statistics Reference Online*, 1-10.

[5] Lapan, R. T., Shaughnessy, P., & Boggs, K. (1996). Efficacy expectations and vocational interests as mediators between sex and choice of math/science college majors: A longitudinal study. *Journal of Vocational Behavior, 49*(3), 277-291.

[6] Lauritzen, S. L. (2001). Causal inference from graphical models. *Complex stochastic systems*, 63-107.

[7] Malgwi, C. A., Howe, M. A., & Burnaby, P. A. (2005). Influences on students' choice of college major. *Journal of Education for Business, 80*(5), 275-282.

[8] Margaritis, D. (2003). Learning Bayesian network model structure from data: Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

[9] Maringe, F. (2006). University and course choice: Implications for positioning, recruitment and marketing. *International Journal of Educational Management, 20*(6), 466-479.

[10] Moogan, Y. J., & Baron, S. (2003). An analysis of student characteristics within the student decision making process. *Journal of further and Higher Education, 27*(3), 271-287.

[11] Moogan, Y. J., Baron, S., & Bainbridge, S. (2001). Timings and trade-offs in the marketing of higher education courses: a conjoint approach. *Marketing Intelligence & Planning, 19*(3), 179-187.

[12] Noble Calkins, L., & Welki, A. (2006). Factors that influence choice of major: why some students never consider economics. *International Journal of Social Economics, 33*(8), 547-564.

[13] Owen, A. L., & Jensen, E. (2004). Learning About Learning: Students' Course Choice.

[14] Pearl, J. (2009). *Causality*: Cambridge university press.

[15] Perry, B. B. (2003). *A genetic algorithm for learning Bayesian network adjacency matrices from data.* Kansas State University.

[16] Robinson, R. W. (1977). Counting unlabeled acyclic digraphs *Combinatorial mathematics V* (pp. 28-43): Springer.

[17] Spirtes, P., Glymour, C., & Scheines, R. (2000). Causation, prediction, and search 2nd ed: Cambridge, MA: MIT Press.

[18] Van Deuren, R., & Santema, S. (2012). How to choose your minor. *Decision making variables used in the selection of a minor by undergraduate students from a Dutch university of applied sciences. Working papers, 6*.

[19] Vrontis, D., Thrassou, A., & Melanthiou, Y. (2007). A contemporary higher education student-choice model for developed countries. *Journal of Business Research, 60*(9), 979-989.

[20] Worthington, A., & Higgs, H. (2004). Factors explaining the choice of an economics major: the role of student characteristics, personality and perceptions of the profession. *International Journal of Social Economics, 31*(5/6), 593-613.

[21] Yang, S., & Chang, K.-C. (2002). Comparison of score metrics for Bayesian network learning. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 32*(3), 419-428.