Вісник Тернопільського національного технічного університету
https://doi.org/10.33108/visnyk_tntu
Scientific Journal of the Ternopil National Technical University
2023, № 1 (109) https://doi.org/10.33108/visnyk_tntu2023.01
*ISSN 2522-4433. Web: **visnyk.tntu.edu.ua***

UDC 004

TECHNOLOGIES FOR DESIGNING AND PROGRAMMING BIG DATA IN E-LEARNING

Roman Khrabatyn; Viktoriia Bandura; Natalia Shkolna; Yuri Khrabatyn

Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine

Summary. *Recently, e-education around the world is developing rapidly and the main problem is the timely provision of students with quality educational information. A significant impetus for this is the global epidemic of covid-19.*

The problem of implementing e-education cannot be solved without analysing the large flow of information coming into the information environment of e-education from participants in the educational process – students, teachers, administration, etc. In this environment, there are a large number of different types of data, both structured and unstructured, which are difficult to process by traditional statistical methods. The aim of the study is to show that the development and implementation of successful e-learning systems requires the use of new technologies that would allow the storage and processing of large data streams.

Large amounts of disk space are required to store large data. It is shown that to solve this problem it is expedient to use cluster technology NAS (Network Area Storage), which allows to store information of educational institutions on NAS - servers and to have access to them from the Internet. To process and personalize Big Data in the e-learning environment, it is proposed to use technologies MapReduce, Hadoop, NoSQL and others. The article provides examples of the use of these technologies in the cloud environment. These technologies in e-learning make it possible to achieve flexibility, scalability, accessibility, security, confidentiality and ease of use of educational information.

Another important problem of e-learning is the discovery of new, sometimes hidden, relationships in big data, new knowledge (data mining), which can be used to improve the educational process and increase the efficiency of its management. To classify electronic educational resources, identify patterns (patterns) of students with similar psychological, behavioural and intellectual characteristics, the development of individualized curricula in the article it is proposed to use methods of big data analysis.

The article shows that to date, many software applications have been developed for big data mining. These software products can be used for classification, clustering, regression and network analysis of educational information. The application of these methods in e-education will allow teachers to receive timely information about students, to respond quickly to any changes in the learning process, to make timely changes to educational content. The obtained results of the research are offered to be used for development of recommendations at creation of electronic courses in higher and secondary educational institutions of Ukraine.

Key words: *e-learning, big data, big data analysis, data personalization, big data management, MapReduce, Hadoop, NoSQL, data mining in e-learning.*

https://doi.org/10.33108/visnyk_tntu2023.01.072 *Received 09.01.2023*

Introduction. Recently, e-learning has become a trend in the education systems of many countries. Thus, in the United States, e-learning has become the most promising strategy in the national education system. This is evidenced by the data of the American e-education consortium «Sloan», according to which, in general, in the fall of 2014, 5.8 million students studied remotely, of which 2.85 million studied on-line from all courses, and 2.97 million – for some courses of the curriculum.

Globally, seven trillion dollars were spent on the development of e-learning in 2011 and, according to analysts, each year will increase by 25% annually.

Over a long period of time in the systems of electronic education of educational institutions accumulates a huge amount of information about various aspects of the educational

 Corresponding author: Roman Khrabatyn; e-mail: romankhr@yahoo.com

process: students, their success and attendance, teachers and their research, educational and administrative activities, educational content (text, audio, video), etc. This data must be efficiently stored, processed and analyzed. New technologies, often referred to as Big Data technologies, are needed to process large archives and large data streams.

The term Big Data refers to large and complex data sets that can be structured or unstructured and take up a very large amount of disk space. In the field of e-learning, big data covers three aspects: volume, speed and diversity.

Data Mining in e-learning for big data analysis. Data Mining technologies began to develop in the middle of the last century and only at the beginning of this century these technologies began to be used in education. One of the founders of the use of these technologies in education is Ryan Baker (Ryan Baker) – a professor at Columbia University [8]. These technologies are called EDM (Educational Data Mining). During this period, in connection with the increasing use of information technology in education, the amount of processed information increases sharply, the algorithms for processing this information are improved.

EDM technology is based on the concept of patterns of behaviour and personal qualities of students [9]. The use of these technologies in the field of education allows to find out which subjects cause great difficulties for students, which tests they cope better with, what form of classes they prefer, which topics they are most interested in and how to optimally build a curriculum so that students acquire those competencies which he will need in the field of his future professional activity.

