A Bayesian approach in students' performance analysis

Conference Paper · July 2018

DOI: 10.21125/edulearn.2018.2498

2 authors:

Azarakhsh Hamedi
Tampere University

Amir Dirin
Haaga-Helia ammattikorkeakoulu

A Bayesian approach in students' performance analysis

Azarakhsh Hamedi[1], Amir Dirin[2]

[1]Tampere University of Technology, Tampere, Finland
[2]Haaga-Helia University of Applied Sciences, Helsinki, Finland;
[1]azarakhsh.hamedi@tut.fi
[2]amir.dirin@haaga-helia.fi

Abstract

Psychological and environmental aspects have a great influence on students' performance in higher education. Analysing and modelling the performance of students helps educational institutes to improve the quality of their course offering. Furthermore, it helps students to perform better in their studies and therefore has a better competence. Bayesian networks can model the causal interactions and statistical relationships of a system's variables with a graphical demonstration, which is easy to interpret. Using such a model and having evidence about one or many of performance indicators of students, it is possible to investigate the status of other indicators in the model. It is also possible to predict the intervention impact on one or more indicators on the other parts of the network. Development of machine learning techniques for Bayesian networks in the recent years makes it possible to discover the knowledge of a domain automatically using the collected data. In this study, the Bayesian networks are used to model the causal relation of student's performance factors and then the model is used to classify the students according to performance and explore the effect of the intervention on each individual student. The results indicate that by reducing the fear-of-failure factor by 30 percent will impact the overall students' performance and reduce the academic withdrawal factor.

Keywords: Bayesian Approach, Data Analysis, Interference

1 INTRODUCTION

Students' performance at the educational institutes depends on many factors. These factors are classified as an internal, which is based on individual psychology and external factor such as educational environment, family setups, economy, and other contextual factors. In addition to all of these factors, economic recession, educational financial challenges and national policy also affect to educational performance. Many of these factors if not all have experienced by students during last five years in Finland. These factors are often the core reason for students drop out from the universities. Awareness of these factors by degree program and educational institutes help to anticipate and impact the overall students' performance. In other words, to predict and prevent discontinuation, behavioural engagement, and academic performance help to reduce the academic withdrawal rates.

This study is based on gathered questionnaire from BITE degree program students. The questionnaire is based on Niemivirta [1] study. For analysing the data, we applied various statistical analysis, pattern recognition, and machine-learning algorithm. We elaborate how a degree program may anticipate drop out by affecting to some key dropout factors such as fear-of-failure and academic withdrawal. In this paper we only cover the analysis based on Bayesian data analysis [2] by using Bayesian Lab software

This study is based on the following eight scales on motivational factors adapted from [1] .

1- Mastery-intrinsic orientation, which measures students' inner motivation and interest in studying.

2- Mastery-extrinsic orientation, which measures students' motivation to perform well in their studies.

3- Performance-intrinsic orientation, which measures students' tendency to outperform their peers.

4- Performance-extrinsic orientation, which measures students' tendency to avoid making mistakes or fail in the course.

5- Avoidance Orientation, which measures students' efforts regarding school work (make as little effort as possible as for school work)

6- Academic Withdrawal, which measures students' tendency to give up studies or to dropout from school.

7- Fear of failure, which measures students' fear towards academic performance.

8- School value, which measures how students appreciate and value academic institutes.

A Bayesian Network (BN) is a graphical model, which is representing the qualitative and quantitative relationship between variables of a system. The qualitative part is represented using Directed Acyclic Graphs (DAGs), in which nodes represent variables and arcs are showing the dependencies between variables. The qualitative part is a compact representation of the joint probability distributions, which is factorized based on the dependency relations depicted in the DAG [3]. Bayesian networks allow us to use the information from a subset of variables (nodes) in the system (network) to make rational decisions about any other subset of variables in the system [4].

1.1 Missing values

All the real life datasets have missing values to some extent. The reason for that can be a failure in the recording system, human mistakes, errors in sensors or a non-response in a survey. Missing values can be classified in four classes. *Missing Completely at Random* (MCAR) is the case that missing mechanism of the variable is totally independent of other variables in the system. The second class is called *Missing at Random* (MAR) in which the missing values are dependent on other observed variables. The next class is called *Missing Not at Random* (MNAR) or *Not Missing at Random* (NMAR) and it is referred to missing mechanisms, which are dependent on an unobserved variable. The last class of missing values can be called *Filtered* values which are not missing in reality. These are the values, which their existence is depending on the values of other variables [5].

