The Mathematical Model for Ranking Students of Online IT Courses

Ivanna Dronyuk [1][0000-0003-1667-2584], Volodymyr Verhun [1][0000-0003-0683-0841]

and Natalia Kryvinska [2][0000-0003-3678-9229]

[1] Lviv Polytechnic National University, Bandery 12, 79013 Lviv, Ukraine
ivanna.m.droniuk@lpnu.ua, vverhun@gmail.com
[2] Comenius University, Odborajov 10, 82005 Bratislava, Slovakia
natalia.kryvinska@univie.ac.at

Abstract. The article deals with mathematical models for ranking online IT course students. There are four types of factors are investigated. Nonacademic factors are taking into account at the first time. For constructing criteria, Boolean functions are used. Criteria of success studying are extended to the studying of a group of students. On the base of mathematical model, the information technology is constructed. The schemes of information technology and construction are presented. Creating information technology prognoses the successes of online IT course students studying. The result of the application is time and others resources for accompaniment of studying minimizing.

Keywords: Mathematical model, Boolean criteria, Intellectual analysis, Information technology.

1 Introduction

The information technology industry plays an important role in the growth of the economy in Ukraine as a whole. The approximate number of professionals involved in the industry has exceeded 90,000 and experts expect the numbers will be double every 3 years [1]. The only one way to involve and engage the workforce into IT is education.

Lifelong learning is essential for continuous growth. In order to get the right skills needed to boost career employee should pay a lot of attention to. This leads to the appearance of many education-related issues. The first issue is the process of creating a curriculum and training schedules. Due to the rapid change in technology and approaches to software development, training programs do not fully cover the needs of the industry. The industry does not have the opportunity to wait a long time for training and retraining of employees. Therefore, the ability to quickly and autonomously develop programs and high adaptability are critical skills of any educational institution.

Another problem is the entry level of candidates who want to start studying for any programs. Knowledge at the very initial stage of training is critically important in

order to make the right decisions about involving candidates to the training and to create more personalized programs. In turn, such specific educational programs reduce learning time, are more likely to complete the entire cycle and gain the necessary knowledge and skills.

Another problem is the schedule and timing of training. Students need to learn a lot of material in a short time. Accordingly, good self-study skills are very important.

Given the above, it can be concluded that companies operating in the IT market need fast personalized programs with the opportunity to choose the best candidates even during the training. And this training should be done online in order to save time and resources.

The main purpose of this study is to investigate possible non-academic factors which may affect to the initial level of candidate and determine the factors and create mathematical model based on discovered factors which should be considered in selection process as a student ranking system. The participants' related data from online training is considered. Marks and test results are not considering since it is distance learning. The ranking system is needed in order to make decision regarding further job placement at any time of study even there are no actual feedback.

2 Analysis of problem

Predicting the success of students of online courses is one of the most common tasks in current research on the subject of educational data mining. Mostly studies focus on predicting final results and predicting students dropout. The studies select different data sets with different attributes. It should be noted that sampling the necessary and correct data is one of the biggest problems in these studies. For forecasting, the authors examine sets of academic, non-academic, and social background factors. According to the data mining process, data are pre-processed and machine learning tools are used to solve classification or clustering tasks. Figure 1 shows the areas involved in the data mining process of the training programs.

The most related field areas of the graph is learning analytics (Analysis of Learning) which can be defined as the measurement, collection, analysis and reporting of data about students and their contexts, for understanding and learning purposes and optimizing the environments in which occurs, therefore, Educational data Mining may share many attributes of all and each one of the surrounding areas [2].

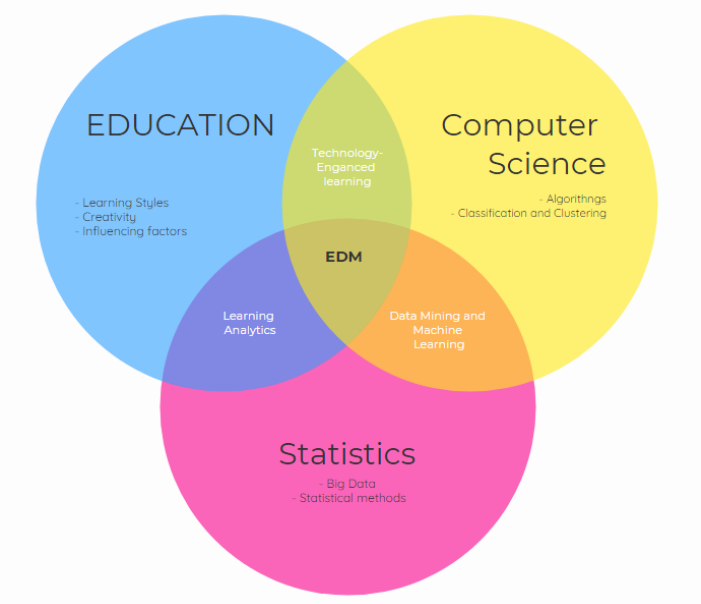

Fig. 1. Main areas of educational data mining researches [2]

