

Data Science and Big Data Technologies Role in the Digital Economy

Sergey V. Novikov

Moscow Aviation Institute (MAI), 125993, Volokolamskoe highway, 4, Moscow, Russia

Abstract – This article explores the role of Data Science and Big Data technology in the modern digital economy. The author states that large and medium companies from retail trade and service sector show increased interest in using them. These technologies are actively used by banks, mobile operators and large manufacturing companies to analyze data on equipment failures and to reduce downtime, which allows reducing costs. The role of Big Data technology is to be a liquid product and a necessary condition to increase the profitability of enterprises through personalized customer service and predictive analytics. For today's Russian digital economy, it is very important to legalize a single definition of Big Data and to achieve the emergence of special data exchanges.

Keywords – Big Data technology, digital economy, machine learning, innovative technologies, project management based on Data Science

1. Introduction

The technological era in which we are living now can be called “digital transformation”. Computers, telephones, and many other technologies have become cheap, portable, and fast, so they are everywhere. Today, everything is changing very fast. The main driver of these changes is the "digital" consumer. The speed of change in customer preferences, is increasing the demands on the quality, functionality, and design of products.

DOI: 10.18421/TEM92-44

<https://doi.org/10.18421/TEM92-44>

Corresponding author: Sergey V. Novikov,
Moscow Aviation Institute (MAI), Moscow, Russia.


Email: svnovikovmai@mail.ru

Received: 21 March 2020.

Revised: 04 May 2020.

Accepted: 10 May 2020.

Published: 27 May 2020.

 © 2020 Sergey V. Novikov; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The article is published with Open Access at www.temjournal.com

Impressions are now important to the consumer. Consumption patterns are changing, from which emerges the obvious trend: interaction via smart devices. Business operates at high speeds, increasing demands, saturation of traditional markets for companies, increased competition, and emergence of new technological competitors. Outdated technologies do not bring much profit for a long time, their modernization is expensive and requires relatively large resources for their maintenance [1].

The global market for Big Data is growing very fast. The most active data generators and consumers are the public sector, healthcare, manufacturing, and retail. In Russia, the entrance of Big Data is uneven and the banks are leading the way as users. In the coming years, insurance companies, petrochemicals, and metallurgy may join them. Analysts believe that the economic effect from the introduction of Big Data technologies will reach 300 billion rubles by 2024. The volume of the data is growing exponentially. In 2018, users and businesses around the world have generated 33 zettabytes of information, in 2025 there will be 175 zettabytes. The main growth drivers are entertainment platforms, video surveillance systems, Internet-connected equipment, and tools for productivity growth. They all operate with Big Data. Digital transformation increases the speed and flexibility of value creation processes and allows adaptation to the high pace of market changes and the requirements of the digital consumer. This increases the requirements to the competence of company staff, ability to work with new technologies, approaches, and methods of project management: development and operations (DevOps), continuous integration (CI), continuous delivery (CD), Scrum/Kanban, user experience (UX) / user interface (UI)-testing, etc. It also involves the introduction of flexible forms of campaign management, reducing hierarchy levels and improving the quality of communication interaction. As the processes of digital transformation take place everywhere, it is necessary to remember that Russia does not fall behind the general trend. At the same time, it is difficult to predict what will be the remote result of the digitalization of the society and the economy. In this situation it is very useful to choose

a common path along which all links of the state, economic, and social structure of the country will move in a coordinated way towards digital transformation. Undoubtedly, the national program "Digital Economy of the Russian Federation" will be an additional impetus in digital development [2].

