

DATA MINING FOR THE E-LEARNING RISK MANAGEMENT

Oksana GUSHCHINA

ORCID:0000-0003-2381-8537

Department of Applied Mathematics and Computer Science
Togliatti State University
Togliatti, Russia

Andrew OCHEPOVSKY

ORCID: 0000-0002-5270-5390

Department of Applied Mathematics and Computer Science
Togliatti State University
Togliatti, Russia

Received: 24/07/2018 **Accepted:** 04/03/2019

ABSTRACT

The article shows the role of data mining methods at the stages of the e-learning risk management for the various participants. The article proves the e-learning system fundamentally contains heterogeneous information, for its processing it is not enough to use the methods of mathematical analysis but it is necessary to apply the new educational methods of data mining. It determines the basic types of e-learning risks for the elimination or minimization of which proactive manners are applied that aimed at testing and planning other actions. The article gives the rationales for the use of various data mining methods at the different stages of the risk management: 1) for quality authentication of the risks the classification based on using the method called “decision tree” is applied; 2) for making the analysis of risks the brainstorm method for mind map creation is used. The mind map displays the e-learning risk groups that are then rated on the prioritization criteria with the help of expert evaluation method; 3) for the assessment of probability of risk impact on organization of e-learning the expert evaluation method, cluster analysis and bow-tie analysis are used. The article shows that the data mining methods are able to not only classify educational risks but also identify the causes and anticipate possible impacts on the final outcome. Having a great deal of information obtained through the process of the e-learning risk management and using it in data mining, it is possible to determine the reasons and the taken actions depend more the minimization of risks and strengthen the effectiveness of the e-learning.

Keywords: e-Learning, educational risk, risk management, data mining.

INTRODUCTION

Internet and development of telecommunications technologies allow to exchange the real-time information, it determines the new generation of e-learning instruments. E-learning is considered as the new context for education (Dolzhenko, 2016), through the manipulation of a huge amount of information, describing continuum of educational interactions (any educational action of the trainee in the e-learning is monitored and fixed in multiple databases, log-files, personal accounts etc.), it brings the need in constant updating of data for using it anytime and without any problem. Thus, currently the most relevant questions are the analysis of this information, the possibility of new knowledge extraction from already presented in databases, containers etc. Data mining (Gorlushkina, 2015) allows to understand students, to find out the difficulties in training course organization, to identify the most effective techniques of presentation of training material etc. These data can be used by to make the effective solutions for running the educational process. But the problem arises concerning data extraction, processing and structuring.

Extraction methods of the hidden data are discussed in data mining science discipline appeared at the intersection of artificial intelligence, data analysis, automatic classification and statistical methods of data processing (Romero 2013; Dringus 2005; Monk 2005). Its integration into e-learning systems is examined under the conception of educational data mining (EDM). The development of e-learning systems can lead to multiple risks related to analysis of multivariable learning data.

The aim of the present investigation is to show that at different stages of e-learning risk management various methods of data mining can be used. These methods can not only classify learning risks but determine the reasons of its occurrence and predict the consequences of their influence on the final result that will predetermine risk minimization and raise effectiveness of e-learning.

DATA MINING METHODS

The term “Data mining” is defined as “...bunch of detection methods in data previously unknown, practically useful and interpretive knowledge of educational process and its participants to support decision making” (Khlopotov 2014). The term educational data mining (EDM) is defined as “...an area of science related to development of techniques for unique data type research coming from educational sphere and its usage for better understanding of students and environments they have” (Scheuer 2012). Educational data mining methods are based on statistics, machine learning and database theory. The main activities of this area are: data mining usage for Intelligence Tutoring Systems support, analysis of education processes, visual data mining and visual education process pattern. The analysis of the scientific literature in the field of using the methods of data mining showed that this problem is interesting to many modern researchers. For example, in (Ceylan 2015) the authors propose a searching model system related to student success in the form of classifiers, each of them is learned with different dataset with hundreds of thousands of lines in relation to sections. Received classifiers would serve as an advisory system for students who want to choose courses prior the registration in the semester. (Kamisli 2017) provides a social network analysis and mood analysis for students of the open and distance education systems by using tweets. With the help of statistical and data analysis methods, tweets analysis allowed to analyse students attitudes. In (Herlina 2017), the role of the K-Means algorithm for classifying students learning activities using e-learning was showed. This algorithm helped to form student activity and improving student abilities cluster. An approach based on minimal spanning tree for clustering e-learning resources is proposed in (Wu 2016). The developed clustering method can classify students into groups so that a homogeneous classification can increase the learning effectiveness. In addition, the authors believe that the clustering of e-learning resources would be valuable for decision making in terms of learning strategies and learning design for intellectual learning. The system proposed in (Apurva 2017) used the ontology-based document clustering approach, that is based on a two-stage clustering algorithm. Combining the clustering of documents with ontology would help to create better clusters that would not ignore the semantic connection between words. This proves that the clustering approach that uses the semantics of the documents for term weighting produces better results than the approach without semantics. (Rawat 2019) justified the use of cluster analysis for classifying a new student into the corresponding class and recommending relevant courses using various evaluation metrics. This advisory system should help students to improve their academic progress and, therefore, the overall learning process. (Hussain 2018) used machine learning algorithms to identify students with a lack of involvement in the course. The results showed that a decision tree classification gives better performance in the context of accuracy. The system model for clearing and drilling learning data (Liang 2017) allows to form a student profile, as well as student attitudes to learning and the endurance of academic behavior by using a similarity algorithm and Jaccard’s coefficient. In accordance with e-learning resources and student behavior, the authors presented an intellectual leadership model that would help to improve the e-learning platform and student knowledge. The authors (Panyajamorn 2018) have proposed effective e-content development method that would include effective indicators to identify potential students and improve their abilities. This method is based on data analysis, which included a quantitative research approach with the use of paired Student’s test to determine differences in estimates before and after testing; analysis of variance (ANOVA) for the analysis and search for the relationship between improvements in grades. (Wu 2018) uses a text data mining to automatically identify student interest in an open academic environment. The authors suggest a method for computer-aided engineering of student interests based on a combination of the content generated by the

student and its dynamic interaction with other educational resources. This method was used as a basis for the student theme model, in order to find out not only the interests of student knowledge (interest in creating a content), but the interests of the student collector (interest in collecting a content generated by others).