Many higher education institutions are investigating the possibility of developing predictive student success models that use different sources of data available to identify students that might be at risk of failing a course or program.

In research [3], authors proposed a backpropagation neural network model to predict retention and college GPA of engineering students. This method is able of modeling two outcomes in the same model. The data of 1470 firstyear engineering students considered for the research. The predictors of the models include seven affective measure factors and eleven high school history factors. Authors proposed to include into model the non-academic factors such as leadership, surface learning, motivation, meta-cognition, expectancy-value. In the multi-outcome model, the authors reached the accuracy of retention prediction as 71.3%. Artifcial neural networks are widely used for such kind of researches. The authors of research [4] used conventional statistical evaluations to identify the factors that likely afect the students' performance. The neural network has been modelled with 11 input variables. Levenberg–Marquardt algorithm is employed as the backpropagation training rule. The neural network model has achieved a good prediction accuracy of 84.8%. Using predictive modeling methods, it is possible to identify at-risk students early and inform both the instructors and the students. Many researchers have been applied data mining methods to analyze, predict dropout students and also optimize finding dropout variables in advance. The authors of research [5] analyzed and measured the correlation between demographic indicators and academic performance to predict student dropout using three single classifiers. Such non-academic factors as parent's income, parent's education level, gender and age as student's dropout predictors has been included into

prediction model. Authors combined algorithms with Ensemble Classifier Methods using Gradient Boosting as meta-classifier and got about 79.12% prediction accuracy with proposed set of data.

There are a lot of approach for discovering student success predictors and dropout rate and risks. According to the literature review [6] the count of researches is constantly increasing. And the main area if research remaining the identification of influence factors to student success.

3 Description the factors, which have influence for the successes of studying

In our study the 4 types of factors are being investigated that influence students' online learning success: social factors, learning styles, factor of interest, and creativity index.

Let's take a closer look at the factors we have highlighted, in particular, the basic social factors that were examined in our study. Gender was considered as a qualitative factor by description (M or F). Additional education (for instance IT courses, online programs) was taken with the boolean description (Y or N). It was decided to divide the age factor into 3 sub-factors: age less than 22 years, age between 22-29 years, excluding initial and final values, and age more than 29 years. The division was considered due to the constant increasing the count of employees who decided to extend theirs already gained professional skills and become a software engineers. Therefore another factor was introduced - factor of any work experience presence, described by boolean value. (Y or N).

Data for determining social factors was determined based on the CV summary of the online IT course trainees who submitted their enrollment.

The next group of factors that have an impact on learning success is learning styles. In addition to the Johnson Questionnaire, a special survey was conducted to determine the student's learning style when enrolled. According to the methodology of this survey, four qualitative traits of learning style were identified: activist, reflector, theorist, pragmatist. Also, according to this survey methodology, each of these attributes can have the following clear meaning: S(Small), L(Low), M(Medium), VL(…), V(…), which corresponds to the level of expression of a given learning style.

The next factor considered in this study is the factors of interest. The basis for the formation of this factor were messages of trainees from any communication channels during training chat and forum. Figure 1 shows a diagram of the method of determining the index of interest. Initially, from any communication channel the listeners' messages has been selected for further analysis. The next step is an intellectual analysis of these data, the messages are classified, their weight is determined and the digital values of the quantitative characteristics of the factors of interest are formed on this approach.