2. Theoretical Basis

Both the introduction of new technologies and working with Big Data is transforming businesses. It is not only about a new position, Chief Data Officer (CDO), but also about the constant training of all the staff, from ordinary employees to top management. New realities have also changed the rules of the game and companies are accustomed: working on a project in the field of Data Science does not always have commercial benefits, but it has positive results for the team and subsequent tasks. Data is a key asset in the era of digital economy, as well as people and finances. But if the HR function is responsible for managing people and finances (controlling, risk management, treasury, and other departments), then the data have not been historically controlled by anyone. They seem to be common, but at the same time they are nobody's. Now it is changing, Data Governance practices are being implemented in companies, and the position of the CDO is being introduced. At the moment, the CDO is one of the top officials in the company [3]. He plays a lot of roles: expert in advanced technology, business strategy, and enabler of ideas related to data usage. He is also responsible for data quality and transparency, including various regulatory bodies. The CDO develops a strategy on how the organization's competence in data analysis will be developed and what skills to develop in employees. He also controls the availability of the necessary set of technologies in the organization, which allow efficient extraction, storage, processing, and visualization of the data [4].

As practice and researches show, there are three widespread variants where the CDO can appear in the company: department of direct subordination of the CDO, in subordination at the Chief Information Officer or inside other business department, for example, the financial one. The location of this department depends on many reasons such as: who was the first to offer the idea, what relations with business have been built, who at the moment of introduction was the readiest for it... The practice of implementing the role of the CDO comes from the West: the majority of companies with this position are located in the United States of America or Western Europe [5]. Typically, these are large

corporations that have a lot of data at their disposal and want to manage it more efficiently, as well as financial organizations that have been placed in a rigid framework of standards and regulations, especially after the global financial crisis in 2008. Some Russian organizations, especially subsidiaries of foreign companies or the companies whose shares are traded on major international venues, take over some of these standards and regulations. But this has not yet resulted in the widespread adoption of the CDO role in Russia [6].

Organizations now have a shortage of qualified staff: few educational programs train data analysts, and good professionals tend to migrate towards global Informational Technology industry leaders. There is a struggle for every digital talent. Before, large mainframes available only at enterprises were needed to solve computational problems, and it was difficult to develop them independently. Technologies are cheaper now, computing power is growing, and everybody needs an average laptop to find something in the subject area. The world is moving towards Open Source, packages for analysis can be downloaded for free and legally [7]. There are also educational programs from commercial and non-profit organizations. Any person, if they have free time and desire, can in a short time learn a level that allows them to apply for starting positions in almost any company. It is not enough just to approve the role of the CDO and Data Governance policy or to staff with high quality data-scientists. For effective work on Data Science initiatives, it is necessary to change the culture within the company [8].

The culture of businesses, which was formed before the wide spread of Data Science, is usually based on a conservative approach to risks, minimizing costs, and improving operational stability. Previously, companies tried to use only time-tested business models and technologies, and key competencies were often outsourced to save money. This occurred because the markets were formed [9]. It is impossible to carry out such a large-scale transformation only by sending individual employees to retraining courses. It requires a more fundamental and thorough approach. It requires deeper involvement and support from the Chief Officer of Human Resources (COHR), who oversees all aspects of the HR management and industrial relations policies, practices and operations for the high-tech organization. Using all the skills and capabilities that the HR has, he will help the organization to create a balanced and end-to-end development program that addresses all of these challenges. One of the examples of solving this program is the creation of an "umbrella" employee development program in the direction of Big Data [10]:

- annual overview training for top management, where executives learn about the capabilities of Data Science and successful cases from around the world;
- basic training for middle management aimed at showing the benefits of using Data Science and Machine Learning tools for different types of tasks;
- hackathon for mixed teams of managers and data scientists where they learn to interact with each other;
- regular online competitions for data maintainers on external data sets;
- annual advanced case machine training for data scientists;
- individual training for managers and data-scientists.

3. Methodology

A project in Data Science is similar to software development: in both cases the goal is to gain economic benefits or create additional value for clients. To achieve it, the code is written. That is why managers often expect the same straightforward results from Data Science processes as in software development, and success is measured exactly by whether they have achieved the desired goal [2].