There are many approaches to handle the missing values. With datasets with large amounts of data, it is possible to remove the records with the missing data. This approach is also called *likewise deletion* or *case-wise deletion*. Koller and Friedman [6] had shown that this approach can potentially change the distribution of the variables and lead to magnificent amount of bias in the dataset. The other approach to process the missing values is to replacing them with some fixed values. The replacing value can be coming from an expert's knowledge or, for example, can be the mean or median of the variable with missing value. The third approach can be replacing the missing values with some inferring method. Static imputation is the first method in this group, in which, the missing values are replaced with a random draw from the non-missing vales of the same variable. The resulting distribution will be similar to likewise deletion, but with no decrees in the amount of data and not additional bias. The second method in this group is structural expectation Maximization method. This method replace the missing variables according to the available values and relations between them.

1.2 Expectation Maximization method

In the case of datasets with some missing values, Expectation Maximization (EM) is a method to replace the missing values using the available values and the parameters calculated using them. This method consists of two steps. The first step uses a maximum likelihood estimator to find an expected value for the missing value, based on the parameters calculated from the other incomplete data. Since our data is in discreet form, the counting technique can be used to estimate the parameters. The estimator replaces the missing values with a number in the range between 0 and 1. Then, in the second step, a new set of parameters are calculated using the real data and the estimated values. At this point one iteration of EM algorithm is finished. Using the new parameters, a new set of likelihoods will be calculated for the missing values and the procedure will continued. In each iteration, the likelihood value for the network will be compared with the previous iteration, and if it does not change, it means that the algorithm reached to a local maximum. The method guarantees reaching to a local maximum [6].

1.3 Information theory based metric

Entropy is a formal quantification of uncertainty [7]. It shows how even the probability distribution of a random variable is. In other words, entropy is the measure of information one can get, in average, from distributed variable in a domain. One of the interpretations of entropy can be calculated using Shannon's [8] formula: $H(X) = -\sum_i p(x_i) \ln p(x_i)$. This formula calculates the number of bits needed to describe the random variable X. Observation of other predictive random variables can increase the amount of information and consequently the entropy value. The entropy of a random variable given the observations of another random variable is called conditional entropy. The difference between the marginal entropy of a variable and the conditional entropy of the same variable given the observations of another random variable is called entropy reduction or mutual information. Mutual information can show us what will be the benefit of observing a particular random variable in predicting the variable of choice. In this way, we can find out which variable has the most predictive importance. Mutual information in a non-negative and symmetric value. According to Yao [7], conditional entropy and mutual information can be used to determine one way associations between variables.

1.4 Learning the structure from the data, the EQ Framework

Finding the best network in the search space of all possible network is NP-hard, and the heuristic search algorithms can easily trap in local minima. To tackle this problem, Munteanu and Bendou [4] developed the EQ framework to use the space of essential graphs [9] of an equivalent class [10] of BNs to search for a suitable graph. The EQ framework proposes a local scoring scheme. In each essential graph, transformation algorithms create non-empty subsets of graphs, called *instantiable graphs*, by making small changes in the structures in the essential graphs set. A small transformations means suppression or addition of only one single edge. On the other hand, the score of a Bayesian network can be decomposed as the sum of the local scores calculated from a node and its parents. Using this fact, they calculated the change in the score of the whole BN after performing the small transformation. The procedure of learning the structure in this framework is as follows:

1. Setting constraints in order to avoid making non-instantiable graphs from essential graphs in the transformation process and creating rules according to this constraints
2. Making operators for creating new instantiable graphs by adding or removing edges and V structures according to the rules and calculating the change in the local score for each of them
3. Creating the essential graph corresponding to the created instances and
4. Calculating the score for the found equivalence class and compare to others to find the best structure

1.5 Network selection and validation

The EQ framework uses a metric called Maximum Description Length (MDL) to finding the best Bayesian networks structure that describes the joint probability distribution. MDL is a two components value consisting of the number of bits required for representing a model and the number of bits to represent the data given that model. The best solution has the lowest value for MDL. The MDL value in this study is calculated as $\alpha DL(B) + DL(D|B)$. The minimum value for the first component is when we have a set of fully unconnected nodes. On the other hand, the minimum value for the second component is when all nodes in the network are connected, i.e. fully connected network. Then to minimize the sum, we have to find the best trade-off between these two which is controlled by the value of α [5].