The table 1 shows the educational data mining methods with short descriptions (Baker 2010; Bienkowski 2012; Romero 2006).

Table 1. The main educational data mining methods

Method	The description of method
Prediction: classification, regression and density evaluation	Model development in which the output of a single aspect of data (predictable variable) is based on the combination of other aspects (variable-predictors) (Herlina 2017)
Clustering	Finding points that are naturally connected and data separation on clusters. It is used when a data structure is unknown (Wu 2016)
Analysis of the relations	Finding out the relationship between different variables in the database set and for the algorithm encoding for future use. The different types of methodologies for analysis of the relations are used: search for strong rules, analysis of temporary patterns, correlation data mining, analysis of causes (Apurva 2017)
Data processing for assessing	Authentication and evaluation of data and judgment. Compilation, visualization and interactive interface to support decision making are used (Rawat 2019; Hussain 2018)
Outlier detection	Data discovery which deviate greatly from remaining values (Panyajamorn 2018)
Text mining	Finding and extracting useful patterns, models, rules, trends from unstructured documents. It includes classification, clustering, extraction of the concept, sentiment etc. (Wu 2018)
Model study	Study of pre-compiled models (developed with the help of EDM-methods of forecasting, clustering, knowledge engineering). Application of models to data and its use as the component of other analysis (Liang 2017)
Data conversion to the form acceptable to person	Data conversion to the form acceptable for person usage (Kamislı 2017)

Description of presented methods shows that data mining is a powerful tool for information extraction and its transformation to the form needed for subsequent use (Hanna 2004). Using of the methods can solve a lot of problems:

1. evaluation and deployment of the materials on the base of usage and performance data;
2. supporting consultations on complex issues;
3. performance measurement after the given consultation and advice;
4. assistance in searching for suitable concourse etc.

In addition, global trends, dynamic environment, difficulty of the problems requiring greater efficiency, adaptability, integration and coordination of all of relevant design process and implementation of the e-learning systems.

In this context, the probability of occurrence of events that could influence on achieving goals has appeared, in other words, the probability of occurrence of non-performance risk or appearance of a specific threat or expected losses (Matveenko 2012). The occurrence of the risk relates to uncertainty that is the lack of knowledge of facts and the lack of uniqueness. Since the e-learning system does not allow to avoid influence of uncertainty factors, the problem of risk management, linked to its registration and regulation, needs to be resolved (Chubarova 2005) by decision making focused on performing constant risk assessment and action planning in order to reduce any risk.

e-Learning Risk Management

The term “risk management” in the e-learning area means the idea that characterized quality of stable educational activity, more specifically, initial situation, preparation and decision-making processes under the delivery and receiving educational services (Chubarova 2005). In the e-learning, risks may be associated with general computer threats (Barik 2012) that include:

1. breach of confidentiality: an unauthorized party gets the access to the assets presented in the e-learning system;
2. violation of integrity: unauthorized access to materials used in the e-learning system;
3. denial of service: prevention of access privileges by violation of traffic during the e-learning system users activity;
4. virus attack: a malware changes the code in case to crash other programs;
5. problem with the requirements and organizational documents;
6. theft and unauthorized data changes;
7. lack of access to servers and personal computers.

As the result of the above-mentioned threats for various e-learning system participants (Authors, Students, Managers, Tutors and System Designer/Administrator) different types of educational risks may develop. Substantiated classification of risks allows the clear identification of each place in the general e-learning system where each risk is assigned its policies and techniques: identification, analysis, constant evaluation, monitoring and active control until the risks are resolved or turn into the problems that could be processed (Ruggeri 2013).

Figure 1 shows the main stages of risk management process that may be offered as building blocks of repetitive and dynamic process of risk management in education.

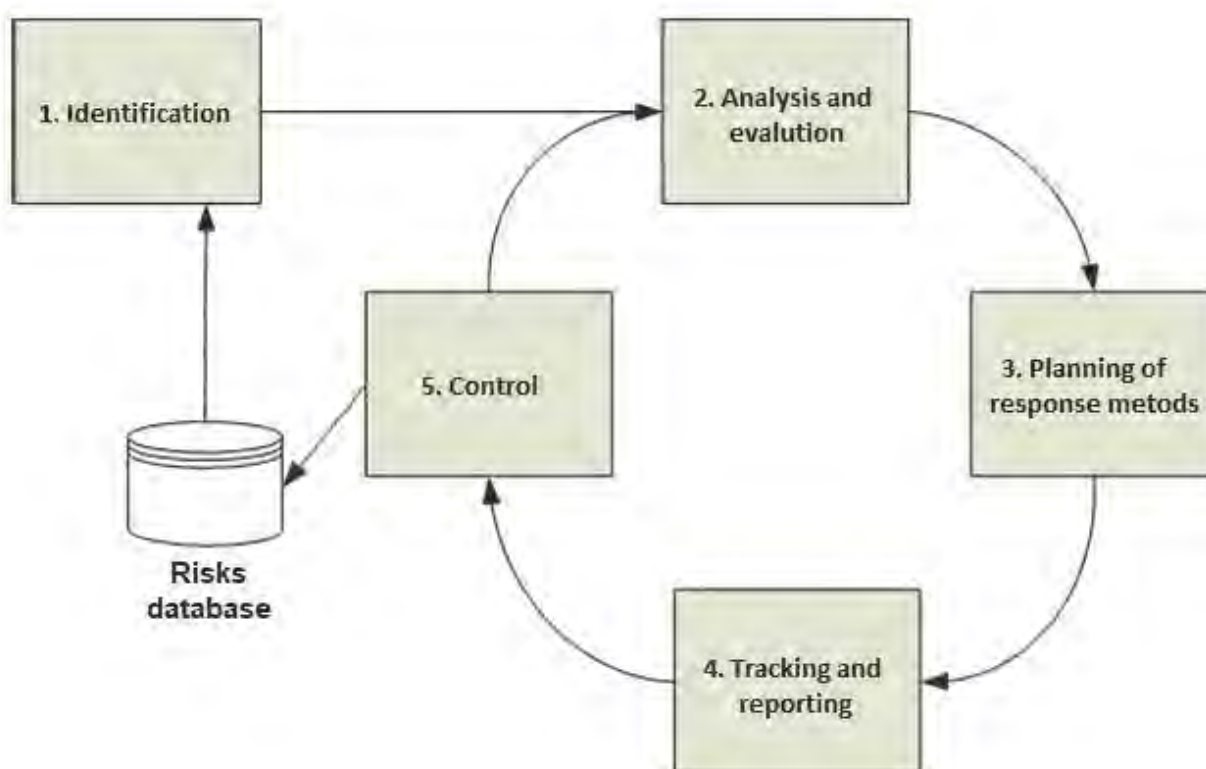


Figure 1. Main stages and processes in the area of educational risk management

Application of Data Mining Methods at the Stages of the e-Learning Risk Management

Consider the main stages of e-learning risk management and demonstrate the application of data mining methods for evaluation and minimization of educational risks.