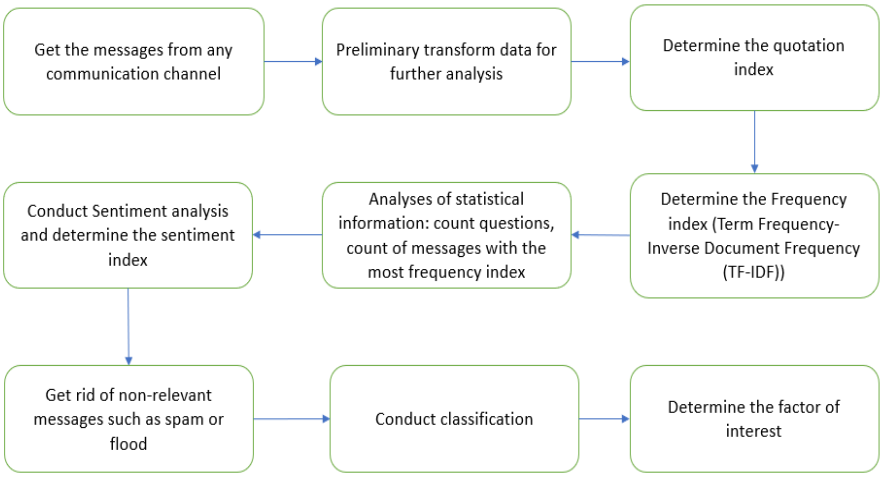

Fig. 2. A schema of finding the factor of interest method.

The last group of factors that we have considered is the group of factors associated with the creativity. Creativity is known to be a determining factor in the creative industry. Therefore, this factor is very important for the students of the online IT course. To evaluate this factor, we identified one parameter, which we called the Creativity Index. The creativity index in our studies is a boolean type characteristic (Y or N).

4 Creating mathematical model

4.1 Description factors for a model of successful studying

To construct a mathematical model, we introduce the boolean domain as a set consisting of one and zero: $B^1=\{0;1\}$, a also multidimensional Boolean space as a Cartesian product of one-dimensional $B^n = \coprod_{i=1}^{n} B^1$. The addition and multiplication operations will be the usual boolean addition and multiplication.

Four groups of factors were identified to describe the model, namely social, educational, interest and creativity. To characterize the first group of factors of a research object, we introduce the set of social factors $S=\{\vec{S}_1;\vec{S}_2\}$, where $\vec{S}_1 \in B_p$, and $\vec{S}_2 \in R^m, where\ p+m \leq n.$

We will assume that the factors of vector $\vec{S}_1$ correspond to the characteristics having the binary definition and the factors of the vector $\vec{S}_2$ describe characteristics having numerical values.

We describe the second group of factors that characterize learning as a vector $\vec{L} \in R^k$.

To describe the third group of factors that characterize students' interest in learning, consider the vector $\vec{Z} \in R^q$, and to describe the fourth group of factors that characterize the students' creativity, use the vector $\vec{C} \in B^r$.
Let's introduce the cumulative set X of definitions of all factors of the model, which is given by the following formula:

$$X = \{B^p, R^m, R^k, R^q, B^r\}, \quad (1)$$

Where N=p+m+k+q+r determines the total number of factors selected to predict the success of a course student's learning. Then the success of the trainee training can be determined on the basis of the following Boolean objective function

$$F: X \to B^1. \quad (2)$$

Successful training can be described by the formula

$$F(x) \equiv TRUE, \text{де } x \in X, \quad (3)$$

where X and F are given by the (1), (2).

4.2 Mathematical model for a group of students

If there are K trainees in the training group, then the following functions will characterize the overall success of the trainee group.

$$\Psi = F_1 \wedge F_2 \wedge \ldots \wedge F_K, \quad (4)$$

Where the symbol $\wedge$ means conjunction, and each of the functions Fi is given by (2). therefore, function (4) is conjunctive, which means that it is true when all the trainees have successfully completed the training.
Another feature that describes group learning is disjunctive:

$$\phi = F_1 \vee F_2 \vee \ldots \vee F_K \quad (5)$$

where the symbol is a conjunction and each of the Fi functions is given by (2), so this function will be true if at least one listener successfully completes the courses.

4.3 Construction a criteria

From the practice of conducting training on courses it is possible to distinguish two basic criteria for determining the success of completion of training of the listener. The first criterion corresponds to the fact that the trainee has successfully completed all the training tasks before graduation and did not drop it, that is, the function F of (3) for the trainees can be specified in the form $F^{(1)}$

$$F^{(1)}(x) \equiv TRUE\ , \quad (6)$$

where $F^{(1)}$- has the content of a Boolean function that depends on all the factors that have been determined and will become true if the trainee has successfully completed the training.

The second criterion corresponds to the fact that the listener successfully passed the final test, that is, the function F of (3) for this case is written in the form$F^{(2)}$

$$F^{(2)}(x) \equiv TRUE\ , \tag{7}$$

where $F^{(2)}$- has the content of a Boolean function that depends on all the factors that have been determined, and will become true if the listener has successfully passed the final test.