The other measure to check the fitting of the Bayesian network model with the data set is to check the Contingency Table Fit (CTF) value. In case of Bayesian network, a network that is fully connected, i.e. none of the conditional independencies is considered in forming the network, is the best representation of the contingency table. On the other hand, an unconnected network assumes that there is no dependency between variables and is the worst representation of the contingency table. By comparing the resulting network for each structural coefficient with these two networks, we can have a measure of how good the network in describing the dataset. The CTF value then can be calculated by comparing the entropy of the current network with the entropy of a fully connected network and the entropy of an unconnected network [11].

1.6 Inference in Bayesian networks

Inference is to find the marginal probability distribution of a node, after performing changes in the distribution of the other nodes in the network. There are several methods to calculate this marginal probability, namely sum-product algorithm, junction three algorithm, etc. [6]. In the case of this study, we used the junction tree algorithm. This algorithm consists of seven steps [12] and starts with moralizing the network, meaning transforming the directed graph to an undirected graph. This is done by adding an undirected edge between the parent nodes and then changing directed edges to undirected edges by adding an edge in the reverse direction. The second step is to triangulating the graph. This means for any cycle in the undirected graph, if the number of nodes are bigger that three, we should add an edge to a pair of non-consecutive nodes in that cycle, which is called a chord. Now, in the third step, we can form the junction tree. Junction trees are tree graphs created using a hypergraph formed from the cliques of the triangulated graph in last step. A hypergraph is a set of all nonempty subsets of a graph. A junction tree should have another property which is called running intersection property or junction property, which is that the intersection of any two nodes in a path in a junction tree should be contained in every node in that path. In the fourth step, the conditional probability distribution (CPT) tables are used to assign potentials for each clique; the potential is the joint probability distribution of that clique. Since the main reason for forming the junction trees is to apply a message-passing algorithm, in the fifth step, the algorithm defines a root node to start the message passing procedure from it. After setting the root node, the algorithm uses one of the message passing algorithms in graphical models to pass the changes in the nodes all the way to the leaf node and backward. Therefore, the message-passing step, step six, consists of two messages and the junction tree guarantees the convergence of the algorithm. In the seventh step, we use the result of the last step, which is the modified joint distribution of each clique to calculate the marginalized distribution of node of choice.

2 RESEARCH QUESTIONS AND RESEARCH METHODOLOGY

The main objective of this research is to investigate the Business Information Technology (BITE) students' motivational factors. This information helps the department to plan a proper strategy to anticipate educational withdrawal and increase students' educational performance and reduce the dropout rates. To achieve the above objectives we pursuit to answering the following question.

How changing in a fear-of-failure can make an influence on the other variables such as performance and academic withdrawal?

The data gathered by two methods also defined as qualitative data and quantitative data. Quantitative research quantifies opinions, behaviours and other defined variables. The data gathered by quantitative approach is analysed using parametric tests, while qualitative data does not. The distribution of parametric tests is powerful than nonparametric test, however, nonparametric test is much more flexible. [13]

As one of the data analysis techniques, data mining is the process to structure the raw data and formulate or recognize the various patterns in the data through the mathematical and computational algorithms. Data mining helps to generate new information and unlock the various insights. To analyse the gathered data, we applied inference in Bayesian networks.

3 DESIGN

This study is based on the information gathered from BITE degree Programme students during Autumn 2017 using a questionnaire. The data are collected by handing over the questionnaires to 100 BITE degree program students in various semesters. The questionnaire contains thirty different statements. Each statement has a 1-7 point Likert-scale where students are asked to choose how much they agree or disagree with that statement. The questionnaire is designed to collect information about students’ motivation factors. In order to measure the relations between different factors, we defined eight clusters, each of which consisting of three related statements.