An example of the use of EDM can be the following task: Are there examples (templates) of grades obtained by graduates who would then be able to find a job that meets their requirements within a short time after graduation.

The main objectives of the use of Data Mining in education are: 1. Classification – assigning objects (observations, events) to one of the previously known classes. Many different models are used for classification in Data Mining: neural networks, decision trees, the method of k-nearest neighbours [10].

Mathematically, the classification problem can be written as follows. There are some set {X} descriptions (characteristics) objective and, and a set of classes {Y}.

There is some objective function «f» that converts «X» to «Y» in the training sample {Xm}. Xm = {(x1, y1), (x2, y2), ..., (xm, ym,)}, where x1, x2, ... xm are vectors of feature features of objects, and y1, y2, .. ., ym is the name of the classes to which the corresponding sample objects belong.

You need to build an algorithm that converts the set X to Y, able to assign a new (non-sample) arbitrary object X to one of their classes B.

For example, the degree of similarity of objects, and hence the probability of their belonging to the same class, can be determined on the basis of the distance between their points in the feature space (the method of k-nearest neighbours). The smaller the distance between the feature vectors, the more similar the corresponding objects are.

Another method used for classification problems is the decision tree method, which is designed to break down the original data into groups until sets consisting of homogeneous (similar) data are obtained. Graphically, it can be represented in the form of a tree (hierarchical) structure, in the nodes of which decisions are made and there is a branching (English: branching) – division into branches (English: branches), depending on the choice made. In the nodes there is a branching of the process, ie its division into so-called branches, and the final (or, what is the same, terminal) nodes are called leaves (English: leafs, leaf nodes), in which the final result (decision). The data of the end node belongs to one class.

Examples of the use of classification problems in e-education are:

1. Classification of electronic educational resources (by functional feature that determines the value and place of educational resources in the educational process, by the text

of the resource, by the nature of the information presented, by the form of presentation, by purpose, etc.); classification of test tasks (according to different levels of difficulty, taking into account the individual pace of work, taking into account the individual capabilities of the student). When solving this task, it is possible to adjust the number of proposed tasks depending on the level of development of the student.

2. Regression, including forecasting tasks. The use of regression methods will simulate the impact of one of the parameters on another, to establish the dependence of the output parameters (objective functions) on the input variables (factors). This allows you to find out if you really achieve the desired result if you change the value of the selected parameter. Solving regression problems will allow to predict the results of final exams, the level of competencies of the graduate, the demand in the labour market and the level of their wages after employment. On the basis of which it is possible to identify the degree of influence on the formation of such factors as the need for specialists; university resources (including funding), the degree of introduction of information and telecommunications technologies in the educational process, the level of staffing of universities, teachers' salaries, etc.

3. Clustering – the division of the whole set of objects (observations, events) that are closest to each other in a number of features or properties, into clusters. Similar objects should appear within each cluster, and objects should be different in different clusters.

As a result of application of the cluster analysis it is necessary to solve the following tasks:

1) Select objects to cluster.
2) Identify the many criteria by which objects will be evaluated in the sample.
3) Apply one of the methods of cluster analysis to create groups of similar objects (clusters).
4) Visualize the results of the analysis.

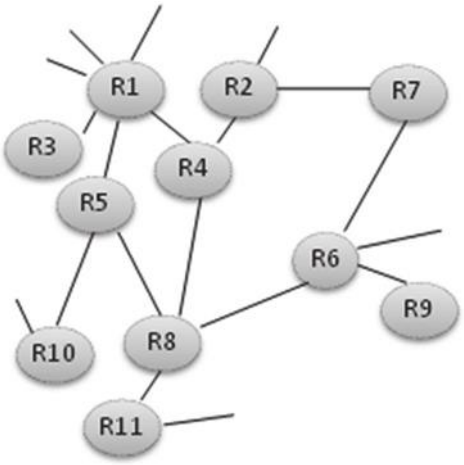

Figure 1. A fragment of visualization of student interaction in a social network

After obtaining and analysing the results, it is possible to adjust the selected metric and clustering method to obtain the optimal result. Clustering tasks in education can be used, for example, to identify students with similar psychological, physiological, behavioural, and intellectual characteristics. It is possible to identify how these behavioural patterns (patterns) affect success in different activities, which teaching methods are effective in relation to students with different stereotypes of thinking and psyche. Based on the clustering data, it will be possible to develop individualized curricula for individual groups of students, taking into account the duration of training, the trajectory of the material, the degree of complexity of tasks and other characteristics of the discipline.