Identification of Risks

Identification of risks is the repetitive process as the e-learning evolved within its service life new risks may occur or information about them will appear and the risks will be known (Khodyreva 2017). The risks should be proper described on the stage of identification. Description format is meant to be consistent for clear and explicit understanding of each risk in order to support the effective analysis and develop the response plan.

For the qualitative risk identification educational data mining method is used – classification, which aims to define category and class for each risk. As shown in Pena-Ayala (2014) works most of research in the area of EDM use different classification algorithms that are: decision trees, k-nearest neighbor algorithm and naive Bayes classifier (Bishop 2006).

K-nearest neighbor algorithm for classification of e-learning risks is used. Under this algorithm all of the risks are applied to the class that is most common among others, whose classes are already known. Put in the basis of method realization the multiple selection based on the voting among the experts: r-vote is the class number to which r-closest neighbor is related. Define class weight G_r as approximate quantity of risks in the class (1).

$$w_r = \frac{n_r}{k} \quad (1)$$

where k-nearest neighbor belongs to class, to class, ..., to class, so that

$$n_1 + n_2 + \dots + n_N = k.$$

Let us formulate several types of rules about reference educational risk to some class:

1. simple majority rule: new risk belongs to that class, which risks replace over half of place in selection (2):

$$w_r = \frac{n_r}{k} \geq 0.5 \quad (2)$$

2. simple majority rule: new risk will belong to the class with the large number of elements in selection of k-nearest neighbor, in other words, to the class that obtains the largest number of votes (3):

$$r = \arg \max \{n_r\} \quad (3)$$

3. weighted majority rule: new risk will belong to the r-class which number is defined by the following formula (4):

$$r = \arg \max \left\{ \frac{w_r}{\rho_{min}^{G_r}} \right\} \quad (3)$$

- 4) single sample mean rule: new risk belongs to class to which the mean distance is minimal (4):

$$r = \arg \max \bar{d}(x, G_r) \quad (4)$$

Figure 2 shows the result of outlined rules usage for e-learning risk classification obtained by the experts using analysis of literature (Barik, 2012; Bozinoff 2014; He 2013; Ilyina 2016; Khodyreva 2017). All of the risks were classified as follows: author's risk, tutor's risk, manager's risk, system designer's risk, student's risk, and other risks that may not belong to any of the mentioned groups.

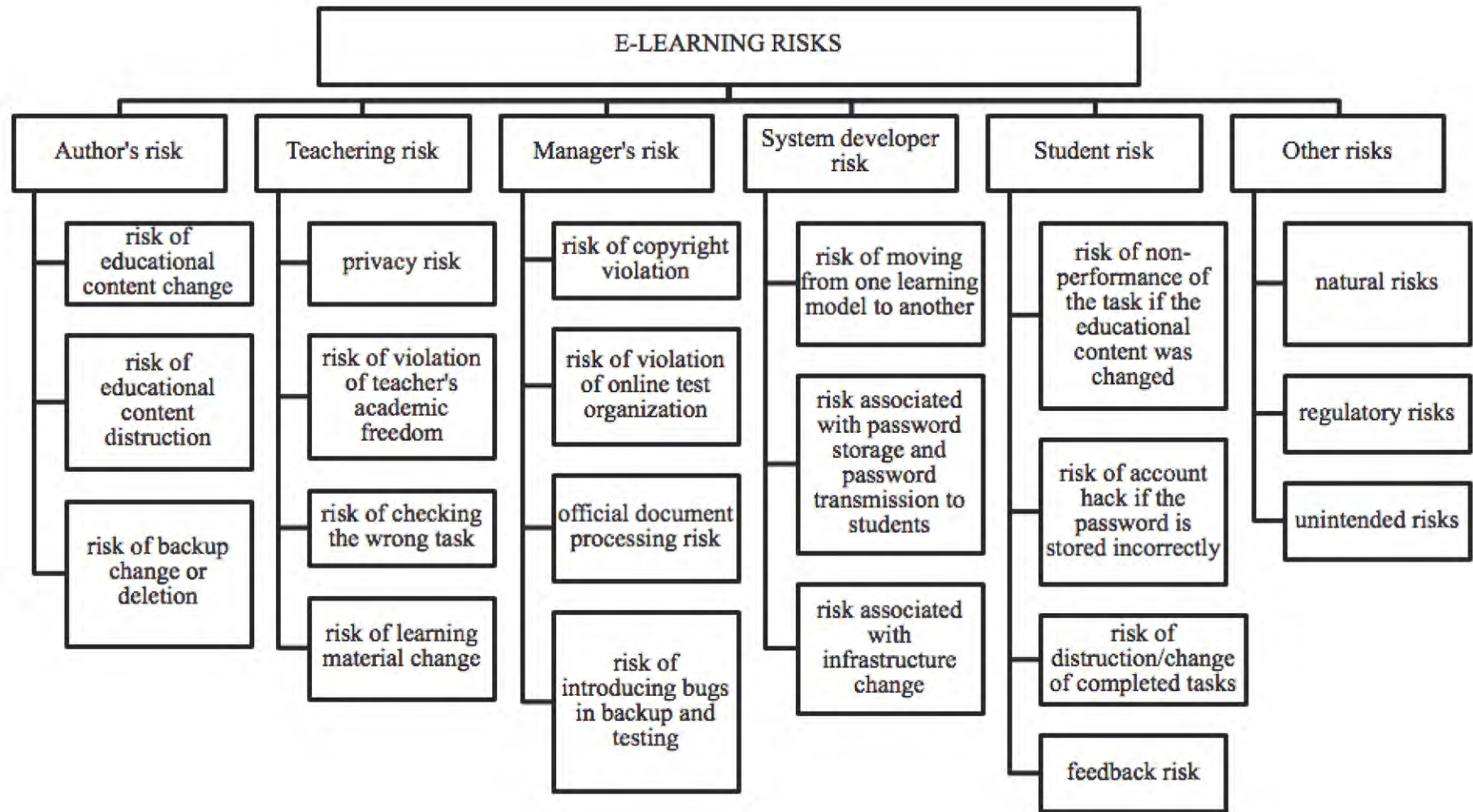


Figure 2. Classification of e-learning risks

1. *Author's risk* – all of the e-learning risks associated with implementation and presentation of learning material.