One of the most common approaches to data analysis is cross-industry process for data mining (CRISP-DM), where each task is presented as a certain hypothesis that can be accepted or rejected. Furthermore, work on the hypothesis goes through typical stages: evaluation of the required data set, determination of success criteria and quality metrics, data preparation, modeling, and result evaluation. Most often the work turns out to be cyclical; some stages may be repeated several times [11]. If it turns out that the created model improves the current process or product of the company, it is put into operation. But it also happens that the efforts spent do not give a positive result: for example, due to limitations on the model's application or poor data quality. Such situations should not be considered as a negative outcome, because the project team has gained experience and knowledge that it will apply to further work on other hypotheses [12].

CRISP-DM is a standard process for examining data. The lifecycle model for data analysis consists of six phases. The sequence of them is not strictly defined. As a rule, in most projects you have to go back to the previous phases and then move forward again. Data research life cycle phases are the following (Figure 1.) [13].

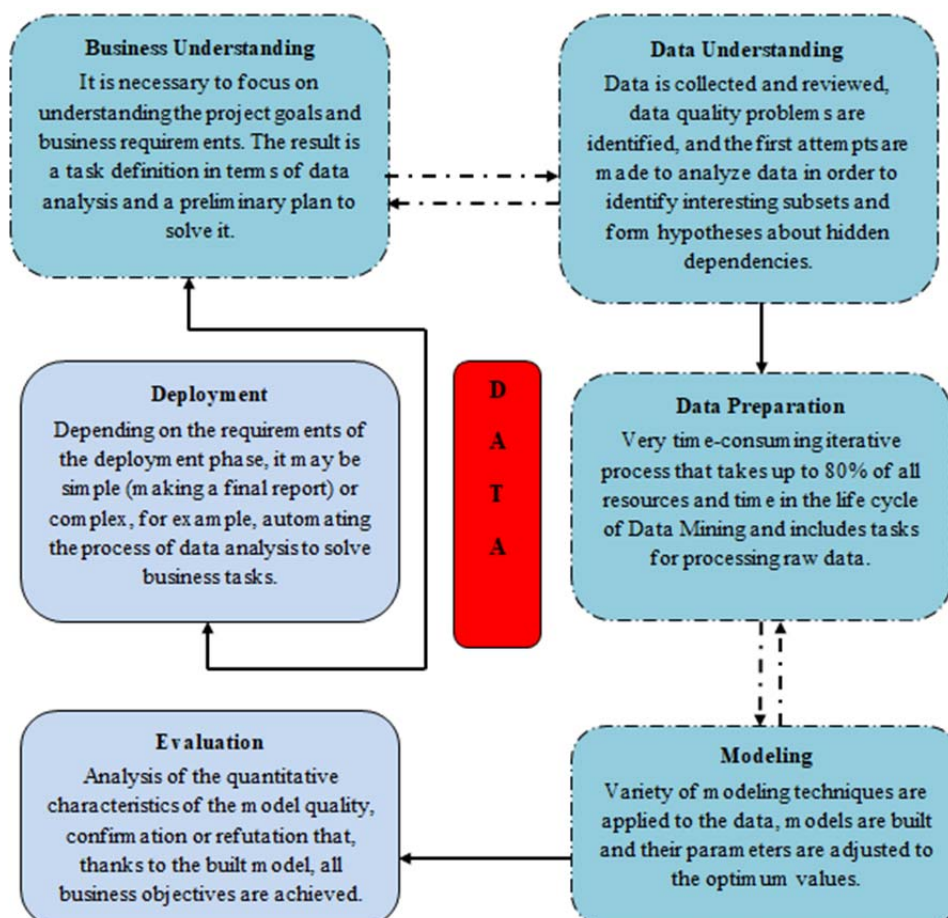


Figure 1. CRISP-DM data cycle phases in the study process

1. Business Understanding. Phase objectives:
 - to determine business goals;
 - to evaluate situation;
 - to determine objectives of data analysis;
 - to make a project plan.
2. Data Understanding. Phase objectives:
 - to collect initial data;
 - to describe data;
 - to research data;
 - to check data quality.
3. Data Preparation. Phase objectives can be performed many times without any predetermined order:
 - to select data (tables, records, and attributes);
 - to clear data, including converting it and preparing it for modeling;
 - to derivate data;
 - to merge data;
 - to customize data.
4. Modeling. Phase objectives:
 - to select method of modeling;
 - to make a test for the model;
 - to construct the model;
 - to evaluate the model.
5. Evaluation. Phase objectives:
 - to evaluate results;
 - to review process;
 - to determine the next steps.
6. Deployment. Phase objectives:
 - to plan the deployment;
 - to plan the support and monitoring of the deployed solution;
 - to make the final report;
 - to make the review of the process.

Ready to enter the game, the outcome of which is not known in advance, is a new requirement, not typical for the classic approach, where a project that has not achieved the business goal is a failure in advance. The advantages are obvious: the culture of tolerance for negative results liberates employees and creates a productive atmosphere, which in the long run still leads to positive results [14]. In the case of banking, a successful model can be exemplified by predicting indirect indicators and historical data on client income. On the basis of the obtained data, the bank can offer an individual loan product without official confirmation of income.

4. Discussion

A lot has been said about the digital economy after the Address of the President of the Russian Federation to the Federal Assembly on December 12, 2016, and already by April 25, 2017 the Ministry of Communications prepared a version of the program "Digital Economy of the Russian Federation" and sent it to those responsible for making proposals for

changes to the submitted draft. In a growing storm of events dedicated to such an important topic, one can notice some inaccuracies in the understanding of the term "digital economy" itself, or even misunderstandings of the main accents that need to be properly placed. In the heat of the discussion and exchange of competent opinions, the three basic components of economic digitalization are sometimes forgotten, and all three things that should not be overlooked can be found in the very notion of the digital economy [15].