4. The next important area of big data analysis is the analysis of social network data. The fact is that most of the time students spend outside of school – in social networks, where, communicating with peers, exchange information, collaborate, create shared online content, which can then be used in the educational process. Due to the fact that students are more comfortable in the social network than in reality (not under the control of the teacher), the information extracted from social networks, unlike other sources, may be more objective. From social networks, you can get information about the student's connections with his peers, his interests, movement on the network, activity (frequency of entry into the network and time of his stay in the network), etc.

The analysis of social networks is based on the mathematical theory of graphs, which is presented in the work of the Hungarian mathematician Erdos [11]. Mathematically, a network is a set of nodes (in our example, students of educational institutions) connected by lines that characterize the relationship between nodes. Each relationship connects several nodes. The first fundamental studies of social networks date back to 1979 and are reflected in the work of Wellman [12]. They developed cluster modelling algorithms and basic metrics for the analysis of social networks.

The first step in network analysis is data visualization. Visualized graphs allow you to identify the nodes that are closest, to find dense clusters of activity. One of the main parameters of the graph is «Degree centrality», which expresses the ratio of the number of strings of a certain node to the total number of other nodes. If for some node this parameter is equal to 1, it means that this node is connected to all other nodes of the network, if it is equal to 0, then this node is isolated. This indicator shows the degree of «celebrity» of the node, shows that the student has a great influence on others. Figure 2 shows a fragment of a visualized graph for analysing student interaction on a social network. The figure shows that the user of the network R1 has a higher degree of centrality than the user R3 (R3 leaves fewer communication lines than the node R1)

Software applications for data mining in e-learning. To date, many software applications have been developed for data mining. Thus, for classification problems, an algorithm called C5.0 is widely used, which is a standard decision-making procedure developed by programmer J. Ross Quinlan [13]. There is a free version of the package on the Internet (http: // www.rulequest.com), which implements this algorithm. With this package you can solve the problem of classification from different areas, including e-education, you can break down educational information on the basis of a large number of levels. This package can be used to analyse not only numerical but also nominal data; provides processing of the missed data. Only the most important features of objects are used to build a decision tree (selects from many factors only those that strongly influence the result of classification). The weakness of the application C5.0 is that the classification tasks require a relatively small sample size, even small changes in the training sample greatly affect the result. In general, this package is considered more efficient than other applications for classification.

For classification problems, another algorithm called CART (classification and regression trees) is widely used - for classification and regression using the decision tree [14]. Unlike C5.0, in the CART algorithm, decision nodes have two branches (binary representation of the decision tree). The result of the classification may be, for example, the answer to the question:

«Will the one who is successfully passed the exam pass or not?»

The next widely used algorithm for classification problems is the SVM (Support vector machine) method. For classification problems, this algorithm uses a hyperplane to divide the data into two classes. That is, if we have a set of data about each student: grades received for the semester, missed classes, the level of activity in the classroom, the degree of homework,

etc., you can use this data as input to the algorithm. Each of these parameters is a measurement in n-dimensional space. SVM displays these parameters in this space and finds a hyperplane to divide this data into classes.

Algorithms are widely used for data clustering problems: k-means and c-means. The first of these algorithms is used for hierarchical clustering problems. In hierarchical clustering, at the very beginning, each object is placed in a separate cluster, then, these clusters are combined into larger clusters, a system of nested partitions is built. That is, at the output we get a tree of clusters, the root of which is the whole sample, and the leaves – the smallest clusters. The optimality of partitioning into clusters is determined by the requirement of minimizing the rms error of partitioning (k-means).

The second algorithm is used for fuzzy clustering, when each object is associated with a set of real values that show the degree of assignment of the object to the cluster with a certain probability.

The next algorithm used for data clustering purposes is the EM-algorithm (Expectation-Maximization), which is iterative. In the first step of iteration (expectation) the probability of belonging of each data point to a cluster is calculated, in the second step – (Maximization) model parameters are updated according to the cluster distribution carried out in the previous step. Based on the EM algorithm, we can make assumptions about the contents of the cluster, and which cluster should include the new data. The software implementation of the algorithm can be found at http: // www. mathsisfun.com/data/index.html.

MySpace software application is used for Social Network Analysis (SNA) [15]. The analysis of social networks is aimed at studying the relations between people, considers social interaction in terms of network theory. These terms include nodes (individuals within a network) and links or references (representing relationships between individuals).