The authors have responsibility for development and implementation of learning material so a threat that the compile material can be referred or processed without the consent of the authors arises. Author's responsibility is to protect learning material from any unauthorized use, modification or re-use of data in different contexts related to e-learning. Author's lectures, video clips, documents, learning activities and etc. can be changed/destroyed so it is in the author interests to make sure that users are getting educational content without any changes and in a full manner.

Besides, risks of educational content may arise in e-learning associated with a number of reasons: bad quality of e-learning materials; outdated and inaccurate information; absence of pattern and lack of visualization; uniformity of tests; lack of learning activities of different complexity and absence of examples of tasks execution.

2. *Teaching risk* – all of the e-learning risks associated with the use of different educational methods, examinations and final grades.
 - risk for student and teacher's privacy related to usage of different forms of learning materials discussion, for example, via online-forum where the documents are retained on electronic file on a server;
 - risk which includes standardization of test questions and the list of questions that abridge teacher's academic freedom;
 - risks linked with checking executed tasks associated with deception;
 - risks of learning material change while transmission to the students (He 2013).

The teachers have responsibility for learning material so they can develop educational content in accordance with requirements of the course.

3. *Manager's risk*. In the e-learning system should be some governing board setting rules for its creation and functioning.

The main managerial risks in the e-learning associated with dealing with legal issues: copyright, online-testing, sending of official documents etc. Managers should see about entry onto the course and cancellation of registration as the need arises, development of backup plan and testing of process of annealing, otherwise, when needs arise it would be hard to update data.

Manager's risk is the risk that divides responsibilities in such questions as:

- saving passwords of all of the servers and routers;
 - recording of daily traffic;
 - surveillance of constant electrical power supply of server and other network based devices;
 - control of authorization, in other words access strategy (reading, recording and doing) for students and other participants for effective performance of the e-learning system.
4. *System developer risk*. In the e-learning system the courses classify into different modules. In the case of market demand requirement one educational model may be changed to another that leads to the solution of different problems for support and implementation of a new one.

The quality hardware components for modeling, development and delivery of the e-learning components need: high performance web-servers and databases, high-speed Internet with quality LMS and also safe infrastructure that are able to support several users and networking applications.

Developer risk associated with storing of passwords in the open code of the application and its transmission to students and teachers.

5. *Student risk*. The main participants of the e-learning are students who are learning and sharing their knowledge between each other. The student is obliged to know learning material but the risk of unsuccessful examination if the learning material was changed by hackers may arise.

There are following types of risks for a student:

- risk of storage of information for login (identification of users and passwords);
- risk of unsuccessful authorization for gaining access to the e-learning server if the information for login was not used correctly;
- risk of wrong usage of feedback mechanism;
- risks of phishing when on a counterfeit website students are encouraged to enter their confidential information.

6. *Other risks* in the e-learning:

- natural risks (problems of an energy source, problems with the cell phone reception, equipment undercapacity etc.);
- regulatory risks (threats fuelled by swindle, blackmail, theft etc.);
- unintended risks caused by computer errors, power cutoff, process errors, etc.

Analysis and Risk Assessment

Analysis and risk assessment of the e-learning are performed for the purpose of transformation gained from identification of the information into the one that allows taking decisions. In the process of the qualitative analysis of the risks a number of assessments of possible adverse effects caused by detected factors are produced. In the process of quantitative analysis indicator values of the probability of menacing events caused by the occurrence of risks are determined and specified.

E-learning system uses Data Mining methods for knowledge extraction by analysis of information that is available in a form of data created by users. The basic principle of data mining is unused data extraction that can be used in the future. Data mining passes through several stages:

1. data processing and selection. Internal and external databases provide data for relational description between them;
2. data update. When positions and values are identified, it is necessary to delete useless data and convert processing data into accurate.
3. data coding. In the process of data coding all of the data can be used for statistics, artificial intelligence or machine learning and the results will be attained quickly.

Given the specificities of the e-learning, ranking method was used for qualitative risk analysis, with its help grading of factors according to its influence on targeted indicator was performed. Each factor was associated with a weighting factor that makes quantitative evaluation of impact degree. The expert report was made by the group of L-experts, each of which arranged private criterion of risks in order of importance: from 1 (the most important private criterion) to n (the lowest private criterion in importance etc.). Then the transformation of ranks was made as follows: rank 1 – received m-appraisal, rank 2 – received m-1 etc. until the rank m that received appraisal 1. Let's define received appraisals r_{ik} , where i is expert, k is criterion. The results of the expert survey can be compiled in a table 2.

Table 2. Criterion scores of educational risks

Experts	Criteria			
	F_1	F_2	...	F_m
1	r_{11}	r_{12}	...	r_{1m}
2	r_{21}	r_{22}	...	r_{2m}
...
L	r_{L1}	r_{L2}	...	r_{Lm}
Σ of appraisals	r_1	r_2	...	r_m

Appraisals summarization of received criteria with the formulae (5) for the determination of a weighting factor (6) was made:

$$r_i = \sum_{j=1}^L r_{ji}, \text{ where } i = 1, \dots, m \quad (5)$$

$$\lambda_i = \frac{r_i}{\sum_{i=1}^m r_i}, \text{ where } i = 1, \dots, m \quad (6)$$

With the usage of the weighting factor risk matrix was received that, with combining two indicators (probability and operation) for defining the importance of the risk, defined many educational risks required increased attention. Array cells were “thresholds”: unacceptable, average and low risks that were then classified in order of influencing on effectiveness of the e-learning system organization. Using array computations the analysis of the e-learning risks was made in order to identify the root causes of its appearance. This analysis was made with the use of method of creative generation of ideas – “brainstorm” and the method of expert evaluation.

