

Research of Net Users by Means of Big Data Technology

Irina Vdovychenko^{1*}, Svitlana Khotiskina¹, Mykola Khodukin², and Ilyos Mamanazarov³

¹National University, Kryvyi Rih, Ukraine

²State University of Economics and Technology, Kryvyi Rih, Ukraine

³Ministry of Economic Development and Poverty Reduction of the Republic of Uzbekistan, 100174, Uzbekistan

Abstract. The presented article offers the review of means, methods and possibilities of one of the most welcomed modern information technologies – Big Data. The popularity of the technology is grounded in the work. The sources of the big data, the speed of their appearance, amounts of information in the world are examined. The main specializations of the employees for the operation in the field of the structuralized and non-structuralized data of big amounts and the significant variety are marked. The scientific and practical directions and the structure of Big Data are analyzed. The examples of the Big Data use's technology in the world are examined. The research results of the big data, received from the Ukrainian sites, are presented in the article. The possibilities of the programming language R, as one of the leading instrumental means of the Big Data technology, were used in the work. The research aim was to define the attendance frequency's dependence of different sites by the young people, aged from 17 to 35 years old, upon their sex, provision of job, incomes and age. The authors present the obtained results and offer directions of the activity, where they may be used.

1 Introduction

The amount of information grows very quickly today. The scope of information, used by the modern individual, are not only books, magazines, films and web-pages. Besides it, the vast number of data, appearing in the operation process of enterprises, banks, institutions and so on, is hidden from the user. The following ones refer to them:

- ✓ data of all-possible accounting;
- ✓ information of production output and sales;
- ✓ histories of transactions in banks;
- ✓ data of the sick persons and medicaments at the hospitals;
- ✓ data of the location and routes of all the cars in the auto-parks;
- ✓ vast quantity of research results;
- ✓ data of observation of planets and stars;
- ✓ data from the cameras, fixed at the enterprises, plants, in the roads and streets of the cities to observe the traffic of people and cars;
- ✓ data are contained in telephones, TV sets;
- ✓ and so on and so forth.

“The humankind produces the gigantic quantity of data through the nets, the public transport and the Internet-purchases without any stop. Their volumes impress the mind. We load 95 million of images and video, 340 million tweets and 1 milliard of documents every day. We produce 2,5 quintillions of bites per a day in general” [5].

The analytical company IDC presented the report in 2012, where it was predicted, that the information volumes would be doubled every 2 years during the

next 8 years. The data amount in the world will reach 40 ZB (1 ZB = 10^{21} byte) for the nearest 7 years and it means that every resident of the Earth will have 5200 GB of data at the disposal.

The analogical data are presented by the company IBS, which tells that the amount of the digitalized information grows along the exponential curve and, according to the forecasts, the mankind will form 40-44 Zettabytes of information by 2020. The data represent the vast amount of information of different type: pictures, video, multimedia, text, geo-data, web-magazines, machine code. All the information is located in various storehouses and is hardly subjected to the analysis with the help of the traditional methods. The specialized technologies are used for that.

The traditional methods of information analysis cannot overtake the vast volumes of the constantly growing and restored data, - the fact, which finally opens the road for the Big Data technologies. The term Big Data has the relation to the selection of data, which size exceeds the possibilities of the standard data bases in processing, storage, control and analysis of the information. It's the instrument and the method of data analysis.

2 Background

The very term Big Data began to be used several years ago. The term “big data” was introduced by Clifford Lynch, the editor of the magazine “Nature”, in 2008. The problem of the rapid growth of the world volumes of information was examined in the magazine. It's clear,

* Corresponding author: vivin2015@knu.edu.ua

that big data existed before as well. According to the words of the professionals, the most part of the data flows of more than 100 GB per a day belongs to the Big Data category.

The first investigations on the data separation from data were conducted already long ago. The mathematical theory, concerning the reveal of the definite data on the indistinct field of data was elaborated still in the 70-80-s. The practical use of the theory was limited by the possibilities of the computer engineering.

The Big Data notion today has already entered the everyday life reliably throughout the world. The most part of these data is collected through the Internet.

Big Data — are the totality of approaches, methods, tools that are used for the operation with the structuralized and non-structuralized data of the significant amounts and great variety for the solvation of the definite tasks and objectives. They are welcomed on conditions of the continuous growth of information. “Big data serve as an alternative to the traditional systems of the data base control and to the decisions in the frames of Business Intelligence. These methods, that allow fulfill the distributed processing of information, may be applied both to the vast selections of data (for example, the contents of all the pages in the Internet) and to the small ones. Big data have an important significance, because the big quantity of data leads to the more exact analysis, which, in its turn, guarantees the more efficient decisions-taking” [5].

The main difference of the big data lies in: their amount, speed and variety.

The data were selected from one source and were kept in the single format for quite a long while. At present, the data may be of any forms and sizes: text, sound, video, graphics. Thus, Big Data grant the possibilities for the use of the most varied data and their combinations, for the creation of the new methods of data collection and processing in future. The unification of the heterogeneous information storehouses for the construction of the complex inquiries will be efficient.

The important characteristic feature of the big data is the speed of the new information generation, it's also important, how rapidly the data are accumulated from different systems for the next processing.

The Big Data sphere may be conditionally divided into two categories – the elaboration, i.e. the creation of systems for calculations and processing of the data scope (Big Data Engineering) and the analytics, i.e. consolidation and the analysis of data (Big Data Scientist / Data Analyst). They are different in the character of tasks.

Working with Big Data, you should be able to process the vast information scopes and to interpret the results, to able to fulfill the previous processing of data, to create clusters, to analyze information, to comprehend the mathematical statistics, the analytics, the theory of algorithms, the numerical methods, the theory of probabilities and to have good knowledge of the higher mathematics.

