



EXPERIENCE IN USING BIG DATA TECHNOLOGY FOR DIGITALIZATION OF INFORMATION

Andrey Aleksandrovich
Rodionov

Doctoral Student at the Institute for
Staff Advanced Training and
Statistical
Researchandre-rodionov@mail.ru

Rabim Alikulovich Fayziev
Professor of the Department of
Mathematical Methods in Economics
Tashkent State University of
Economicstdu@yandex.ru

Saidakhror Saidakhmedovich
Gulyamov

Academician, Doctor of Economics,
Professor, Head of the Department
"ICT and Digital Economy" of the
Institute of Advanced Training and
Statistical Research at the State
Statistics Committee of the Republic
of
Uzbekistanacademicgulyamovss@gmail.com

ABSTRACT

This article summarizes the experience of using big data in data digitization for statistical analyzes and forecasts. The authors of the article also prepared analytical information on ongoing research in various sectors of the economy using Big Data technology. A technique for processing unstructured data and systematizing information is given, followed by the construction of predictive models that can be used in the upcoming census of the population of Uzbekistan.

CCS CONCEPTS

• **Big Data technology**; • **Statistical research**; • **Forecast**; • **Analysis**; • **Big data**; • **Statistics**; • **Structured and unstructured data**;

ACM Reference Format:

Andrey Aleksandrovich Rodionov, Rabim Alikulovich Fayziev, and Saidakhror Saidakhmedovich Gulyamov. 2022. EXPERIENCE IN USING BIG DATA TECHNOLOGY FOR DIGITALIZATION OF INFORMATION. In *The 6th International Conference on Future Networks & Distributed Systems (ICFNDS '22)*, December 15, 2022, Tashkent, TAS, Uzbekistan. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3584202.3584266>

1 INTRODUCTION

The concept of Big Data technology includes huge arrays of structured and unstructured data. A prerequisite for the data to fit this definition is that it must be truly large. With the help of a system of automated tools, it is possible to process such data and information. Such data can be used to conduct statistical research, make forecasts, as well as to make correct and rational decisions. Today, there are discussions and disputes about a strict and clear definition of Big Data technology.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICFNDS '22, December 15, 2022, Tashkent, TAS, Uzbekistan

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9905-0/22/12...\$15.00
<https://doi.org/10.1145/3584202.3584266>

Despite the fact that the term "Big Data" itself was proposed back in 2008 by Clifford Lynch, editor of the journal Nature, and included in this concept that "more than 150 gigabytes of data per day must be processed here," many scientists still do not come to a consensus regarding this definition [1]. After all, the criteria for determining that have already been proposed may not be enough for a technology that is a promising direction for the economic development of corporations and companies.

In order to work with such technology, it is often necessary to attract specialists from a variety of areas. Indeed, in order to obtain structured information from such a huge amount of data, which can and should be used for analysis in the future, a lot of effort is required [2].

That is, a specialist in the field of Data Science should be involved. Data Science is the application of scientific methods when working with data to find the right solution. In a broad sense, the natural sciences are based on Data Science. For example, a biologist conducts experiments and analyzes the results to test his hypotheses. He must be able to generalize private observations, exclude accidents and draw correct conclusions [3].

Such an employee works with data in the same way as a scientist in any other field. It uses mathematical statistics, logical principles and modern visualization tools to get the result. An employee in the field of Data Science processes data arrays, finds new connections and patterns in them using machine learning algorithms, and builds models [4]. A model is an algorithm that can be used to solve business problems. Global search engines, recommendation services, voice assistants, autonomous trains and cars, facial recognition services - all created with the participation of data scientists.

Data analysis is part of the job of a data scientist [5]. But the result of his work is a model, a code written on the basis of analysis. This is the main difference between a data scientist and a data analyst. The first is an engineer who solves a business problem as a technical one. The second is a business analyst, more immersed in the business component of the task. He studies needs, analyzes data, tests hypotheses, and visualizes the result.

Data Science works for both startups and large corporations. In the first, specialists work alone or in small teams on individual tasks, and in the second, they implement long-term projects in conjunction with business analysts, data analysts, developers, infrastructure administrators, designers, and managers. The project

manager with analysts takes on most of the work: communicates with the business, collects requirements, and forms the terms of reference. Depending on the level and principles of work in the company, the Data Scientist participates in negotiations or receives tasks from the project manager and analysts. The next step is data collection. If the company does not have processes for obtaining data, the data scientist solves this problem as well. It introduces tools that help to automatically receive and pre-cleanse, structure the necessary information. Data markup is also a way to put things in order. Each entry is assigned a label that can be used to determine the class of data: is it spam or not, the client is solvent or not. Algorithms are rarely used for this task; labels are affixed manually. Well-labeled data is valuable.