Mind Map, which was produced by psychologist Tony Buzan (Muller 2007) and presents convenient and effective imaging team of mentality and alternative recording of the clear reflection of association links, was used to visualize the brainstorm (Khakimov 2016).

Figure 3 shows the Mind Map that reflect the process of general systems thinking using schemes (Nozdrina 2012) (Fig. 3) for showing the groups of e-learning risk that then were ranked by the priority criterion by the method of expert evaluation.

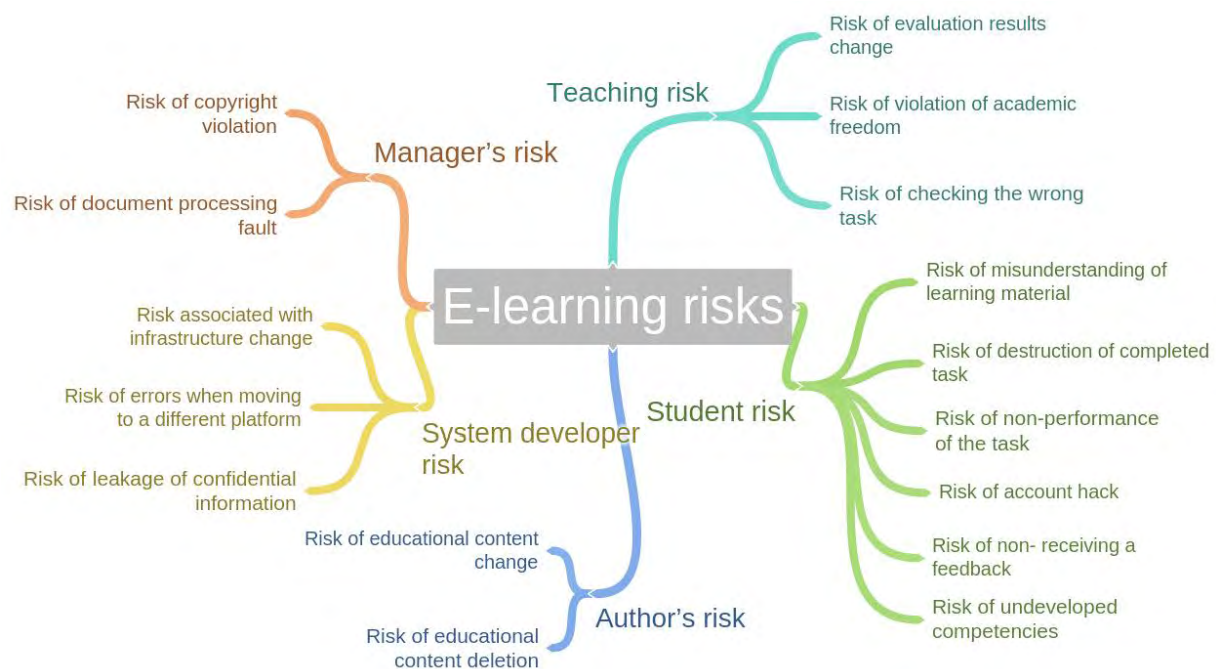


Figure 3. Mind map of e-learning risks

The table 3 shows average of expert evaluation received by 100 respondents who needed to evaluate the rank of each level by ten-to-one scale, where 1 is low and 10 is critical risk. On the basis of these exponents expert evaluation of probability of risk influence on e-learning organization was rated.

Risk \ Parameter		Probability of risk occurrence	The degree of influence on the e-learning organization	The degree of influence on getting an education by a student	Probability of losing students	Degree of influence on the quality of educational services
	<i>Weight</i>	<i>0,2</i>	<i>0,15</i>	<i>0,1</i>	<i>0,2</i>	<i>0,35</i>
Teaching risk	Risk of evaluation results change	3,5	0,35	1,2	0,6	1,8
	Risk of violation of academic freedom	2,45	0,65	1,05	0,3	0,4
	Risk of checking the wrong task	1,65	0,85	0,15	0,2	0,4
Author's risk	Risk of educational content change	2,45	0,445	0,9	0,3	0,4
	Risk of educational content deletion	3,5	0,3	1,35	0,8	1,6
Student risk	Risk of misunderstanding of learning material	2,8	0,7	0,15	0,8	1,6
	Risk of destruction of completed task	1,35	0,8	0,15	0,2	0,2
	Risk of non-performance of the task	2,75	0,6	0,15	0,8	1,8
	Risk of account hack	3,25	0,35	1,2	0,8	1,8
	Risk of non- receiving a feedback	2,45	0,65	1,35	0,6	1,6
	Risk of undeveloped competencies	2,75	0,45	1,5	0,3	0,4
System Developer risk	Risk associated with infrastructure change	2,45	0,55	1,5	0,6	0,8
	Risk of errors when moving to a different platform	1,4	0,85	1,35	0,7	1,2
	Risk of leakage of confidential information	1,35	0,8	0,15	0,1	0,6
Manager's risk	Risk of copyright violation	1,55	0,75	1,2	0,2	0,4
	Risk of document processing fault	3,15	0,3	1,2	0,6	1,6

Table 3. The results of the e-learning risks analysis through the expert evaluation method

For the interpretation of received results with the help of mathematical statistics:

1. Generalized expert opinion was derived and average value of risk was defined as (7).

$$S_j = \frac{\sum_{i=1}^m \alpha_{ij}}{m_{kj}} \quad (7)$$

Let m be quantity of experts evaluate j -risk;

i is the expert number, $i = 1, \dots, m$;

j is the risk number, $j = 1, \dots, n$.

2. Average rank for risk universe was defined as (8).

$$\bar{S} = \frac{\sum_{j=1}^n S_j}{n} \quad (8)$$

where n is the number of risks.

3. Deviation of average rank of j -risk from average rank of risk universe was defined as (9).

$$d_j = \bar{S} - S_j \quad (9)$$

4. Kendall's coefficient of concordance or coefficient of multiple correlation of ranks for the determination of compatibility of expert opinions was defined as (10).

$$W = \frac{12S}{m^2(n^3 - n)} \quad (10)$$

where m is the number of experts in group,

n is the number of factors,

S is variations in rank matrix under the average one.

