Big data: some ways to solve the problems of higher education

O A Denisova, G A Kunsbaeva and A S Chiglintsiva

Ufa State Petroleum Technological University, 1, Kosmonavtov st., Ufa, 450062, Russia

E-mail: denisovaolga@bk.ru

Abstract. The article deals with issues related to the transition to digitalization of higher professional education, as well as with the widespread use of distance education technologies. The analysis of methods for monitoring the quality of education of university students is carried out. There are five types of data that require the application of big data analysis methods: personal data; data on the interaction of students with and between e-learning systems; data on the effectiveness of educational material; administrative data; forecasts. To solve a number of problems, we propose the use of Education Data Mining technology, which will allow us to design digital education management systems based on data and ways to systematize them for making organizational, pedagogical and managerial decisions. The analysis of educational data is based on the prediction of models, the identification of learning structures, and the identification of relationships. The possible ways of developing the individual trajectory of a student studying at the university, his further career.

1. Introduction

In connection with the transition to the digital economy, the development of the methodology and technology of big data analysis, the revision of the professional development of the teaching staff, there is a need to apply the methods of big data analysis in educational activities. This is mainly due to new approaches in the field of higher education, which focuses on the individual characteristics of students and the development of each student individually (individually oriented approach). New technologies in the field of big data analysis are becoming a unique tool for teaching each student. This provides new opportunities for improving both the effectiveness of education and the management at different levels of the education system itself.

Big data, as a concept in the scientific world, began to be talked about in the 40s of the last century. According to the researcher and entrepreneur Gil Press [1-3], the growth of data volume, and, consequently, the problem of data processing, appeared in 1941, which coincided with the publication of the term "explosion of information" in the Oxford Dictionary. "Big data – quantities, signs or symbols that a computer operates with, and can be stored and transmitted in the form of electrical signals, recorded on magnetic, optical or mechanical media" [2; 3].

A broad public discussion of the phrase "Big Data" (Big Data) began after the September 3, 2008, special issue of the scientific journal "Nature" [3; 4], in which Clifford Lynch noted the explosive growth in the volume of world information and used the term "Big Data". In describing this term, he referred to any heterogeneous data sets exceeding 150 gigabytes per day, and suggested that new tools and technologies under development would help expand this concept. Thus, one of the important issues that will be discussed in the scientific community in the coming years is the importance and
promising possibilities of processing and applying large amounts of data for the global economy and science.

In 2012, D. Boyd and K. Crawford define "big data" as a technical, scientific, and cultural phenomenon that includes three components: technology, analysis, and mythology. Technology refers to maximizing the computing power and complexity of algorithms for collecting, analyzing, linking, and matching huge amounts of data. The analysis involves creating sets of big data images to define economic, social, legal, and technical standards. Mythology is the representation of large data sets into a higher form of knowledge and information that can generate ideas that were not previously possible [5].

Big Data Science is a set of methods and technologies that allow you to extract information from large amounts of data [6; 7]. Big data is a specific technology that allows you to analyze and extract new knowledge from unstructured data sets. Big data allows you to use a set of mathematical and statistical tools to form a universal data structure and extract useful information, using it in a wide variety of fields of activity. This requires a set of special knowledge and skills from researchers. Big data analysis can accelerate the solution of technical, economic, social, research, and educational problems. For example, big data allows you to automate information systems, helping to systematize information, reducing the routine part of the teacher's work to a minimum. Also, on the basis of big data, it is possible to analyze the state of the education system in order to improve its quality. Data is collected from search engines, Internet services, surveys, content analysis, and observations during training. The defining characteristics of big data are, in addition to their physical volume, related to the importance of information, its purpose and contribution to the information problem being solved. Therefore, from the point of view of the features of big data - the physical volume, the speed of data growth and the need for their rapid processing, the possibility of simultaneous processing of several types of data, depends on the choice of methods and technologies for analyzing big data.

Today, an urgent task is to combine into a single database various information about the education system and the quality of its work, collected at all levels of its organization. Such data should be generated automatically, without increasing the amount of work of the employees of the educational organization. It is important to provide access to the generated database for any participant in educational activities, as well as for managers at all levels to make management decisions. This will allow, mainly, to carry out timely monitoring of the state of education, taking into account the results of the analysis of the quality of education, which will allow to objectively assess the general state of Russian education, track trends in its development, determine the impact of standards and educational literature on the quality of education, identify areas that require improvement and improvement of the level.

2. Methods
In connection with the COVID-19 pandemic, which began in 2020, there was an urgent need to switch the training of students and schoolchildren to a distance format. Thanks to the development of the distance learning system, it has become possible to analyze information using big data methods, which, in our opinion, provide researchers with much more information than the traditional educational process. This is primarily due to the active use of data collection and storage technologies, as well as the growth of search queries of students in the electronic educational environment.