Firstly, data is what stands behind the first word in the term "digital economy". Whether we admit it or not, the new round of economic systems is based on even more massive and qualitative data collection and analysis. At present, this is becoming objectively possible and feasible, and thus will gradually increase in line with the needs of the economic actors to support their own solutions everywhere and in everything. Business and its consumers, the public and state regulators surround themselves with digital data sets not for idle curiosity, but because it is convenient and professional. While we do not rule out temporary misunderstandings about the role of data, and even taking many mistakes in handling it, we cannot deny the importance of gradually moving towards total and widespread use of data-driven decision-making. The performance of any qualified professional can now be improved many times over with structured digital information technologies [16].

The data allows to digitize the business reality around us and build high-level models for in-depth retrospective business analysis or for predictive "fast" analytics where time to make decisions is limited to fractions of a second. Data opens up new facets of events and phenomena, helping practitioners to confirm or disprove the hypotheses of theorists. In turn, it is the digital data that forces economies to bear significant costs and to build the infrastructure that is vital to their full use. Data centers, high-speed communication networks, sensors and distributed computing power are the technical side of the overall digital data globalization process. Subject applications, complex algorithms, trainable neural networks, cryptographic protection, preservation of integrity of data units, productive cloud servers, interactive infographics and indicator panels represent another broad front of the development of the processes of collecting and processing of "figures" [17]. In the endless range of technologies of the work with data, the possibilities of moving them both among IT-systems and among different subjects have gained special importance. The ways of data transmission via protected and public schemes are actively growing and transforming. Of course, open data from states and communities, as well as private business data offered

for sharing are playing an increasingly important role [2].

It is the data that forces businesses to rethink their behavior and to be equipped with effective tools to tame them and use them for their own pragmatic purposes. As a result, a high number of innovative projects and work teams that methodically and gradually create data are processing it and finding new ways to extract additional value from seemingly stretched and incapable arrays. Thanks to their information nature, these digital streams can be grinded and combined endlessly. Perhaps a business has not yet had such a period when its management received at its disposal a lot of resources. Subjects became accustomed to save available resources and it was difficult to find new ones. Now, for the first time, they have faced not a deficit, but an excess. By analogy with the terminology adopted for natural resources, one can speak about infinitely large explored reserves of digital data. This state of affairs will undoubtedly lead to a significant transformation of the economy. There is a need for fundamentally different instruments and approaches. However, in order to properly extract data and work with them in a qualitative way, it is necessary to understand what this is really for [3].