5. General assessment of coefficient of concordance by the calculation of Pearson's chi-squared test was made (11) (Gafarova 2015).

$$\chi^2 = \frac{\sum_{j=1}^n d_j^2}{\left[\frac{1}{12mn(n+1)} + \left(\frac{1}{n} - 1 \right) \sum_{i=1}^m T_i \right]} \quad (11)$$

where m is the number of experts,

n is the number of risks,

T_i is the coefficient taking into account occurrence of related variants of i -expert that is defined as (12).

$$T_i = 1/12 \sum_{q=1}^Q t_q^2 - t_q \quad (12)$$

where Q is the number of bands (forms of duplicate elements) of i -expert evaluations;

t_q is the number of elements in q -band for i -expert (the number of duplicate elements).

The coefficient of concordance showed that among experts were groups with good consensus of opinions but these opinions are opposite and altogether have neutralized one another. For this reason additionally k-means clustering was used (Karun 2013; Kavitha 2017), the purpose of using of which is to split the e-learning risks into classes (fig. 4), each corresponding to specific risk group – unacceptable, average and low risks.

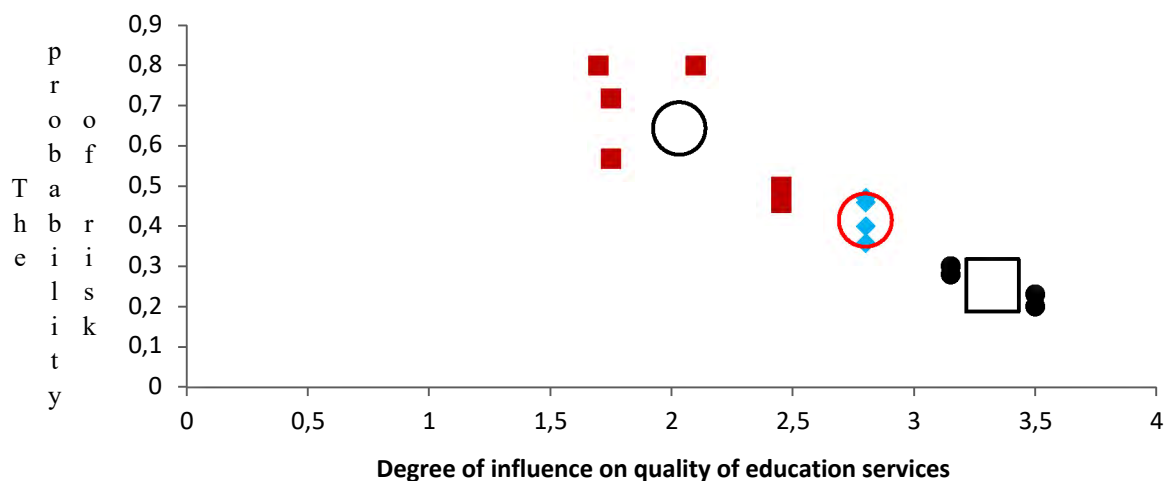


Figure 4. The result of e-learning risk clustering

Cluster analysis made it possible to combine e-learning risks into classes (clusters) so in one class are the most similar risks and the risks of different classes are vary from each other. Cluster analysis identified a group of risks that has maximum effect on organizational efficiency of the e-learning.

To sum up, at the end of analytical work about the e-learning risk evaluation were received:

- the list of risks grouped by priority;
- the items list which required additional analysis due to the high probability of occurrence and increased level of influence on the e-learning organization. Received list of the e-learning risks requires additional constructive analysis. Bow-tie analysis is applicable for schematic description and analysis of “Hack of student account” (Fig. 5) and also for reasons and consequences that allow to evaluate probability, level of negative effect and correlation between risks.

Planning of Ways of Responding to Risks

Planning of ways of responding to the e-learning risks is the regulated procedure of development of a plan to minimize negative consequences of the risk impact. Four basic methods of response to risks can be marked: avoidance, minimization, transfer/insurance, acceptance (Sultanov 2016).

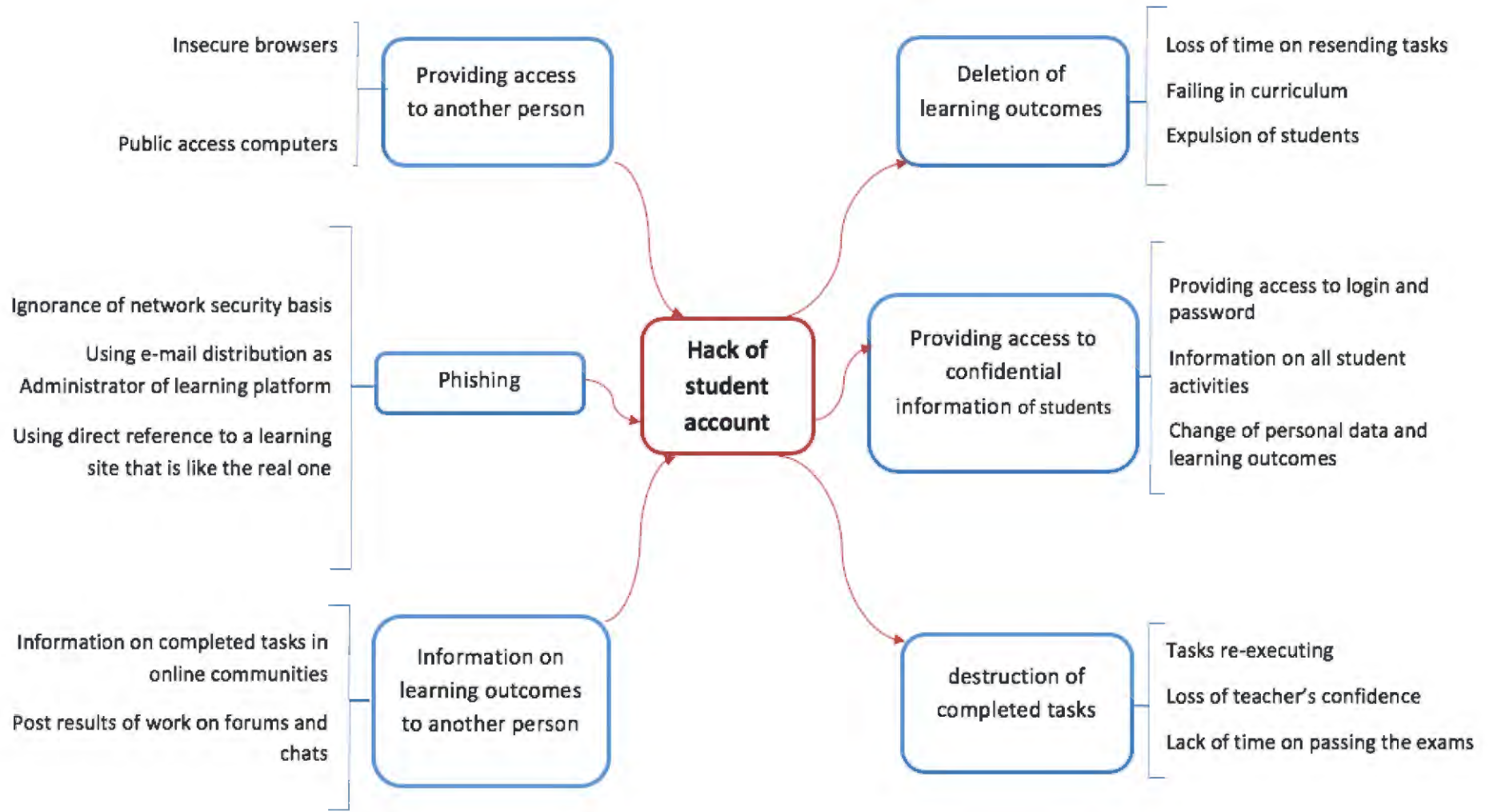
With the use of these methods for planning the ways of responding to e-learning risks model of dependence containing the list of measures to minimize received the e-learning risks was created. This helped to manage effectively e-learning risks.