The questionnaire. The questionnaire is hand over as a hardcopy and students are given 10 minutes to fill-in the form. The data is then converted to excel sheet for further analysis. We applied various statistical analyses, pattern recognition, and a machine-learning algorithms to predict the future trend. In this paper, however, we only present the Bayesian analysis of the collected data.

4 IMPLEMENTATION

To form the clusters and find the dependencies between clusters and creating the Bayesian network, we used Bayesialab software from Bayesia S.A.S [14]. The workflow for this process is as follows:

The process starts with importing the dataset created for the purpose of statistical analysis. The Bayesialab's data import tool can import CSV format files and visualize the variables and values. As some of the participants were reluctant to answer a question hence some values are missing in the dataset. Therefore, we have to use one of the missing value processing methods, which is presented above, to replace them with relevant values. Since the missing mechanism is unknown, the Structural Expectation Maximization method has been chosen for processing the missing values. In the modelling phase, we represent the data for each motivation factor, with a cluster created out of the three corresponding statements. To form the clusters, we put the variables of each cluster into a class and used variable clustering tool in the software to form a latent variable that represents each cluster. The software created a latent variable for each class and set it as the central node for of a naïve Bayesian network. Then it finds estimates of the parameters of that latent parameters using expectation maximization method.

Before using the unsupervised machine-learning algorithm, we used the experts' knowledge of the domain, to set constraint for the learning algorithm and prevent finding irrational causal relations in the network. We grouped the clusters to internal and external motivational factors. According to the literature, clusters like Fear of Failure, Performance-approach orientation, Performance-avoidance orientation, and School value can be considered as external factors. Conversely, Academic withdrawal, Avoidance orientation, Mastery-extrinsic orientation, Mastery-intrinsic orientation are internal factors. A rational assumption can be that internal factors cannot influence external factors, but the external factors can influence the internal factors. Therefore, by forbidding directed edges from all internal factors towards external factors, we can avoid finding irrational relations in the network. Furthermore, we forbid any direct connections between questionnaire variables, so the algorithm find only the relation between latent variable representing the clusters.

In the next step, we used structural coefficient (SC) analysis tool to find the best structural coefficient for learning the structure. In figure 1 we ran the tool for 100 structural coefficient values between (n=1 and N=0.01) and set the tool to calculate the CTF value for each network.

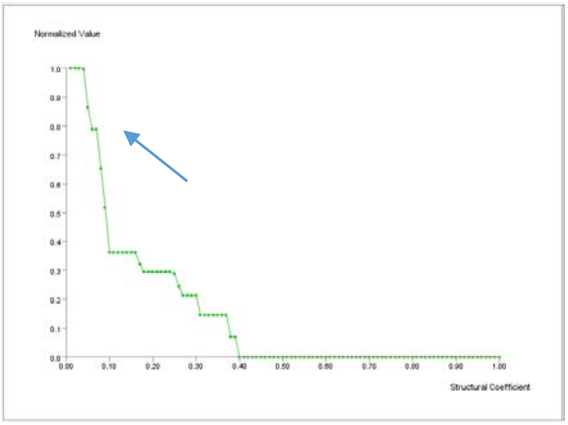

Figure 1 - Structural analysis curve

The Y-axis of the graph is the normalized CTF value of a network corresponding a SC value on the X-axis. As shown in the Figure 1 with the blue arrow, the contingency table fit value reaches to 76.6% for the SC value 0.07, which shows that the resulting network is a good fit for the joint probability distribution of our database.

Using the SC value found in the previous step, we use the EQ unsupervised structural learning algorithm to find the desirable Bayesian network structure and parameters. The resulting network is shown in the Figure 2.