Second, needs are "economy" in this context. Active informatization transforms consumer behavior. They, armed with different levels of information quality, become sophisticated and demanding. For marketers, there are difficult times when, in order to really get to the mind and heart of their clients, they have to delve into various details and intricacies of social communication. Personalization of products and services while maintaining a confident scale of production: is it possible now and in the future [6]. Deepening in the needs and preferences of each individual consumer causes genuine interest in a particular way of producing goods. The first successful experiments of a personalized "smart" plant operates only on robots and following the orders. Marketing is gradually and steadily approaching the essence of economic interaction, the main driving force forcing each person to engage in economic interaction, to the needs, and this fact transforms marketing [18]. Gradually, from sales and annoying advertising it goes the way to joint consumption and algorithmized technologies. Let there be errors and obvious failures on this way. But the skill is honed, the consumer is studied, the hypotheses are tested, and the driving force of needs brings to life automated production, lightning-fast and multifunctional robots, new materials and products, unthinkable combinations of convenient and accessible services. The power of needs transforms the way we consume, freeing us from the cargo of property and offering in return the

convenience and comfort of sharing. And what's more, digitalization brings with it extremely interesting possibilities of combined satisfaction of needs that initially seem to contradict each other. This is also made possible by the interaction of intelligent physical and logical things. Man uses a machine and algorithmic network capable of meeting his needs quickly and holistically [1].

Consumption can do more. Based on the mass demand and the resources released, special forms of investment and effective projects to develop new products "by subscription" are emerging in the markets. In turn, they really and strongly change the very understanding and meaning of value, price and asset value. Free forms of real and virtual resources in electronic form penetrate into life, as exemplified by cryptocurrencies. And it is at the point of intelligent consumption that the digital economy has an enormous opportunity to engage and connect other global spheres such as society, cultural values and traditions, history and religion, art and the environment, etc. Only now we seem to have at our disposal the potential to somehow ennoble primitive consumption into a manageable zone of total comfort for everyone, without contradicting humanitarian and traditional values [16].

The economy is transforming, and its technologies, which tend to ways of selling or exaggerating the consumer effect of the sold goods, will lose efficiency. Products and services turn into open ones. Digital trust allows you to control the origin of a product or service, as well as verify their authenticity, authorship, quality and other important parameters. Mobilization of all kinds of control devices operating within a single global information network will invariably lead to significant changes in the approach to the delivery and sale of a variety of products and services. There are many alternative solutions, some of which are provided almost or conditionally free. And the very speed of distribution of information about a product and service, about the manufacturer and seller level out any tricks to hide important consumer characteristics. But innovative products and services, especially those linked in a whole network of interacting physical and logical things, for the consumer are really becoming complex and fundamentally different in the way they are used from traditional goods and services. Those who offer fundamentally new "digital" ways to meet new and old needs, just have to create and master new principles of communication with clients and work in the markets. It is necessary to visualize needs and products, show successful consumer experience, explain and educate users, provide constant and convenient support. But this is a slightly different model and technologies, transactions and businesses that are almost impossible without digital platforms

and an expanded understanding of consumption principles [10].