Risk Management, Report and Control

At the stage of risk management were made the reports that showed the results of responding: risk was allowed; additional corrective actions for risk solution were taken.

At the stage of e-learning risk control database with received educational risks was formed. The database is being continuously updated and contains innovative models of risk management created with the usage of data mining and focused on e-learning assessment.

Figure 5. Bow-tie analysis for description and analysis of risk from reasons to consequences



GENERALIZATION AND CONCLUSION

Due to the development of the e-learning monumental flows of heterogeneous data (Duke 2011) were showed that need to be analyzed. In addition, many factors or risks could have a negative effect on developing e-learning, the probability of its occurrence is to be esteemed and minimized, in other words – managed.

In our view, traditional methods of mathematical analysis cannot allow the variable solution of the e-learning risk management goals that combines with averaging over selection conception led to operations on fictitious variables and lack of analytical criteria for authenticity of relations and regularity in multi-dimensional data etc.

We believe that for the diverse data processing represented by quantitative and qualitative data, data mining can be applied, the main feature of which is the interpretation function of regularities that is the basis of creating the rules for the object entry into equivalence classes. In addition, they are used in solving problems whose regularities cannot be accurately described in the language of statistical or other analytical mathematical models.

Data mining methods not only identified the e-learning risks, but also determined the causes of their occurrence and predict the possible consequences of their impact on the final result. Therefore, having a large amount of received in the process of the e-learning risk management information and subject it in Data Mining, it is possible to determine the reasons and driving forces, measures, taken actions minimization of risks and strengthen the effectiveness of the educational process in e-learning system are depend, as well as to identify factors that do not have a statistically significant impact or impact negatively to the outcome.

BIODATA and CONTACT ADDRESSES of AUTHORS



Dr. Oksana GUSHCHINA, is an Associate Professor at the Department of Applied Mathematics and Computer Science, Institute of Mathematics, Physics and Information Technologies, Candidate of Pedagogic Sciences (2003), Associate Professor (2009). Research interests: technologies of data science, e-learning and distance learning, AR and VR experience. The writer of some undergraduate programs in the field of Computer Science. The author of more than 50 academic papers and 10 study guides.

Oksana GUSHCHINA

Address: Department of Applied Mathematics and Computer Science, Institute of Mathematics, Physics and Information Technologies
Togliatti State University 445020, 14, Belorusskaya St., Togliatti, Russia
Phone: +7 8482 539181,
E-mail: ok_mih@mail.ru



Dr. Andrew OCHEPOVSKY, is a Head of the Department of Applied Mathematics and Computer Science, Institute of Mathematics, Physics and Information Technologies, Candidate of Engineering Sciences (1994), Associate Professor (2007). Research interests: data sciences, e-learning, distributed systems and computing. The writer of some undergraduate and master's programs in the field of Computer Science. The Head of Engineering Research Center «IT Student». The author of 54 academic papers, two of which were published in international academic journals.

Andrew OCHEPOVSKY

Address: Department of Applied Mathematics and Computer Science, Institute of Mathematics, Physics and Information Technologies
Togliatti State University 445020, 14, Belorusskaya St., Togliatti, Russia
Phone: +7 8482 546473,
E-mail: ochepovsky@mail.ru