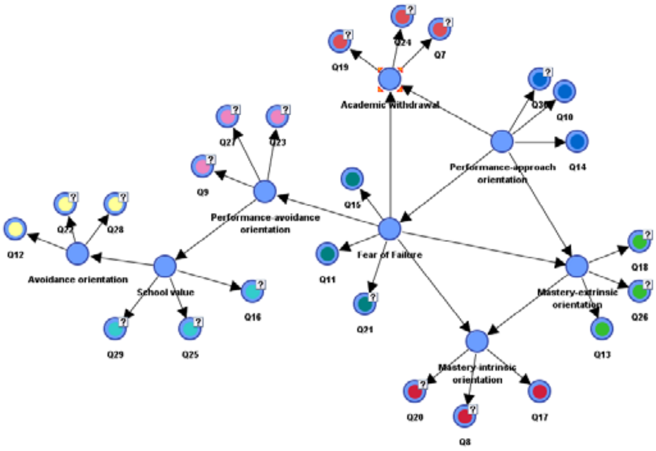

Figure 2 - The Bayesian network structure

To evaluate the Bayesian network model found in the previous step, we used the Multi Target Evaluation tool in the Bayesialab software. The tool assumes each of the nodes in the network as a target node and then calculate several metrics such as precision values for that node. The precision value calculates the percentage of correct prediction of the values for each variable by the model. The table 1 shows the results of the report.

Table 1 - Precision values

Mean Precision	
Mean	64.14%
Standard Deviation	22.94%
Minimum	26.84%
Maximum	100%

Table 1 shows the mean precision value is the unweighted average of normalized number of correct prediction by the model. This shows that the model can predict the values of dataset correctly in 64.14% of the times in average.

After forming and validating the Bayesian model, we tried to investigate consequences of intervention on some of the motivational factors on the others. Among the motivational factors, we decided to focus on the fear of failure and check the effects of this intervention on academic withdrawal and other factors, which have a close association with academic withdrawal. The aim for doing this is to take some measures, e.g., by reducing fear of failure to increase students' academic performance and as a results impact on overall academic withdrawal.

In the first step, we found the four most effective factors_on the academic withdrawal based on the value of mutual information between them. These four most effective nodes on academic withdrawal are mastery intrinsic orientation, mastery extrinsic orientation and performance-approach orientation. Figure 3 demonstrates the Academic Withdrawal as the target node with the four most effective nodes in yellow.

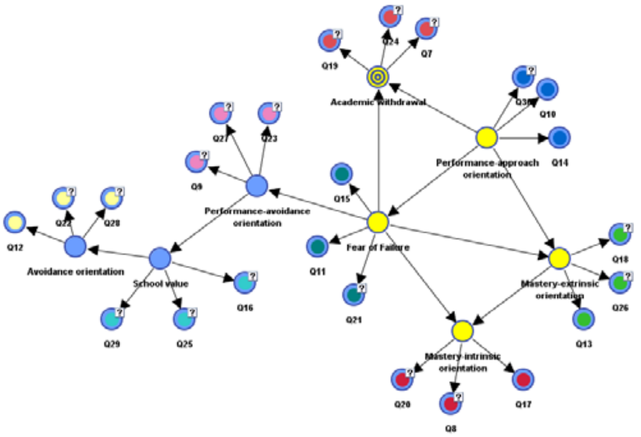

Figure 3 - The four most effective nodes for Academic withdrawal node

Then we can see the marginal probability distribution of each selected node in the figure 4. Each of the nodes have seven states, which is the result of data clustering step. The expected value, or the mean value, of each node is shown as "Value" in each box. The target node, academic withdrawal, is displayed in red.

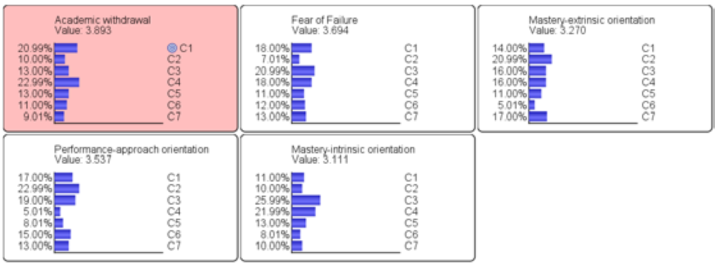

Figure 4 - Marginal Probability distribution of selected nodes

Then we used Pearl's [15] graph surgery concept to perform an intervention on the Fear of Failure factor. Concisely, Intervention in this case means that we want to set some exact value for a node, regardless of the effect of the other nodes on it. Therefore, we can remove all the incoming arcs from that node's parents. Using this, we try to predict the result of an intervention of the fear of failure factor by 30%, on academic withdrawal and the four other three other most effective nodes on it.

