Big Data technology: assessing the quality of the educational environment

O A Denisova
Ufa State Petroleum Technological University, 1, Kosmonavtov st., Ufa, 450062, Russia
E-mail: denisovaolg@bk.ru

Abstract. The paper shows the prospects for the development of digital technologies of Big Data, which are proposed to be used as a means of qualitative assessment of education, as well as its effectiveness. The use of intelligent systems that work at the level of the best teacher in the future will be available in any educational organization. This will make it possible to overcome the "educational inequality" and reduce educational barriers for people with disabilities. If the methodology of traditional teaching is created based on the personal experience of the teacher, then the methodology based on big data analysis technologies is the result of the experience of a huge number of users. With the help of big data, you can create methods that are adapted to a large number of students; personalize content; choose learning modes, that is, learning becomes adaptive and personality-oriented. The educational program is transformed from the traditional approved text format to the format of personalized online content, which dynamically changes depending on requests. In addition, approaches to monitoring and evaluating both the educational process itself and the results of education are changing. As a result of the application of Big Data technologies (multidimensional statistical analysis), new methods of forecasting are proposed for use, when a combination of known data predicts the desired unknown; structure identification and clustering; network analysis. Big data is a huge volume of information, big data analysis methods are the systematization of this information, but they can never replace the teacher who gives his own and "reads" the emotions of the student.

1. Introduction
In the modern world, the ability to manage big data is at the forefront, this is due to the daily growth of information in various areas of human activity, and new opportunities for the use of digital technologies are emerging. It is necessary to process, systematize, and store huge amounts of data, which leads to new problems—an increase in the power of computer technology, as well as the speed of processing and transmitting signals [1; 2]. The term Big Data was introduced relatively recently, however, it is quickly being introduced into the modern education system. Based on big data technologies, it is possible to improve and simplify the principles of assessing the quality of the educational environment. Processing and storing information about the effectiveness of an educational environment with big data requires a lot of hard disk space. With the right approach, digital big data technology can simplify the process of tracking student markers and identifying problem areas of learning. You will be able to respond to changes in the learning process by extracting and analyzing data at an automated level to achieve flexibility, scalability, availability, security, confidentiality, and ease of use of training information. The capabilities of Big Data are not yet sufficiently used to...
improve the quality of educational activities. However, the rapid development of digital technologies can make Big Data an effective tool in assessing the effectiveness of education [3; 4].

2. Methods
The idea of "big data" (Big Data) appeared in the last century, however, the intensive growth of interest in the use of big data technologies in Russia arose in connection with the transition to the digital economy. Big data technologies are used in the social sphere, economy, trade, medicine, as well as in educational activities. According to search queries, they make it possible to analyze the purchasing power and interests of customers, monitor the situation and plan tourist trips to the most popular places, and, consequently, offer the consumer relevant information [5-7].

Due to the introduction of digitalization in the educational environment, this has both positive and negative sides. The positive aspects include the fact that the transition to an electronic format will allow you to make management and organizational decisions more effectively, and quickly post information for general access. Most students switch to individual training plans, that is, an individual trajectory of each student is built. In addition, in parallel with classroom classes, classes conducted in a remote format, there are now a huge number of online courses that you can take and get additional points when certifying the discipline you are studying [8; 9]. Big data technologies allow you to evaluate the knowledge gained by students in a remote format, keep records of academic performance. The positive aspects can also be attributed to the fact that children with a limited state of health can receive general secondary or higher education without leaving home [10; 11].

The role of full-time classes and an experienced teacher is difficult to overestimate. Each teacher plans and conducts their classes using their own experience or the experience of colleagues, but, unfortunately, the time of training sessions is limited, so it is not always possible to bring the material to each student within the class. Therefore, outside of the lessons, each student can additionally find the material they need using search queries. That is, in this case, computer systems are an additional and effective means for obtaining knowledge.

Big data technologies allow you to analyze and process search queries and the experience of a huge number of students and teachers, as a result of which you can offer an effective teaching methodology or technology, since they are the product of generalization of mass activities [12].

In addition to improving teaching methods, big data technologies allow you to create personal content according to the needs of each student. The negative aspect of using digital educational technologies is that the computer is an inanimate object. He can't track a person's reaction, and understand whether the student has learned the material. Therefore, a computer can never replace a teacher. It is only a means to gain knowledge.