MySpace can be used to interpret and analyse the structure and relationships of students in solving common problems or in interactions with different means of communication.

With the help of log files, you can track information about those who study on social networks (Twitter, Facebook, etc.) – measure the number of messages in chats and links to different sections of the research topic, contacts between participants and the number of questions asked by teachers.

Cloud technologies for storing big data e-learning. Large amounts of disk space are required to store large data. Other projects of Hadoop technology are applications: spark apache and apache storm [19]. These programs make it easy and reliable to work with NAS (Network Area Storage), which involves connecting storage devices directly to a local or distributed computer network that uses the TCP/IP protocol. This network allows users to store files on NAS servers and share them with a browser or its network address. Operating system The NAS cluster infrastructure consists of several interconnected storage (repositories) of information, allowing users to share information available there and search. Figure 3 shows the NAS architecture.

According to experts, the volume of the NAS technology market will reach $ 7 billion by 2017. More than others, these technologies are used by the United States and Western Europe, but in the coming years in the Asia-Pacific region is also expected to grow by 14.1% [16].

Given the vast amount of information that is generated every day around the world, it is not surprising that organizations are looking for more efficient and cheaper storage devices, which are now networked storage, to which you can add new drives if necessary. However, it is already clear that with the advent of the era of cloud technology, big data analysis and the Internet of Things, real-time access to data requires a new approach to both data storage optimization and access to that data.

Modern cloud technologies to support the requirements of "Big Data" storage and software offer users storage optimization, security, flexible delivery methods and scalable infrastructure.

Clouds can contain not only huge amounts of raw data, but also data in their original format. New technologies allow them to be processed when needed. For example, Java-based Hadoop allows analysts to store vast arrays of data by hosting them on a large number of low-cost servers and then, using MapReduce on a Java virtual machine (JVM), to coordinate, aggregate, and process data [17].

In connection with the above, it should be noted the joint project of the Madrid Research Institute IMDEA Networks, the Polytechnic University of Madrid and the King Juan Carlos University project Cloud4BigData [18]. The project aims to integrate and integrate differentiated and specialized technologies into a single, unified platform, takes into account the requirements of cloud data technology, «Internet of Things» and «smart» – technologies and allows e-learning to achieve: flexibility; scalability; accessibility; Processing in parallel handles unlimited real-time data streams, gradually zoom in and save the infrastructure without data loss. Spark is a project of the University of Berkeley Laboratory (2009 USA), used to process unstructured and poorly structured data. It consists of a kernel and a number of applications for query processing (Spark SQL), settings for processing streaming distributed data (Spark Streaming), a set of machine learning libraries (Spark MLlib), software for distributed graph processing (GraphX).

In addition to Apache Spark and Storm, another project of interest to Java developers – DeepLearning4J [20], designed to build a machine learning library for Java and Scala, integrated with Hadoop and Spark, can import data from the network and create a multilayer neural network, use learning algorithms with and without a teacher to solve problems. This project can be used to solve the following e-learning tasks: face or image recognition; voice search; language recognition (converting it into text); regression analysis of data, etc.

All these tasks require the processing of large amounts of data using modern technologies and software, such as Map Reduce, NoSQL, Hadoop and others.

Conclusion. The concept of big data and methods of their analysis have recently been used to manage enterprises in the manufacturing sector. These technologies can also be applied to e-learning, including data processing decision-making, financial planning and student performance monitoring. Big data allows you to save the learning experience, give a picture of each student's learning. By analysing this information, with the help of Data Mining, e-course specialists can develop individual learning trajectories and adapt the learning process to meet the needs of each student. Data analysis will improve the model of the student, will allow researchers to obtain detailed information about the characteristics of the student or his condition, such as knowledge, motivation, to study what factors affect the increase of learning material, and what hinders this process. Data Mining in e-education will allow teachers to receive timely information about students and respond quickly to any changes in the learning process, make timely changes to educational content.