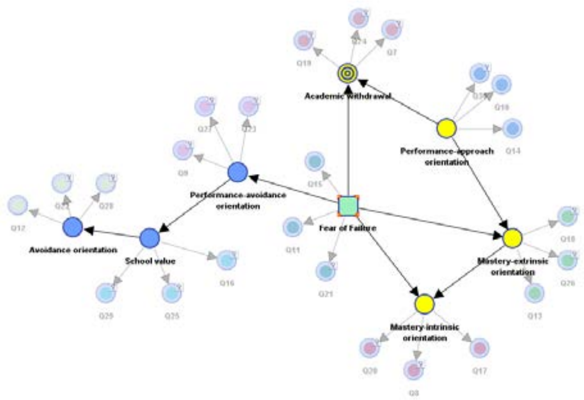

Figure 5 - the network after intervening in node Fear of Failure

5 RESULTS

To perform the intervention, we reduced the mean value of the fear of failure node by 1.094 units, which is equal to 30% of the mean value. To make the change we use the Minimum Cross Entropy (MinXent) [16] distribution estimation method to estimate the new probability after changing the mean value. After performing the inference, the results is as shown in the Figure 6.

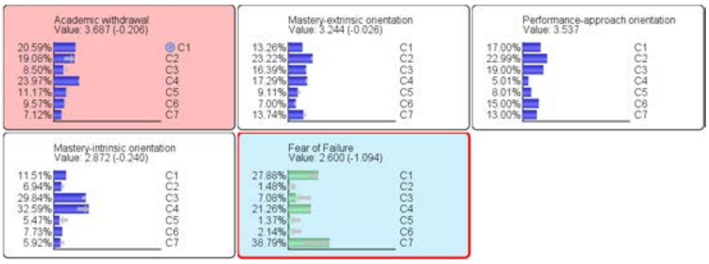

Figure 6 - Nodes marginal probability distributions after intervention

The result of inference shows that increasing in the mean value of Fear of failure factor, can lead to a decrease in the mean value of academic withdrawal by 5.3 percent. This can be interpreted as that the decrease in the fear of failure in students can directly affect the academic withdrawal factor. Based on this result, we have decided to develop methods in order to decrease the fear of failure in the students. Align with this, the mean value of Masteri-interinsic Orientation drops by 7.7 percent. According to the literature, this happens because the intrinsic factors have a direct relation with fear of failure, as shown in the Bayesian network with the direct arrow from fear of failure to mastery intrinsic node. This relation is in such a way that the students with lower intrinsic motivations may not have the tendency to outperform their peers, so they have lower anxiety to fail. This fact is clearly demonstrated by the decrease in mastery intrinsic orientation factor. On the other hand, the mean value of Mastery-extrinsic orientation dropped by only 0.7 percent. This shows that the fear of failure has a minor effect on the extrinsic factors.

Removing the arc from Performance-approach orientation to Fear of Failure caused separation of these to nodes. Therefore, after a changing the probability distribution of any of them, the inference algorithm will not change the marginal probability distribution of the other one.

6 DISCUSSIONS

Educational institutes pursue regularly to develop and improve their educational offerings. A typical approach to achieve optimal quality has been through statistical analysis of students' feedback on each course and quality assurance on each department. The data is often collected and analysed statistically to anticipate and address the existing problems. Unlike teachers who track individual student's educational performance. The educational institute process the data collectively and hence they tread all students collectively. Teachers on the other hand tend to improve their educational offering and course set-ups after each semester, therefore, the often collect and analysis individual students' performance

This main contribution of this paper lays on the following two main points

Individual student's performance in comparison to collective analysis. In the contemporary educational setups, the available technologies enable us to collect data about individual student performance factors. The data collection happens often in each course by teachers and also by school administrations. However, the analysis and the decision-making addressed towards all students collectively. Treating all students in the same manner is not the right approach specifically in multi-nationals and diverse social ethnics' classes. On the other hand, enforcing a collective decision would be effective if the decision impacts to everybody with the same degree of impact. For example, reducing the fear-of-failure factor may affect all students in various degrees, meaning those who have higher fear-of –failure would be affected more than those who do not suffer from fear-of-failure. The collective decisions enforce in all students in this case is more mental which aim to develop the performance of an individual. Student is an individual person, which has psychological and physiological needs and capacities. For example, learning capabilities, education contributions, motivations, and interest. Therefore, addressing a collective decision towards all student must be based on identification common factor, which in some degree impact all students

Statistical Analysis versus Machine Learning Algorithm

Technologies enable us to study the behaviour and performance of each student in various courses. Technologies such as data mining and machine learning not only helps to analysis individual behaviour but also enable us to identify independent and dependent factors, which may affect all students to some degrees.