The analysis of educational data (Education Data Mining) is a field of research related to the use of data mining methods, machine learning and statistics of information produced by educational institutions [2; 6]. The analysis of data related to education allows you to determine patterns from the data generated in the learning process. For example, when students have access to the materials of the curriculum, there is a record of how many times and for how long students accessed this information, in what order the test tasks were completed, the e-book of grades provides data on academic performance. All this information helps to identify patterns in students' learning and, accordingly, draw conclusions about the organization of the educational process, identify possible problems and quickly eliminate them.
In the field of education, there are five types of data that require the introduction of big data analysis methods [4; 8; 9]:

- personal data;
- data on student interaction with and between e-learning systems (e-textbooks, e-courses, page display speeds, page returns);
- data on the effectiveness of the educational material (which student interacts with which part of the content, interaction results, educational results);
- administrative data (attendance, absences due to illness, number of lessons passed);
- forecasts (student participation in this activity, what is the probability of completion of the work).

Big data analysis allows you to work with the study of individual programs, configure training for each student individually and as a whole for the entire study group. It shows which student addresses which part of the lesson, how this interaction takes place, who showed interest, with whom and how he contacts in the learning process, at what stage of the learning process he needs help. In this way, the training itself becomes adapted and personality-oriented, purposeful and reliable, dynamic and thorough.

Educational analytics based on big data changes the perception of the format of educational programs. Based on the analysis of how the material is perceived by students, the conclusion is made: what was effective or ineffective. The result of this analysis is a change in the content of both the course being studied and the lesson as a whole. Therefore, the educational program is moving from the format of the approved text to online content, which is dynamically changing due to the analysis of data processed as a result of students ' interaction with online content [7].

If you monitor and evaluate both the educational process itself and its results, then the interest of the teacher and students increases. This is due to the fact that data analysis allows you to create an individual plan for each student, and the teacher's interest is shown in the ability to get information about groups of students, to have feedback with them (interesting / not interesting, difficult / easy), as well as in the efficiency of resource allocation. The evaluation of learning outcomes can be independent and / or collective based on all data from all interactions. New methods will undoubtedly become an integral part of educational analysis [10]:

- forecasting (the set of known data will allow you to predict the unknown you are looking for);
- structure and grouping detection method;
- network analysis.

All this will create conditions for the development and introduction of mass practice and modern technologies for the analysis and interpretation of big data in education in the educational system. For example, American colleges and universities send letters of invitation to applicants. Each university strives to attract more promising and prepared students. To facilitate the work of the admissions committee, analysts collect and process various types of data from applicants: academic performance, academic interests, grades on final papers, and others. Such activities have proven effective in more than a hundred educational institutions in the United States. In particular, Creighton University in Nebraska managed to save more than 30 thousand dollars on the admission company [3].

3. Results
In Russia, the use of big data analysis in education is at its initial level. The idea of the project is that it is necessary to develop a student ID card that combines several functions: a travel card and a student ID card, a test book and a pass to the territory, a reader's card, etc. With the help of the map, you can collect data about the time and location of the student. By creating a personal account, students can maintain their academic performance, communicate with teachers, find out the schedule of classes,
have access to the electronic library, to the university’s curriculum and the opportunity to contact them. All these services will allow you to collect and process data, generate additional recommendations for improving the learning experience in universities [11].

One of the most striking examples of the introduction of big data analysis methods in the field of education is the company Skillsoft, which provides educational materials to more than 20 million users on its online learning platform. Thanks to big data analysis, Skillsoft has managed to adapt the educational resources offered to each of its millions of customers, taking into account their skill level, specific needs and business interests. This allowed the company to achieve positive results in the field of user interaction with the educational environment [12-14].

It is obvious that access to big data helps to get instant feedback from students on lessons, determine the pace of implementation of the lesson plan, track the progress of students and understand their individual needs. Thus, Big Data identifies and helps to eliminate gaps in students’ learning, offers a tool for obtaining information about their academic performance and the conditions for training teachers and the management of an educational organization in real time. The platform analyzes each student’s academic performance, identifying strengths, weaknesses, and knowledge gaps, and generates appropriate sets of problems to help solve them. For example, if a student is ill or absent from class for some other reason, then after accessing the educational platform (website) in their personal account, they can study the desired topic using an electronic library or in the form of chat messages with their teacher. After reviewing the missed topic, you must pass the test and get the result in the form of a report, then additional points will be entered in the electronic journal of the educational platform. Digital education platform - information that enters the digital diary directly on the website. The system provides information, displays statistics, and provides all metrics.

Big data analysis opens up prospects for creating a new positive learning experience - all participants in the educational process will be able to freely exchange information with each other. Big data can help students choose educational products that match their personal and social needs.

Big data analysis, based on the user’s interaction with educational and social platforms, will allow you to determine the student’s abilities and aptitudes at an early stage. Further, an individual educational trajectory of the path can be built, which will contribute to the development of the necessary skills of students, determined taking into account their motivation and needs.

Big data analytics can enable an educational organization to track the professional progress of its graduates by targeting thousands of different sources, from their social media profiles and photos to posts in the media and on organizations’ websites. This will allow you to keep up-to-date information about the status and qualifications of graduates of educational organizations at all levels, support the graduate community through the management and provision of additional educational services, and bring the student portfolio to a new level of significance.