Until 2011, Big Data technology with subsequent data analysis was used only for statistical needs and purposes. But, since 2012, there has been a sharp increase in the amount of data that required digitalization, and therefore, the need for this technology has increased dramatically [6]. It took another couple of years for huge companies to pay close attention to this technology. Since 2014, such major brands as Oracle, IBM, Microsoft, Google, Facebook, Apple and many others have become interested in technology. They needed this technology in order to rationally deal with the digitalization of data and their systematization. In addition to these corporations, many higher education institutions and government agencies use Big Data technology [7].

2 RESEARCH METODOLOGY

The selected research methods include the method of observation, which is presented in this paper. More and more companies are starting to think about how to implement Big Data technology in practice. An interesting point is the fact that giant companies such as Microsoft, Apple, Facebook and Google can obtain data about their users and their activity using applications that are installed on their mobile phones and other gadgets. This became possible due to the fact that a huge number of people use the services of such companies. For example, G-mail is used by more than 1.6 billion people, the Android platform is installed by 2.7 billion users, and Youtube is used by about 2 billion people. Among the interests of companies is the activity of Internet users. For example, corporations are interested in what users mark and "like" in social networks. They collect statistical data and send statistics about user activity to various services - for example, taxi services and other companies. Thanks to Big Data technology, this has become possible. It allows you to organize and send a huge amount of necessary information. It is important to note that before companies did not have such opportunities, since the unstructured information stored in databases was nothing more than a set of numbers that hardly anyone would be interested in. However, now, thanks to specialists who can work with such volumes of data, it is possible to obtain structured and semi-structured data that are suitable for various purposes and needs [8].

And this technology is unique and useful not only for companies, but also for the users themselves. If at first glance it seems that the use of information about user actions is relevant and useful only for corporations, it is worth considering that thanks to such activities from these companies, the use of services becomes absolutely

free. Also, thanks to the digitalization of large amounts of data, companies have the opportunity to connect special advertising that people will click on. Before appearing on the smartphone screen, the system will analyze what exactly a particular person may like [9].

And it brings huge benefits and profits to companies. For example, Facebook was able to earn approximately \$18.4 billion in 2019 alone. According to company representatives, "each active user of the network was able to bring the organization about \$ 8,000 a year." Only thanks to the high-quality digitalization of data implemented using Big Data technology, such indicators have become available.

Big Data technology does not stand still, and its parts are improving every year [10]. This leads to the fact that the systematization and structuring of "raw" information becomes more and more simple and leads to better results. More and more areas of application are in Big Data technology, as the number of users of programs and the Internet is only growing day by day. In the near future, it is expected that even not so large companies can start using this technology in order to engage in high-quality and useful digitalization of data [11].

Also, Big Data technology helps to develop other industries. For example, it can help develop the education system in the future tense. An example is the state of Virginia, which conducts research on the activities of students who lag behind their peers in terms of academic performance. According to statistics, about 400 thousand people are expelled from US universities every year. The authorities were concerned about such figures. Such performance affects the fact that many higher education institutions are reduced the percentage of budgeting, and they are not given loans for various needs. Thanks to the advanced technology of Big Data, it is possible to identify those students who are lagging behind in the curriculum in time and recommend the right tutor for them in order to catch up with the rest of the students.

With the help of the analysis that is carried out using this technology, high-quality and useful teaching aids are selected that help students improve academic performance. In fact, this technology is of great help to such students, as an individual selection of the necessary books and literature is carried out in order to help them catch up with other students in their studies [12].

However, the role of Big Data technology in education does not end there. It is worth noting that this technology helps in such an industry as career profiling. That is, this means that this technology allows students to identify their talents, and subsequently choose the direction that suits them best. Unlike conventional testing, this technology offers much more opportunities in this regard. There is a collection of really important and useful information for a person, which is subsequently processed [13].