This happens by measuring the degree in independence between the factors formed using the collected data from the students. Moreover, inference in Bayesian networks enables us to investigate the future trends, based on the model created using the historical data.

7 CONCLUSIONS

The main objective of this study was to investigate how to improve the students' performance by reducing the fear-of-failure. The rate of Fear-of-Failure among all students at BITE indicated (22%) in our statistical analysis as significant. Therefore, by this study we pursuit to see the trend how impacting the fear-of-factor as a major impact element towards motivational factors such … In theory as seen in figure XXX in our students increase in some degree (n=XXX%) of our students. Based on this study and the statistical analysis we aim to reduce the fear-of-failure through raising awareness about the degree program, study path, and study environment already before the new students (freshmen) start the school.

ACKNOWLEDGEMENTS

Special thanks to students' and thesis workers in Business Information Technology (Bite) at Haaga-Helia University of Applied Science for conducting the questionnaire.

REFERENCES

[1] M. NIEMIVIRTA, "MOTIVATION AND PERFORMANCE IN CONTEXT: THE INFLUENCE OF GOAL ORIENTATIONS AND INSTRUCTIONAL SETTING ON SITUATIONAL APPRAISALS AND TASK PERFORMANCE," *Psychol. -An Int. J. Psychol. Orient*, vol. 45, no. 4, pp. 250–270, 2002.

[2] A. . Fallis, *Baysian Data Analysis*, vol. 53, no. 9. 2013.

[3] Z. Ghahramani, "an Introduction To Hidden Markov Models and Bayesian Networks," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 15, no. 1, pp. 9–42, 2001.

[4] P. Munteanu and M. Bendou, "The EQ framework for learning equivalence classes of Bayesian networks," *Proc. 2001 IEEE Int. Conf. Data Min.*, pp. 417–424, 2001.

[5] S. Conrady and L. Jouffe, *Bayesian Networks and Bayesia Lab*, vol. 1, no. 1. 2007.

[6] D. Koller and N. Friedman, *Probabilistic Graphical Models*, vol. 53, no. 9. 2013.

[7] Y. Y. Yao, "Information-theoretic measures for knowledge discovery and data mining," in *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, 2003, pp. 115–136.

[8] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. July 1928, pp. 379–423, 1948.

[9] A. Garrido, "Essential graphs and bayesian networks," *Proc. - 2008 1st Int. Conf. Complex. Intell. Artif. Nat. Complex Syst. Med. Appl. Complex Syst. Biomed. Comput. CANS 2008*, pp. 149–156, 2008.

[10] T. Verma and J. Pearl, "Equivalence and synthesis of causal models," *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*. pp. 220–227, 1991.

[11] Bayesia, "Contigency Table Fit and Deviance Formulas," 2018. [Online]. Available: http://library.bayesia.com/display/FAQ/Contigency+Table+Fit+and+Deviance+Formulas. [Accessed: 22-Apr-2018].

[12] D. Kahle, T. Savitsky, S. Schnelle, and V. Cevher, "Junction Tree Algorithm," *Stat*, pp. 1–14, 2008.

[13] J. P. Verma, *Data analysis in management with SPSS software*. 2013.

[14] Bayesia-S.A.S, "BayesiaLab 7 - Bayesian Networks for Research and Analytics," 2018. [Online]. Available: http://www.bayesialab.com/. [Accessed: 15-Apr-2018].

[15] J. Pearl, "A Probabilistic Calculus of Actions," *Uncertain. Proc. 1994*, pp. 454–462, 1994.

[16] A. SHAMILOV, Ç. GRIFTINOGLU, I. USTA, and Y. M. KANTAR, “Determining probability distribution by minimum cross entropy method Determining Probability Distribution by Minimum Cross Entropy Method,” no. September, 2006.