The successful completion of one student's training can then be characterized by the following formula

$$F^{(1)}(x) \wedge F^{(2)}(x) \equiv TRUE\ . \tag{8}$$

Thus, formula (8) expresses a clear mathematical criterion for successful completion of online IT courses by one listener and can be used to build information technology.

4.4 Construction a criteria for a group of students

To describe the successful learning of a group containing K listeners, we substitute formula (6) into expressions (4) and (5), and obtain functions describing the successful learning of a group of listeners.

In conjunctive form we have

$$\Psi(x) = (F_1^{(1)}(x)\ \wedge F_1^{(2)}(x)) \wedge \left(F_2^{(1)}(x)\ \wedge F_2^{(2)}(x)\right)\ \wedge \ldots \wedge \left(F_K^{(1)}(x)\ \wedge F_K^{(2)}(x)\right) \tag{9}$$

In disjunctive form we have

$$\phi(x) = (F_1^{(1)}(x)\ \wedge F_1^{(2)}(x)) \vee \left(F_2^{(1)}(x)\ \wedge F_2^{(2)}(x)\right)\ \vee \ldots \vee \left(F_K^{(1)}(x)\ \wedge F_K^{(2)}(x)\right) \tag{10}$$

In formulas (9) and (10), the variable x∈X from the set (1) contains N various factors that influence the success of learning. Thus functions (9) (10) of the criteria for successful study of a group of K students, which depends on N factors of different nature, are constructed..

5 Creating information technology

On the basis of the developed mathematical model, information technology for predicting the success of students' online learning courses development was developed. Developed information technology implements the collection, polls and storage of statistical and dynamic data of listeners. On the basis of these data, it conducts intellectual analysis of the collected data, classification and generates relevant values of the factors that influence the success of the training. On this basis, information technology ranks the students in the courses. (see Fig. 2). Figure 3 shows the structural

diagram of modelling students ranking system in information technology educational area.

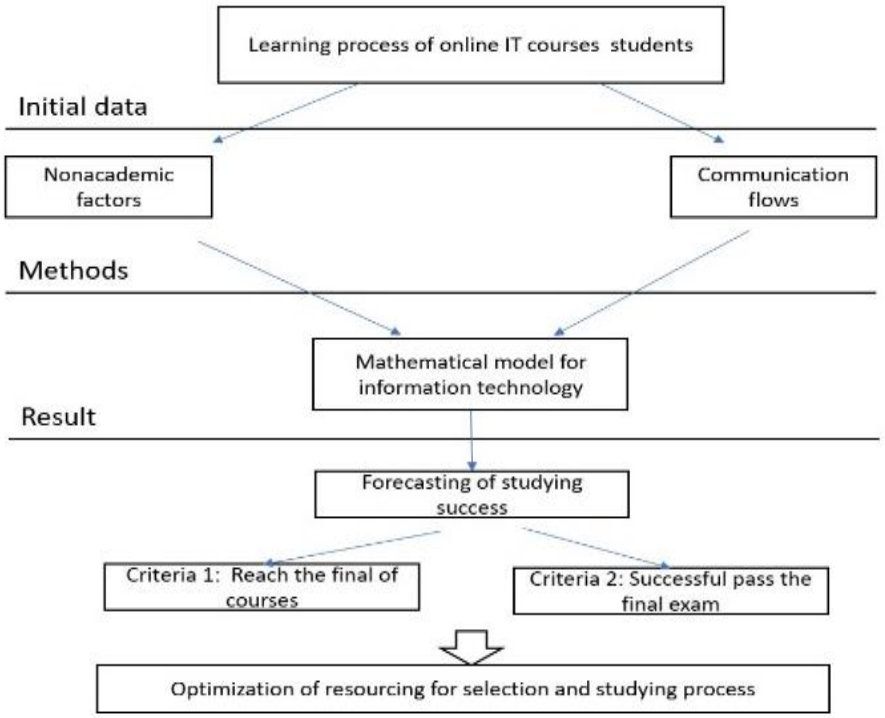

Fig. 3. A schema for creating mathematical model and information technology for online IT courses.