References

1. URL: www.sloan.com.
2. Pat Nakamoto. BIG DATA: The revolution that is transforming our work, market and world. Data Analysis. Kindle Edition, 2013. P. 212.
3. Franks B., Taming the big data tidal wave. John Wiley & Sons, Inc. 2010. P. 341.
4. Mayer-Shenberger V., Kuk'er K. Bol'shie dannye. Revolyutsiya, kotoraya izmenit to, kak my zhivem, rabotaem i myslim. Moscow: Izdatel'stvo "Mann, Ivanov i Ferber", 2014. P. 208.
5. Badarch Dendev. Informatsionnye i kommu- nikatsionnye tekhnologii v obrazovanii. Moscow: IITO YuNESKO, 2013. P. 320.
6. Ian H. Witten, Eibe Frank. Data Mining: Practical machine learning tools and techniques. "Morgan Kaufmann", 2015. P. 229. https://doi.org/10.1109/MIS.2014.42

7. Baker R. S. Educational data mining: An advance for intelligent systems in education. IEEE Intelligent Systems, 2014. 29 (3). P. 78–82. https://doi.org/10.1109/MIS.2014.42
8. Bishop Ch. Pattern Recognition and Machine Learning. Series: Information Science and Statis- tics. 2006. T. XX. 740 p.
9. Dyuk V. A., Samoylenko A. P. Data Min- ing: uchebnyy kurs. Saint Petersburg: Piter, 2001. P. 312.
10. Erdos P., Renyi A. On the evolution of ran- dom graphs. Publication of Mathematics Institute Hungary Academy of the Science. 1960. Vol. 5. P. 17–61. https://doi.org/10.1086/226906
11. Wellman B. The community question. American Journal of Sociology. 1979. 84. P. 1201–1207. https://doi.org/10.1086/226906
12. J. Ross Quinlan: Decision Trees and In- stance-Based Classifiers. The Computer Science and Engineering Handbook. 1997. P. 521–535.
13. Zaytseva T. Pusnaya O. Veroyatnostnye derev'ya resheniy: Programmnaya realizatsiya vreshenyy zadach klassyfykatsyy y prohnozyrovanyia. Lambert Academic Publisher, 2014. C. 96.
14. URL: https://myspace.com. https://doi.org/10.1504/IJTEL.2012.051816
15. Ferguson R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning. 4 (5/6). P. 304–317. https://doi.org/10.1504/IJTEL.2012.051816
16. Asha T, Shravanthi U. M, Nagashree N, Monika M, Building Machine Learning Algorithms on Hadoop for Bigdata. International Journal of Engineering and Technology. 2013. Vol. 3. No. 2. P. 143–147.
17. Qi Zhang, Lu Cheng and Raouf Boutaba. Cloud computing: stateof-the-art and research challenges. Journal of Internet Services and Applications. 1 (1): 2010. P. 7–18. https://doi.org/10.1007/s13174-010-0007-6
18. URL: www.storm.apache.org, URL: www. spark.apache.org.
19. J. Schmidhuber. Deep learning in neural networks: An overview. Neural Networks, 61: 2015. P. 85–117. https://doi.org/10.1016/j.neunet.2014.09.003
20. Belaya kniga Ed. N.V. Tikhomirovoy. M.: Minobrnauki, 2014. P. 137.