3 RESULTS

To give specific examples, it is worth noting that in the United States of America a program has already been invented for determining a student's career profile, which is called "SC-Accelerate". This software product works on the Career Choice-GPS technology, which allows you to analyze the temperament and character of students, identifying their interests and inclinations in various industries and

subject areas. After that, these students receive recommendations on where they can go.

To understand how Big Data technology works, it is worth dwelling on its characteristics:

- 1.Amount of information. According to the definition of this technology, it is necessary that at least 150 gigabytes per day be processed [14];

- 2.The speed of systematization of huge amounts of information and their processing. Displaying and changing data in real time requires truly advanced technological solutions;

- 3.Variety of data types. Data can be in structured, semi-structured and unstructured form. In addition to those data that are represented as a number, it is possible to store graphic and multimedia data;

- 4.Reliability of data arrays. The analysis carried out must be reliable;

- 5.Variability. When selecting information and data, it is necessary to select information that has low variability. For example, there are types of data that, depending on the season, have a certain trend. They are suitable for qualitative analysis [15].

- 6.Value. Depending on what the data source is and their type, we should talk about the different value of the information provided.

4 DISCUSSION

To make correct and rational decisions, it is necessary to analyze all significant factors. That is, if we talk about those data that are relevant for various analyzes, all of them should be included, since the direct and indirect relationship obtained when conducting these analyzes will reveal a trend. With the help of Big Data technology, you can create models for testing certain solutions, products and ideas. The main sources of big data include the following resources:

- 1.The Internet of Things and the devices connected to it. With the help of the Internet of things, you can receive information and data coming from various devices, gadgets and household appliances. This helps to obtain rather voluminous arrays of information that report on any information and indicators [16];

- 2.Mass media and social networks. With the help of social networks and the media, you can also get a lot of information about user actions and other events;

- 3.Data provided by various companies. For example, this may include transfers, transactions, customer data and other information that may be useful in conducting any analysis;

- 4.Information obtained from various instruments and devices. This may be information provided by weather stations. This also includes data from satellite devices [17].

- 5.Statistics of countries and cities. This item can include a lot of information and information regarding various indicators. For example, you can track the trend of indicators such as births, displacements, deaths, and more;

- 6.Data related to medical indicators. This may include images, tests and other indicators. They reflect the patient's medical history, and may also constitute a trend for analysis.

Thanks to modern computers and technologies, you can get instant access to large amounts of information. In order to store data related to Big Data technology, specially equipped data centers are used, which are equipped with extremely powerful server systems [18].

In addition, there are physical servers, and "data and information lakes" - Data Lake, which store a huge amount of information presented in an unstructured form, are also used and applied. Such information may not always be suitable for qualitative analysis, as relationships may not yet be established. To do this, this type of data should be transferred to a semi-structured and structured form.

There is a special framework - "Hadoop", which consists of a large number of programs needed to create and perform distributed computing tasks. To work with Big Data technology, it is necessary to apply new methods of implementation and management, as well as preparing information for analysis [19].

Big Data technology is an extremely interesting and useful statistical tool that finds new areas of application and implementation. For example, in 2021, a population census was held in Russia using Big Data technology. In addition to the fact that with its help it was possible to estimate what the population is, migration flows were also estimated and determined. Unlike traditional methods of conducting a population census, Big Data technology has brought the fact that the state will always have up-to-date and useful information [20].

As the head of Rosstat Pavel Malkov noted: "Artificial intelligence technologies will be applied and used in the analysis of the information that is sorted and transformed by mobile operators. This is the first experience of such experiments at the state level in terms of the use of big data." Also, a contract was signed with the Center for Economics and Infrastructure, which is working in this direction with operators. Such a contract will cost more than 250 million rubles. Also, contracts were signed with companies such as MTS, Tele2, and many others.

The first stage of the population census using Big Data technology was held on October 1, 2020 in various regions and regions. Throughout the country, the census took place from April 1 to April 30, 2021.

Speaking about the analysis of big data, it is worth saying that it happens thanks to 4 main methods:

- 1.Using descriptive analytics [21]. It is an extremely common method for analyzing big data. This includes historical data as well. The main goal is to find out the causes and patterns of success or failure in various fields. This helps to apply the data to the most effective kinds of models. Descriptive analytics uses basic mathematical functions. For example, descriptive analytics is used in sociological research and web statistics data obtained by Google-Analytics services.