Thirdly, management is a qualified system of interaction cooperation, which is designed in theory and practice to predict, plan, organize, execute, monitor and coordinate all large-scale activities to actively use data to meet the growing needs of humanity. And this system today, perhaps, more than ever, is in great need of comprehensive support: scientific, methodological, technological, informational, instrumental, creative, etc. Intensive development of management towards informatization of all its controlled processes and objects provokes the emergence and expansion of new knowledge and new forms of rapid learning. Moreover, distance e-learning centers and game practices of learning new skills are becoming particularly important. Complication of knowledge, increase in their specialization, constant enrichment with practice and even contradiction of different approaches and scientific schools, seasoned with high speed of experience from successful and failed startup projects, now makes professionals to be always in the attention and vigorously follow the events in the target subject area. Making a complete qualified picture in each specific professional field becomes extremely important and difficult. And the professional spheres themselves began to actively mix, change, share and shrink. Therefore, it is necessary to actively learn and participate in expert, advisory and discussion activities. The need for expert support tools and competence updating will grow steadily and confidently [13], [19].

Today specialists and experts are trying to plan their work on digital transformation of economy. But in our current situation it is not easy. It is extremely difficult to model the digital future at a time when even the majority of technologies declared as promising are not yet ready for full use. The probability of implementing most of the expected digital events is very low, and trying to guess what synergy they will give to the economy is quite difficult. And, nevertheless, gradually and progressively, from individual projects and businesses to sectoral and interstate solutions, making mistakes and making original spectacular discoveries, we have to move forward, digitizing globally [16], [20].

5. Results

It is quite understandable that large and medium companies from retail and service sector show increased interest in using Big Data technologies. These technologies are actively used by banks and mobile operators. In addition, they are used by large manufacturing companies to analyze data on

equipment breakdowns and reduce downtime, which helps reduce costs. For example, in the field of flight management, analysis of data sets can increase equipment reliability and reduce the number of failures. But the field of application of Big Data is much wider. For example, another area of application of Big Data technology may be buying a business. It is a question of the widest possible coverage of various information about a company. It is also the analysis of ordered data about [8]:

- physical property features;
- operational data;
- financial data;
- material resources;
- legal data.

The sources of information may be: file tables, traditional Database Management System, and accounting systems. It is also necessary to analyze disordered data: customer feedback, inspection results; incidents, service requests; competitive environment; and IT infrastructure. Here the sources of information can be: tables, schemes, social networks, expert assessments, etc. The result of the analysis of such data can be the so-called object passport, which includes data on location, area, number of floors, permits, inventory data, competitive environment, historical financial data, seasonal factor of sales, etc. But, as it is known, the value of the company is not equal to the simple sum of its tangible assets. Collection and analysis of information on Big Data technology will make possible to estimate the value of intangible assets. They can include [16]:

- manpower, knowledge and skills;
- information resources, databases; organizational and management structure;
- talents, potential employees;
- brand name and reputation;
- developments;
- customer bases;
- relationship with counterparties.

Overseas, it has already become the norm for medium and large enterprises to have a CDO. Most often, he reports directly to the top management. For our companies, even large ones, this position is still a rarity. There is a shortage of specialists in Big Data management and related analytics. In principle, universities in Russia practically do not train such specialists. Young specialists from the IT sphere have to be retrained, moreover, they often have very high requirements on starting salaries considering insufficient qualification. Quite often in practice there is a very low level of knowledge at the majority of heads of the enterprises in the field of management of the Big Data, accordingly, they cannot formulate problems for experts or at all do not

see necessity for such experts. These problems are especially relevant in the light of the State Program "Digital Economy" approved by the Government of the Russian Federation in 2017, since the transformation of the economic system into a digital one is primarily related to the effective management of large data [21].

6. Conclusions

The management of structured and unstructured data with the help of new technologies and tools for their processing is a topic currently relevant both at the level of specific enterprises and at the state level. Thanks to the analysis of the Big Data (i.e. methods of processing large amounts of data for specific tasks and purposes), the possibilities of analytics are greatly expanded and it is possible to obtain valuable information for customers. Using the capabilities of the Big Data, companies can get a lot of useful information about competitors, partners and customers. In the era of multinational companies and international cooperation, the Big Data technologies are no longer a national treasure. The role of the Big Data is to be a liquid product, a prerequisite for improving the profitability of organizations through personalized customer service and predictive analytics. For the Russian digital economy, it is crucial not only to legitimize a single definition of Big Data, but also to achieve the emergence of data exchanges. This will be a fundamental factor in the Russian economy's competitiveness in the world market, as well as a great step in supporting national business within the country.