3. Results
The use of any educational or computer technologies, including big data technologies, will not protect the teacher from communicating with the students on an emotional level, will not free them from empathy, from communication, from conversations that are not related only to the discipline being studied. The teacher in a simple conversation can motivate you to take action, share your life experience. The computer and the Internet in the "teacher-student" system are a tool for the teacher to acquire knowledge, obtain information, as well as a form of presenting this information in the form of presentations, animations, graphs, tables, and drawings. That is, the use of big data technologies allows you to: 1) choose the mode and form of training; 2) use personal content; 3) develop new teaching methods that are adapted to a huge number of students [13-16].

In the near future, big data methods will allow us to change the approaches in the teaching methods of higher education, and training will become more personalized. The student will be able to individually choose the disciplines that he would like to study, will receive individual homework, and will individually pass the test of the acquired knowledge. To some extent, all of the above already takes place in some universities. For example, at Ufa State Petroleum Technological University, students can choose and take online courses, receive homework, and after completing it, upload it to
the student's personal account, and the teacher checks it remotely. Knowledge verification is carried out by testing. The student, sitting at home, can, having received an individual password, take a test on this topic, after completing the test, the result is automatically reflected in the statement. The teacher can remotely view the result, open a ticket, and analyze the correctness of the answers, if necessary.

The second method of teaching, which is also implemented in our university, is the project method. In this case, depending on the task, students are grouped into groups of two or more people. Students receive a task, and after discussion, they distribute the tasks among themselves. Consultations and seminars are held to discuss problems and interim results. Next, the project is defended, the result is presented - a presentation, development, model, computer program, database, patent, etc., depending on the purpose of the project.

Big data technologies are used to monitor student performance. To do this, the deans conduct an intermediate certification of students in the middle of the academic semester, teachers fill out the statements in electronic form, then all the data is collected in a single table. After analyzing the information, the dean's office employees can predict the results of the session even before the session begins, and also take measures for lagging students before the session begins. Similarly, the results of the session are summed up. The statements are filled in electronically. Students who have passed the session without triples are sorted, and a list is formed to be included in the order for awarding a scholarship. This once routine work of the dean's office employees is done through the use of big data technologies.

Another positive aspect of using big data is the introduction of an electronic graduate portfolio. For example, if the electronic portfolio is loaded with all the achievements of a school student (his grades, participation in competitions, Olympiads, conferences, projects, volunteer movement, winning sports competitions, exhibitions, music competitions, etc.), then when choosing a university, taking into account his achievements, as well as his interests (knowing his personal content, search queries, etc.), the computer system will suggest universities where the applicant could enroll. Further, while studying at the university, this portfolio will be supplemented with new achievements, and upon graduation, either the employer will offer the appropriate position, or the graduate will be able to choose from the proposed list of vacancies compiled on the basis of big data analysis.

To assess the effectiveness of education, a quality assessment methodology is needed, which requires the definition of clear and precise criteria. But first you need to keep in mind that the formation has several values, none of which can be ignored when assessing the current state and making forecasts:

- Education - as a branch, as a whole;
- Education - as a training program;
- Education - how students will succeed in the future, their employment and standard of living;
- Education - as the availability of a tool that helps a person to develop;
- Education - as providing the necessary infrastructure.

For everyone-parents, students, teachers, school management, the assessment of the effectiveness of education consists of different factors that have different impact and significance. The main task is to create a tool that will give each user individually and together an indicator that is sufficient and necessary for further decision-making. And sometimes to encourage impulses, to stimulate initiative, to change, that is, to draw attention to problems that until then could have gone unnoticed.

4. Discussion
Identifying triggers is the first, fundamental step in solving a problem. This is the provision of information to the user briefly and clearly, which is estimated in certain sizes, coefficients, indicators of signals. Finding the average indicators that will be important for everyone, for example, parents, is a difficult but important task, although everyone has their own approach to training, education and, consequently, their own criteria for evaluating the effectiveness of achieving a particular result.
Therefore, first of all, it is necessary to determine the main, highest indicators, on which further models of assessing the current state and forecasts of further development should be based: school performance—grades; school performance—absenteeism; school performance — the quality of the tools provided to achieve success (evaluation of teachers, schools, educational materials); health – physical; health - emotional, that is, an assessment of the immediate environment in which the child is located.