Список використаних джерел

1. URL: www.sloan.com.
2. Pat Nakamoto. BIG DATA: The revolution that is transforming our work, market and world. Data Analysis. Kindle Edition, 2013. P. 212.
3. Franks B., Taming the big data tidal wave. John Wiley & Sons, Inc. 2010. P. 341.
4. Mayer-Shenberger V., Kuk'er K. Bol'shie dannye. Revolyutsiya, kotoraya izmenit to, kak my zhivem, rabotaem i myslim. Moscow: Izdatel'stvo "Mann, Ivanov i Ferber", 2014. P. 208.
5. Badarch Dendev. Informatsionnye i kommu- nikatsionnye tekhnologii v obrazovanii. Moscow: IITO YuNESKO, 2013. P. 320.
6. Ian H. Witten, Eibe Frank. Data Mining: Practical machine learning tools and techniques. "Morgan Kaufmann", 2015. P. 229. https://doi.org/10.1109/MIS.2014.42
7. Baker R. S. Educational data mining: An advance for intelligent systems in education. IEEE Intelligent Systems, 2014. 29 (3). P. 78–82. https://doi.org/10.1109/MIS.2014.42
8. Bishop Ch. Pattern Recognition and Machine Learning. Series: Information Science and Statis- tics. 2006. T. XX. 740 p.
9. Dyuk V. A., Samoylenko A. P. Data Min- ing: uchebnyy kurs. Saint Petersburg: Piter, 2001. P. 312.
10. Erdos P., Renyi A. On the evolution of ran- dom graphs. Publication of Mathematics Institute Hungary Academy of the Science. 1960. Vol. 5. P. 17–61. https://doi.org/10.1086/226906
11. Wellman B. The community question. American Journal of Sociology. 1979. 84. P. 1201–1207. https://doi.org/10.1086/226906
12. J. Ross Quinlan: Decision Trees and In- stance-Based Classifiers. The Computer Science and Engineering Handbook. 1997. P. 521–535.
13. Zaytseva T. Pusnaya O. Veroyatnostnye derev'ya resheniy: Programmnaya realizatsiya vreshenyy zadach klassyfykatsyy y prohnozyrovanyia. Lambert Academic Publisher, 2014. C. 96.
14. URL: https://myspace.com. https://doi.org/10.1504/IJTEL.2012.051816
15. Ferguson R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning. 4 (5/6). P. 304–317. https://doi.org/10.1504/IJTEL.2012.051816
16. Asha T, Shravanthi U. M, Nagashree N, Monika M, Building Machine Learning Algorithms on Hadoop for Bigdata. International Journal of Engineering and Technology. 2013. Vol. 3. No. 2. P. 143–147.
17. Qi Zhang, Lu Cheng and Raouf Boutaba. Cloud computing: stateof-the-art and research challenges. Journal of Internet Services and Applications. 1 (1): 2010. P. 7–18. https://doi.org/10.1007/s13174-010-0007-6
18. URL: www.storm.apache.org, URL: www. spark.apache.org.
19. J. Schmidhuber. Deep learning in neural networks: An overview. Neural Networks, 61: 2015. P. 85–117. https://doi.org/10.1016/j.neunet.2014.09.003
20. Belaya kniga Ed. N.V. Tikhomirovoy. M.: Minobrnauki, 2014. P. 137.

УДК 004

ТЕХНОЛОГІЇ ПРОЕКТУВАННЯ ТА ПРОГРАМУВАННЯ ВЕЛИКИХ ДАНИХ В ЕЛЕКТРОННІЙ ОСВІТІ

Роман Храбатин; Вікторія Бандура; Наталія Школьна; Юрій Храбатин

Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ, Україна

Резюме. *Останнім часом електронна освіта в усьому світі стрімко розвивається. Основною проблемою стає своєчасне забезпечення студентів якісною навчальною інформацією. Значним поштовхом для цього є глобальна епідемія covid-19. Завдання впровадження електронної освіти неможливо вирішити без аналізу великого потоку інформації, що надходить в інформаційне середовище електронної освіти від учасників освітнього процесу – студентів, викладачів, адміністрації і т.д. У цьому середовищі існує велика кількість різних типів даних як структурованих, так і неструктурованих, опрацювання яких важко здійснити традиційними статистичними методами. Мета дослідження – показати, що для розроблення і упровадження успішних систем електронного навчання необхідно використовувати нові технології, які б дозволили зберігати й опрацьовувати великі потоки даних. Для зберігання великих даних потрібен великий обсяг дискової пам'яті. Показано, що для вирішення цієї проблеми доцільно використовувати кластерну технологію NAS (Network Area Storage), що дозволяє зберігати інформацію навчальних закладів на NAS – серверах і мати до них загальний доступ з Інтернету. Для опрацювання і персоналізації великих даних у середовищі електронної освіти запропоновано використовувати технології MapReduce, Hadoop, NoSQL та інші. Ннаводяться приклади використання цих технологій у хмарному середовищі. Ці технології в електронній освіті дозволяють досягти гнучкості, масштабованості, доступності, безпеки, конфіденційності та простоти використання навчальної інформації.Показано, що на сьогодні розроблено безліч програмних додатків для інтелектуального аналізу великих даних. Ці програмні продукти можна використати для класифікації, кластеризації регресійного й мережевого аналізу навчальної інформації. Застосування цих методів в електронній освіті дозволить педагогам своєчасно отримувати інформацію про учнів, оперативно реагувати на будь-які зміни процесу навчання, своєчасно вносити зміни в навчальний контент. Отримані результати дослідження запропоновано використовувати для вироблення рекомендацій при створенні електронних курсів у вищих і середніх навчальних закладах України.*

Ключові слова: *електронна освіта, великі дані, аналіз великих даних, персоналізація даних, управління великими даними, MapReduce, Hadoop, NoSQL, data mining.*

https://doi.org/10.33108/visnyk_tntu2023.01.072 *Отримано 09.01.2023*