REFERENCES

- Apurva, Dube, Pradnya, Gotmare (2017) "Semantics Based Document Clustering" // *Int. J. Sci. Res. in Computer Science and Engineering*, vol-5(4), Issue.4, pp. 26-31.
- Baker, R. S. J. D. (2010) 'Data mining for education', *International encyclopedia of education*, no. 7(3), pp. 112-118.
- Barik, N., & Karforma, S. (2012) 'Risks and remedies in e-learning system', *International Journal of Network Security & Its Applications*, no. 4(1), pp. 51-59.
- Bienkowski, M., Feng, M., & Means, B. (2012) 'Enhancing teaching and learning through educational data mining and learning analytics: An issue brief', *US Department of Education, Office of Educational Technology*, no. 1, pp. 1-57.
- Bishop, C. M. (2006) *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer.
- Bozinoff, M., Tankosic, M. (2014) 'E-Learning Risks Management as Competitive Advantage in Institutions of Higher Education'. *Proceedings of the 14th International Conference on Applied Computer Science (ACS '14) "Modern computer applications in science and education"*, Cambridge, MA, pp. 164-170.
- Ceylan, G., Ozturk, G., Kamisli, Ozturk, Z. and Aydin, S. (2015). "Student Success Prediction for the Mega University of Turkey", 27th European Conference on Operational Research
- Chubarova, O.I. (2005) 'Educational risk as an economic category, its essence' *Polzunovskii vestnik*, no. 1, pp. 199-208.
- Dolzhenko, A.V. (2016) 'The content of information competence of the teacher in the environment of e-learning', *Bulletin of the Taganrog Institute named after A.P. Chekhov*, no. 1, pp.121-123.
- Dringus, L., Ellis, T. (2005) 'Using data mining as a strategy for assessing asynchronous discussion forums', *Computer & Education Journal*, no. 45, pp. 141-160.
- Duke, V.A., Flegontov, A.V., Fomina, I.K. (2011) 'Application of technologies for the intellectual analysis of data in the natural sciences, engineering and humanitarian fields', *Proceedings of the Russian State Pedagogical University named after. AI Herzen*, no. 138, pp. 77-84.
- Gafarova, L.M., Zavyalova, I.G., Mustafin, N.N. (2015) 'On the features of the application of the Pearson consensus criterion X^2 ', *ESGI*, no. 4(8), pp. 63-67.
- Gorlushkina, N.N., Kotsyuba, I.Yu., Khlopotov, M.V. (2015) 'Tasks and methods of intellectual analysis of educational data for decision support], *GTR*, no. 1, pp. 472-482.
- Hanna, M. (2004) 'Data Mining in the e-learning domain', *Campus-Wide Information Systems*, no. 21(1), pp. 29-34.
- He, W. (2013) 'A survey of security risks of mobile social media through blog mining and an extensive literature search', *Information Management and Computer Security*, no. 21(5), pp. 381-400.
- Herlina, Latipa Sari, Dewi, Suranti Mrs.and Leni, Natalia Zulita (2017) "Implementation of k-means clustering method for electronic learning model" // *International Conference on Information and Communication Technology (IconICT) IOP Publishing IOP Conf. Series: Journal of Physics: Conference Series*, Volume 930.
- Hussain, M. et al. (2018) "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores" // *Computational Intelligence and Neuroscience*, vol 2018.
- Ilyina, T.S., Zakharov, N.Yu. (2016) 'Management of educational risks', *Vestnik VGUIT*, no. 4(70), pp. 290-295.
- Kamisli, Ozturk, Z., Erzurum, Cicek, Z.I. and Ergul, Z (2017), "Sentiment Analysis an Application to Anadolu University," *Acta Physica Polonica A*, vol. 132, no. 3, pp. 753-755.

- Karun, K., Isaac, E. (2013) 'Cogitative Analysis on K-Means Clustering Algorithm and its Variants', *International Journal of Advanced Research in Computer and Communication Engineering* [online] Available at: <https://www.ijarce.com/upload/2013/april/49%20-%20kavitha%20KARUN%20-%20Cogitative%20Analysis%20on%20K-Means.pdf>
- Kavitha, G., Raj, L. (2017) 'Educational Data Mining and Learning Analytics', *Educational Assistance for Teaching and Learning. International Journal of Computer and Organization Trends (IJCOT)*, no 7(2), pp. 21-25.
- Khakimov, D.R. (2016) 'Application in the educational process of mental maps', *Educational resources and technologies*, no. 1(13), pp. 3-8.
- Khlopotov, M.V. (2014) *Models and algorithms for intellectual analysis of educational data for decision support*, Dissertation cand. tech. sciences, St. Petersburg.
- Khodyreva, E.A. (2017) 'Problems of Risk Management of Innovative Educational Projects', *Concept*, no. 2, pp. 165-172.
- Liang, K. et al. (2017) "Online Behavior Analysis-Based Student Profile for Intelligent E-Learning" // *Journal of Electrical and Computer Engineering*, vol 2017.
- Matveenko, Yu. I. (2012) 'Modern approaches to the study of risk', *Izvestiya TulGU. Humanitarian sciences*, no. 1-1, pp. 165-173.
- Monk, D. (2005) 'Using data mining for e-learning decision making', *Electronic Journal of e-Learning*, no. 3 (1), pp. 41-54.
- Muller, H. (2007) *Compilation of mental maps. Method of generation and structuring of ideas*, Moscow: OMEGA-L.
- Nozdrina, L.V. (2012) 'Risk management of e-learning projects', *OTO*, no. 1, pp. 395-413.
- Panyajamorn T. et al. (2018) "Effectiveness of E-Learning Design and Affecting Variables in Thai Public Schools" // *Malaysian Journal of Learning and Instruction*, vol 15. no. 1, pp. 1-34.
- Pena-Ayala, A. (2014) 'Educational data mining: A survey and a data mining-based analysis of recent works', *Expert systems with applications*, no. 41(4), pp. 1432-1462.
- Rawat, B., & Dwivedi, S. K. (2019). "Discovering Learners' Characteristics Through Cluster Analysis for Recommendation of Courses in E-Learning Environment" // *International Journal of Information and Communication Technology Education*, vol 15, no 1, Article 4, pp. 42-66.
- Romero, C., Ventura, S. (2006) *Data mining in e-learning*, Witpress Boston.
- Romero, C., Ventura, S. (2013) 'Data mining in education', *Wiley interdisciplinary reviews. Data mining and knowledge discovery*, no. 3(1), pp. 12-27.
- Ruggeri, K., Farrington, C., Brayne, C. (2013) 'A global model for effective use and evaluation of e-learning in health'. *Telemedicine and e-Health*, val 19, no 4, pp. 312-321.
- Scheuer, O., & McLaren, B. M. (2012). 'Educational data mining'. In *Encyclopedia of the Sciences of Learning*, Springer US, pp. 1075-1079.
- Sultanov, I.A. (2016) 'Action Plan for Project Risk Management', *Projectimo* [online] Available at: <http://projectimo.ru/upravlenie-riskami/riski-proekta.html>
- Wu, Q. et al. (2016) "Clustering of online learning resources via minimum spanning tree" // *Asian Association of Open Universities Journal*, vol. 11, Issue: 2, pp.197-215.
- Wu, Pengfei, et al. (2018) "Using a Learner-Topic Model for Mining Learner Interests in Open Learning Environments" // *Journal of Educational Technology & Society*, vol. 21, no. 2, pp. 192-204.