The main complexity of the implementation is the number of systems that are data sources, and the number of events and information in them. There were both sources of relatively structured information (exam results; Olympiads; absenteeism; grades) and completely unstructured information (training program; Internet events; names and structure of additional education circles; clicks on an electronic whiteboard).

Therefore, only the stage of data collection of all indicators, taking into account the etl processes, integration of services (and often their development from scratch) and all relational and non-relational databases is very time-consuming and, moreover, a continuous process, since as a result, a request for new indicators and developments is immediately received. For example, the following activities are carried out:

- Monitoring of current academic progress and identification of deviations (received a bad grade; received a bad grade after illness; received a bad grade, and the entire class received good grades);
- Identify learning functions in school (favorite subjects that are well-preserved, that are not skipped);
- Identify which activities are good and bad (written work, answer on the blackboard);
- Building a circle of interests based on visiting circles and additional sections.

Development at the initiation stage: 1) monitoring of extracurricular, extracurricular activities; 2) identification of additional interests based on the interests of the immediate environment; 3) identification of psychological characteristics (extrovert / introvert);

Analysis and identification of patterns using multivariate statistical analysis is the second stage of solving the problem. This is, in fact, creating a "digital twin" of the student, that is, searching for similar ones and comparing them. The following cases are considered here: building school dependence on diseases; building a "portrait" of the student and class, that is, a comparison with classmates; building a rating of teachers, schools, and textbooks.

However, we first examined whether there is a correlation between weather attributes, class attendance, and the scores obtained, as well as the rate at which they change, meaning that not only natural values are analyzed, but also their derivatives.

The result is proven hypotheses:

- Poorly performing and excellent students are less ill than mediocre and good students;
- Junior classes are less sick than middle and senior classes, but in September, senior classes are sick more often;
- The average duration of the disease is two weeks;
- Gets sick on average twice a year;
- Schools get sick in different ways;
- Temperature is an insignificant factor (due to the specifics of the analyzed data - the date of diagnosis in the clinic is taken as the fact of the disease, and not the moment of the actual disease of the student).

Multivariate statistical analysis was used to construct the "portraits". Within its framework, the visualization of the average student's grades against the background of the class, parallel and school is
carried out in the form of a web (by subject), the result is an analysis of the student's level against the background of others:

- How many percent of students in the class (parallel) study better than you in the top 5 subjects;
- How many percent of students in the class (parallel) study worse than you in the top 5 subjects;
- The percentage of students in the class (parallels) with the same proportion of bad grades on the topic.

The result is useful for quickly understanding at what level a child is learning, implicitly comparing their results with the impersonal average student of the class or parallels.

Teacher rating-create a showcase with various indicators for the teacher, calculated on the basis of student performance data, circle data, statistics for the school and the city, here are some of them:

- Number of schools where the teacher works;
- List of schools where the teacher works;
- Average student score from the teacher;
- Number of students;
- Percentage of teachers with a higher average number of students;
- Percentage of teachers of the same subjects, the average score of students who are higher;
- Share of excellent results;
- The percentage of teachers whose share of excellent students is higher;
- The share of teachers of the same subjects, the share of excellent students who are lower;
- The percentage of failures in the lessons.

Building estimates and predictive models is the third stage of solving the problem. The further the planning horizon, the greater the uncertainty. How can you predict the outcome of final exams if the child is only in the middle grades? Or the average score? Or a set of subjects for which the child has a tendency to choose the next profession? How can I determine what factors may affect the incidence of the class in which your child is studying, getting a forecast of the growth of epidemics for the next month? If we cannot anticipate how our actions and decisions will affect the development of circumstances in the following hours and / or days, this will become a very difficult task in the long run.

First, there are more and more incoming flows, exogenous variables that we simply ignore (weather conditions, the distance of the educational institution from the place of residence, and the influence of modes of movement on the overall background of the child's health), due to lack of time, their volume and inexplicability.

Secondly, in this case, it is not you, but your children and/or students, that is, third parties, who are the resulting variable in the regression equation. And this means that there are more and more unknown variables for you.