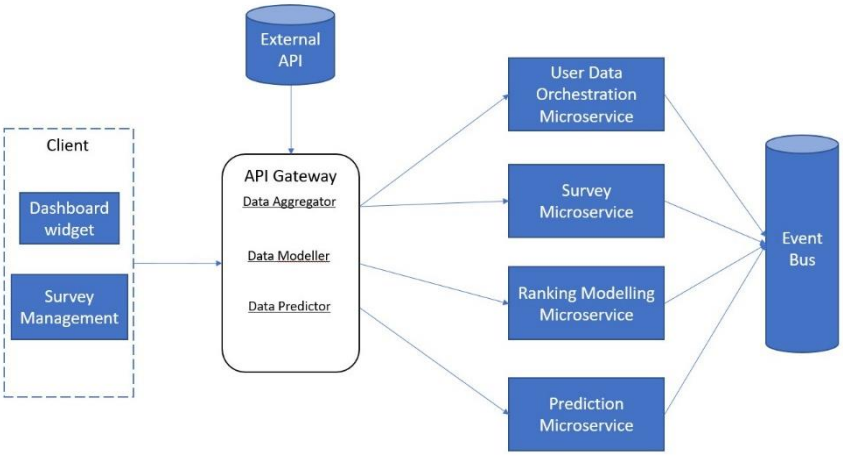

Fig. 4. Information technology structure.

Thus, the paper describes the development of information technology ranking students online courses in programming using non-academic factors and taking into account the principles of a competent approach to learning based on the developed mathematical model, methods and tools of information modeling

6 Conclusion

Building a mathematical description of the criteria for successful training of students online courses provided the opportunity to develop information technology ranking students. The application of the developed information technology creates the following advantages: increasing the level of individual approach to the students' education;

creation of the possibility of adaptation of the training program taking into account prevailing learning styles of students; identify the most active listeners for a better analysis of practical tasks. Also, the automatic collection of necessary information about the course of training provides real-time monitoring. All in all, the benefits outlined reduce the resource and time costs of supporting the learning process

The above benefits of using information technology relate directly to the learning process. However, information technology has several other advantages related to the process of selection of candidates for the course, as well as future employment. Taking into account additional non-academic factors has a positive impact on the decision on the selection of the candidate. The use of automation gives a better understanding of the structure of existing candidates and of the job market as a whole. Information technology enables early decision-making on the employment of a particular candidate. Extremely practically important in the developed information technology is the ability to predict the success of training in real time. Another practical application is the optimization of resources when deciding on the candidate for the vacancy.

Future research plans to consider other non-academic factors for successful completion of online IT courses. And also the study of the influence of a particular factor on the success of the listener

References

1. Software development in Ukraine: 2019-2020 IT market report. URL: https://www.n-ix.com/software-development-in-ukraine-2019-2020-market-report/
2. Romero, C., and Ventura, S.: Data mining in education, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, no. 1, 12-27, (2013).
3. Jin, Q., Imbrie, P. K., Chen, X.: AC 2011-1608: A multi-outcome hybrid model for predicting student success in engineering. Paper presented at the 2011 American Society for Engineering Education, Vancouver, BC. (2011)
4. Lau, E.T., Sun, L. & Yang, Q. SN Appl. Sci. (2019) 1: 982. https://doi.org/10.1007/s42452-019-0884-7
5. Hutagaol, Nindhia & Suharjito, Suharjito: Predictive Modelling of Student Dropout Using Ensemble Classifier Method in Higher Education. Advances in Science, Technology and Engineering Systems Journal, vol. 4 (2019)

6. Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., et al.: Predicting Academic Performance: A Systematic Literature Revi-ew. In Proceedings Companion of the 23rd Annual ACM Conferen-ce on Innovation and Technology in Computer Science Education,(pp. 175–199). (ITiCSE 2018 Companion). New York, NY, USA: ACM (2018) https://doi.org/10.1145/3293881.3295783
7. Dronyuk I., Fedevych O., Stolyarchuk R., Auzinger W. OMNET++ and Maple software environments for IT Bachelor studies PROCS36328, Procedia Computer Science, (2019) , vol.155, pp. 654-659, DOI:10.1016/j.procs.2019.08.093
8. Oleksiv, I., et. al.: Identification of IT Sector Stakeholder's Requirements to Masters Program in Information System in Lviv Region" In: Ermolayev, V., Suárez-Figueroa, M. C., Ławrynowicz, A., Palma, R., Yakovyna, V., Mayr, H. C., Nikitchenko, M., and Spivakovsky, A. (Eds.): ICT in Education, Research and Industrial Applications. Proc. 14-th Int. Conf. ICTERI 2018. Volume I: Main Conference. Kyiv, Ukraine, May 14-17, pp.112-120, (2018).