References

- [1].Gobareva, Y. L., Gorodetskaya, O. Yu., & Kochanova, E. R. (2015). Features of Big Data technology to improve the quality of operation of CRM-systems. *Transport business in Russia*, 5, 62-63.
- [2].Baburin, V. A., & Janenko, M. E. (2014). Big Data technologies in service: new markets, opportunities and challenges. *Technical and technological problems of service*,1(27), 100-105.
- [3].Maltseva, S. V., & Lazarev, V. V. (2015). Big Data E-Marketing Marketing Analytics. *Information Technologies in Design and Production*, 21, 62–67.
- [4].Suleikin, A. S. (2015). Possibilities of using Big Data technology in large retail. *European Research*, 10(11), 77-79.
- [5].Munoz, M. (2013, November). Space systems modeling using the Architecture Analysis & Design Language (AADL). In *2013 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)* (pp. 97-98). IEEE.
- [6].Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F. B., & Babu, S. (2011). Starfish: A Self-tuning System for Big Data Analytics. *CIDR. Fifth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, 11*, 261–272.
- [7]. Michalik, P., Štofa, J., & Zolotova, I. (2014, January). Concept definition for Big Data architecture in the education system. In *2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMII)* (pp. 331-334). IEEE.
- [8].Bulgakov, A. (2017). Big Data в финансах (Big Data in Finance). *Journal of Corporate Finance Research*, 11(1), 7-15.
- [9].Saltz, J. S., & Grady, N. W. (2017, December). The ambiguity of data science team roles and the need for a data science workforce framework. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 2355-2361). IEEE.
- [10]. Avdeeva, I. L., Polyanin, A. V., & Golovina, T. A. (2019). Digitalization of industrial economic systems: problems and consequences of modern technologies. *Economy. Control. Law*, 19(3), 238-245.
- [11]. Hu, V. V., & Kaabouch, N. (2014). *Big Data Management, Technologies, and Applications*. Hershey: IGI Global.
- [12]. Ivashchenko, A. V., & Dvoynina, O. V. (2014). Analysis of super-large data arrays (Big Data) in a single information space of a research and production enterprise. *Science and World*, 12(16), 44-47.
- [13]. Ia, T. V. (2014). Natural and artificial information field. *Mezhdunarodnyi zhurnal prikladnykh i fundamental'nykh issledovaniy-International journal of applied and fundamental research*, 5(2), 178-180.
- [14]. Bayliss, D. (2016). Models for Big Data. In *Big Data Technologies and Applications* (pp. 237-255). Springer, Cham.
- [15]. Golov, N. I., & Kravchenko, T. K. (2014). Designing a data warehouse to solve Big Data problems. *Information Technology in Design and Production*, 1(153), 56-61.
- [16]. Jacobs, A. (2009). The pathologies of big data. *Queue*, 7(6), 10-19.
- [17]. Tsvetkov, V. Y. (2014). Worldview Model as the Result of Education. *World Applied Sciences Journal*, 31(2), 211-215.
- [18]. Chernyak, L. (2011). Big data: a new theory and practice. Open systems. SUBD, 10.
- [19]. Denisov, L. V., Boitsov, A. G., & Siluyanov, M. V. (2018). Surface Hardening in Hydraulic Cylinders for Airplane Engines. *Russian Engineering Research*, 38(12), 1080-1083.
- [20]. Sorokin, A. E., Bulychev, S. N., Novikov, S. V., & Gorbachev, S. I. (2019). Information Science in Occupational Safety Management. *Russian Engineering Research*, 39(4), 324-329.
- [21]. Chekharin, E. E. (2016). Big data: big problems. *Prospects for Science and Education*, 3(21), 7-11.