Third, you may not have all the knowledge that others have historically accumulated in the industry. Even if you spend a lot of time studying the user experience, in most cases it is a large set of different, unstructured information of subjective opinions. Accordingly, scoring models and forecasts allow you to get information about the probability of a particular result, taking into account their conditions, but in an aggregate form that does not allow the disclosure of personal information of other participants in the process. At this stage, such cases are solved using:

- Performance forecast;
Personal recommendations on the pedagogical trajectory - adjustments, additions, the need to pay attention to "saturating" factors;
Personal recommendations on preventive measures for the health of the child, parents, and even the school as a whole.

5. Conclusion
The use of big data technologies and methods of their analysis is promising and opens up wide opportunities for their application. First of all, these technologies allow you to personalize your training. Based on the analysis of the user's personal content, select the information that is interesting to a particular person (a personally oriented approach). The presence of an electronic educational environment in the university, an electronic library catalog allows you to analyze requests, the time of their views, the frequency of requests, as a result, you can dynamically change the content of the material, therefore, the content of both individual disciplines and educational programs as a whole will change. Monitoring the results of the educational process allows you to identify both well-performing students and lagging students. This analysis can be carried out for individual disciplines, which will lead to a revision of the curriculum if necessary. You can analyze the performance of an individual group, stream, faculty, or a single student. If the student does not have time, timely identify the reasons, take action. You can also collect and analyze data on student attendance, identify patterns from the month or time of year, and the reasons for absence from classes. To analyze the educational process, new methods are used, such as forecasting, network analysis, clustering method, and structure method.

Thus, the use of big data analysis methods opens up new opportunities for their application in the educational process, allows you to reach a higher level, quickly respond and meet the needs of the modern information environment. The individual career trajectory of a person "school - university - enterprise" or in other words "student - student - specialist" is built.

References
[1] Silen D 2017 Fundamentals of Data Science, Big Data. Python and Data Science (Moscow: Peter) 354
[2] Elearning market trends and forecast 2017–2021 Retrieved from: https://eclass.teicrete.gr/modules/document/file.php/TP271/Additional%20material/docebo-elearning-trends-report-2017.pdf
[3] Mikhnev I P, Chelnokova A D and Reut A D 2018 Big Data technologies and their application in the field of modern higher education. Development of modern education: from theory to practice: Proceedings of the IV International Scientific and Practical Conference, Cheboksary 64
[4] What is Big data: collected all the most important things about big data Retrieved from: https://rb.ru/howto/cho-to-takoe-big-data/
[5] Big Data technologies will soon change higher education Retrieved from: https://mel.fm/novosti/7249138-bigdata
[6] Banko Yu A and Kokoreva A M 2017 Modern computer threats: what really threatens business? Priority directions for the development of education and science: Materials of the IV International Scientific and Practical Conference, Cheboksary 169
[7] Shatalova V V 2019 Big Data technologies – as a factor of competitiveness of a university teacher in the conditions of modern IT education. BIG DATA and Advanced Analytics = BIG DATA and high-level analysis: collection of materials of the V International Scientific and Practical Conference, Minsk 11
[8] Utemov V V and Gorev P M 2018 Development of educational systems based on Big Data technology. Concept 6 449
[9] Kuznetsov N V 2019 Online education: key trends and obstacles. E-Management 2 19
[10] Aptekman A and Kalabin In 2017 McKinsey & Company. Digital Russia: new reality 133
[11] Belyaev G Yu 2017 Potentially negative trends in the development of modern education and their impact on education as a social phenomenon. *Science today: problems and prospects of development: materials of the international scientific and practical conference* 89

[12] Ivanov V G, Kaibiyainen A A and Miftakhutdinova L T 2017 Engineering education in the digital world. *Higher education in Russia* 12 136

[13] 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

[14] *Informatization of education and methods of e-learning* Materials of the II International Conference Krasnoyarsk Retrieved from: https://bik.sfu-kras.ru/elib/view?id=BOOK1-ББК74/II%20741-504583

[15] *Big Data from A to Z. Part 1: Principles of working with big data, the MapReduce paradigm* Retrieved from: https://ai-news.ru/2017/11/big_data_ot_a_do_ya_chast_1_principy_raboty_s_bolshimi_dannymi_paradigma.html

[16] *Big data technologies in electronic education* Retrieved from: http://dx.doi.org/10.21686/1818-4243-2017-6-41-48