- 2.Using predictive analytics. It helps to predict the most possible developments of events based on the data that are available. To do this, ready-made type templates are used based on objects and phenomena with a similar set of properties. Using this type of predictive analytics, you can track changes in the value of securities in the stock markets. Also, in many cases, the ability of potential borrowers to repay the loan is assessed.

- 3.Prescriptive analytics. This analytics identifies problem areas in business and any other activity. For example, some healthcare centers are using prescriptive analytics to reduce readmissions by about 10-12 percent.

4. Diagnostic analytics. With the help of diagnostic analytics, the causes of what is happening are analyzed. This helps to detect anomalies and unrelated data.

5 CONCLUSION

In conclusion, I would like to note that Big Data technology allows, with the help of data digitalization, to conduct high-precision statistical research. Big data is already becoming part of everyday life everywhere, and is used by many countries in such ongoing activities as population censuses, sociological surveys, and so on. With the help of this technology, a huge amount of data is processed in order to improve the quality of statistical research.

However, it is important to note that such highly accurate ongoing statistical studies can only be performed when structured data are selected. Sometimes the process of structuring data can be very complex, and therefore, special workers are involved who can do such processing of the necessary information.

And, although the process of structuring data is extremely difficult, it is worth saying that Big Data technology fully justifies itself, because its use allows you to monitor in real time the data and information that you need at a particular point in time.

It is important to emphasize the role of artificial intelligence in all this. Although there is still no exact concept and definition of this technology, it is worth saying that using Big Data technology in cases where artificial intelligence is applicable is extremely rational and useful. After all, the more data provided to create any model necessary for conducting a qualitative analysis, the more accurate the result, research or forecast will be.

The role of Big Data technology in forecasting when using artificial intelligence is highlighted. Taking an almost ready-made model and adding a variety of data to it using Big Data technology, you can achieve an accuracy close to 100 percent.

For example, if the task requires tracking seasonal products that have a good trend in terms of sales or another parameter, Big Data

technology will only help increase the accuracy for building a model using artificial intelligence.

However, both artificial intelligence and Big Data technology require knowledge in this area. And, therefore, in order to create predictive and other statistical models, it is necessary to attract such specialists who can make the data more structured.

Also, these specialists understand exactly what data is needed to conduct high-precision statistical analysis. More and more countries are moving to the use of artificial intelligence and Big Data technology, as they are the future. These technologies have a positive impact on the economy, and also allow for accurate analyzes related to statistical, economic and other studies.

REFERENCES

- [1] Anthony Aline. Machine Learning. pp. 124-125 (2019).
- [2] Thomas Hoeren, Barbara Kolany-Raiser. Big Data in Context. p.25 (2017).
- [3] Nasir Raheem. Big Data. pp.255-256 (2019).
- [4] Jim Lindel. Analytics and Big Data. p.157 (2018).
- [5] Khaitmatov U.T., Rodionov A.A. Digitalization of data using Big Data technology (2022).
- [6] Gulyamov S, Rodionov A, Fayziev R, Mukhitdinova M. The introduction of artificial intelligence in the study of economic disciplines in higher educational institutions (2022).
- [7] Mr. Ashok Muthuswamy. Predictive Analytics Using Regression. pp. 344-345 (2020).
- [8] Feti Calisir. Industrial Engineering in the Big Data Era. p.150 (2019).
- [9] Alex Lu. Big Data and AI. p.340 (2020).
- [10] Ian Foster. Big Data. pp.101-103 (2016).
- [11] Russell Walker. From Big Data to Big Profits.p.233 (2015).
- [12] David Pope. Big Data Analytics. pp.105 (2017).
- [13] Hans Weber. Big Data and Artificial Intelligence. pp.200-201 (2020).
- [14] Nam Toai. Future Data and Security Engineering. p.177 (2017).
- [15] M.Bless. Big Data Processing. p.72 (2018).
- [16] Mike Touil. Big Data. p.54 (2019).
- [17] Ch Lok. AI and Big Data. p.15 (2019).
- [18] Sergio Consoli. Data Science for Economics and Finance. p.200 (2021).
- [19] Ulrik Jagar. Data Science Strategy. p.163 (2019).
- [20] Steven Cooper. Data Science. p.100 (2018).
- [21] Neil Wikins. Internet of Things. pp.222-223 (2019).
- [22] Decree of the President of the Republic of Uzbekistan, dated February 17, 2021 No. DP-4996 «On measures to create conditions for accelerated implementation of artificial intelligence technologies».