The free and unhindered circulation of unstructured data and their availability for analysis and processing of information open up opportunities for wider participation of students in processes that were previously considered the prerogative of administrative staff and managers only due to the conservative and traditional nature of the education system. This will allow not only to take into account the needs of the market in a particular region within the higher education system, but also to find new combinations of forms and solutions to improve the educational experience of all participants in the educational process.

4. Discussion
The methodology of big data analysis in the field of education is based on a systematic methodological approach. The system-methodological approach to the analysis of educational data (Educational Data Mining) includes a set of methods of data mining and information statistics produced by educational organizations and educational digital platforms. This approach allows us to design digital education management systems based on data and methods of systematization of educational data for making organizational, pedagogical and managerial decisions in education.

The methodology and technology of educational data analysis are based on a group of methods:
• Methods of forecasting the model (the model predicts the value of the value of interest from the values that acquire independent variables). For example, forecasting the results of the final certification or Olympiad achievements of students based on the analysis of data on current academic performance, participation in programs; forecasting profile selections of students based on data on pre-profile selections and participation in a design and research competition; forecasting Olympiad achievements based on statistics of the types and content of tasks to be solved.

• Methods for identifying learning structures (reveal the data structure, clustering algorithms). For example, the structure of the lesson in accordance with the features of the organization of training of students with different types of educational motivation, the organization of training of students with different types of learning difficulties, the results of the diagnosis of the quality of education.

• Methods for identifying relationships (establish relationships between variables in a data set with a large number of variables). For example, the relationship between the frequency of students' classes, including online, and students' academic results in the modules of the educational program, the relationship between the features of the organization of students' project activities in the classroom and the results of students' project activities.

Thus, three groups of forecasting methods based on data analysis, identifying the structure of data; determining the relationships between variables in data using digital data analysis and technology will help teachers and education managers to obtain metadata (data about data) to make decisions about improving the organization of education and improving the quality of educational outcomes. Improving the effectiveness of teachers, individualizing training, and developing students' skills appropriate for their future profession are all possible through the analysis and methodology of big data and the subsequent conscious application of the results.

5. Conclusion
With the development of Big Data technologies, higher education reaches a higher level, when their application allows you to identify students who find themselves in a situation of expulsion or deserve special incentives. These technologies will allow you to track such situations and help successful students to advance along an individual educational path. The analysis of data on the quality of training will help applicants choose the education and career that best suits their personal qualities, as well as their interest in a professional career.

Educational organizations have yet to learn how to use the methods of big data analysis, but also to find ways to widely apply them in the management of the quality of training. In the future, big data analysis will help implement a mechanism that allows effective real-time interaction between teachers and students, in-depth and comprehensive study of learning models implemented by educational organizations and optimize them based on new knowledge available through big data analysis. All this will help to balance the complexity of the educational process, predicting which students need additional help and support in the process of mastering the educational material. The introduction of new open educational resources will make it possible to accumulate large amounts of data, the analysis of which will contribute to improving the quality of education and the learning experience of students.

References
[1] Fiofanova O A 2020 Big data analysis in education: methodology big data analysis in education: methodology and technologies: monograph (Moscow: Publishing House "Delo" RANEPA) 200
[2] Kravchenko P A 2020 Analysis of the positive and negative consequences of the use of "big data" in the field of educational activities. Student 88 31
[3] Strunin A E 2020 Research on Big data analysis. The scientist's rostrum 4 37
[4] Vitchenko O V, Stryukov M B and Dashko Yu V 2019 Big data analysis as a method of
analytics in business and education. *Intellectual resources - regional development* 5 19

[5] Galoyan O T and Erokhina E V 2020 Methods and techniques of big data analysis. *Trends in the development of science and education* 62 24

[6] Agaev F T, Mammadova G A, Zeynalova L A and Melikova R T 2018 Application of big data analysis methods in e-education. *Innovations in education* 9 83

[7] Paskova A A 2019 Application of Big Data technologies in the educational process. *Pedagogical informatics* 3 23

[8] Nazarenko M. A. 2015 Technologies and methods of big data analysis. *International Journal of Experimental Education* 11 40

[9] Budnikova I K and Mardanova A M 2020 BIG DATA Technologies. *Information technologies in construction, social and economic systems* 3 49

[10] Ishchenko A A 2020 Big data as an information base for the analysis of the quality of education. *Scientific potential* 2 10

[11] Shunina L A and Usova N A 2020 BIG DATA-analysis of approaches to definition and application in education. *Informatization of education and methods of e-learning: digital technologies in education. Proceedings of the IV International Scientific Conference. In two parts. Krasnoyarsk* 266

[12] Denisova O A 2020 Motivation of technical University students to study physics and methods of teaching it in the context of a pandemic. *Journal of Physics: Conference Series* 1691 012025

[13] Kondratenko B A and Kondratenko A B 2018 Prospects for the use of big data in modern education. *Bulletin of the Baltic Federal University named after I. Kant. Series: Philology, Pedagogy, psychology* 1 117

[14] Gorelik O M, Bobrovsky S M, Malysheva E Yu 2016 BIG DATA technologies in the organization of the educational process of the university. *Bulletin of the Volga State University of Service. Series: Economy* 2 173