You may separate the three main specializations of the employees for the operation in the field of data processing:

- ✓ Data Engineer
- ✓ Data Scientist
- ✓ Data Manager

Data Engineer – is a professional in projecting of such data processing systems that operate with the petabytes of data. He masters all the modern technologies and approaches in the sphere of data processing: MapReduce, Hadoop, Spark, Aerospike, Redis, Storm and so on.

Data Scientist - “data researcher” – is able to find the objective regularities in the big data scopes, knows the sphere of machine training well, masters confidently of such instruments as R, Weka, Python + Scikit-Learn + Pandas. The very Data Scientist is able to get the maximal benefit from data and to project the algorithms that will give answers to the necessary questions. But the main force of the data researcher - is in the knowledge of the subject area: he knows where to find so much required “gold grains” [6].

The thorough intellectual analysis of big data allows see the hidden regularities, being unseen for the limited human perception. It gives the unprecedented development possibilities for all the spheres of our life: the state control, education, medicine, telecommunications, economy, transport, production and so on.

You may separate the following directions in the sphere of Data Science:

- ✓ Data Mining – allows solve the tasks of the prognostication, the probability of events and results.
- ✓ Text Mining – allows find the regularities in the text, define its topics automatically and determine its psychological implication.
- ✓ Processing of pictures – allows find images on photos, recognize the text on the picture, reveal diseases, according to photos and so on.
- ✓ Processing of audio-signal – gives an opportunity to give the commands and to receive responses by voice.
- ✓ Recommendation systems – provide the service, corresponding maximally to the interests of the user;

Data Manager – the professional, controlling the project competently. He (she) is obliged to know the possibilities of modern technologies, to understand the subject sphere, to master the special terminology and also to have good skills in the control technologies of information projects.

The Corporation EMC conducted the research, in the process of which it was revealed that the use of Big Data leads to the essential improvement of the decisions-taking processes, increases the competitiveness of the companies and simplifies the control of risks.

Big Data — are the system, the qualitative transition to the composition of the values' chains, based on knowledge. According to the effect, it may be compared with the appearance of the accessible computer engineering at the end of the last century.

The Big Data technologies are often examined for the solvation of the following tasks [1]:

- ✓ prognostication of the market situation;
- ✓ marketing and optimization of sales;
- ✓ efficient segmentation of clients;

- ✓ improvement of goods and services;
- ✓ taking of more grounded decisions on the basis of Big Data analysis;
- ✓ optimization of the investments portfolio;
- ✓ increase of labor productivity;
- ✓ efficient logistics;
- ✓ monitoring of the fixed capital's state.

The similar use of Big Data is evident and clear. But the presented technologies give more opportunity for the realization of the serious psychological tests. You may obtain the most interesting conclusions on the basis of the data synthesis from the social nets, due to the revealing of the hidden regularities.

The principally new effect from the massive use of this approach in the data processing starts to become distinct for the last years. The scientists are searching for the hidden correlations between the researched phenomenon (object, process) and the thousands of the other factors, where the vast statistics, having been accumulated for the long years, was used as the initial data. The use of these empirically closed regularities promises the progress in the development of many scientific directions.

The complex modern Big Data models reveal still more often certain irrational-fantastic, at the first sight, dependencies that allow have a look at far away beyond the borders of the famous scientific picture of the world [1].

Due to that, Big Data are sometimes called as “the new astrology of XXI century”. And this is the result of the smooth transition from the quantity of information to its quality, when the machines become capable to reveal the principally new dependencies, being inaccessible beforehand for the limited human consciousness.

Big Data today are demanded most of all in both: trade and business. Big Data allow earn not only the additional money, but are actively used for the more efficient control of society.

For example, the so-called Predictive Policing has been already used in the USA for the third year. This is the specialized computer system of the USA police, being elaborated on the principles of the “big data” analysis, which helps to predict the splashing time of crimes and also the definite areas of their escalation. Such an approach has quite definite results – the noticeable decline of criminality has been already fixed in the cities, where the presented system was introduced. For example, the number of the arrests in Santa-Crus was increased by 55 %, but the quantity of robberies and the cars stealing was reduced by 10-15% [1].

No doubt, the predictive abilities of Big Data have the impressive possibilities. The company Farsite fulfilled the forecast by the Big Data technology before the presentation with the cinema reward “Oscar”. The analysis was made on the basis of the data scope, prepared on the historical cut of last 40 years of the American cinema industry, the statistics of presentation with the different cinema rewards, the frequency of the actors/films quoting in press. As a result – the exact

hitting of the target for all the nominations. And there are many similar examples.

2.1 Problem Positing

At present, Big Data is a special approach – the ideology of information processing, being used for the processing of big scopes of the “raw” data. The evident transition occurs from the initial function into the new functional form. If they were simply the calculation methods that allowed process the fantastic volumes of entering data beforehand, then, at present, they are the complex self-studying algorithms, which ideally allow not only efficiently “press” the information, but are capable of becoming complicated and improving independently. After such a qualitative jump, the instruments, created on the basis of these technologies, become the general and the universal ones, penetrating far beyond the borders of the Internet. Big Data technologies open the new approaches, technologies, the possibilities of use [4].

The analytical companies point out the interesting distortion in the use of Big Data: many world energetic companies already collect and accumulate the statistical information, spending hundreds of millions dollars for that. Only 1% of the allocated sums is spent in general for the very processing and data analysis at that. Naturally, the real effect from the Big Data is practically absent at such position of the affairs [2]. This paradoxical fact reveals the main problem of the Big Data current state, its narrowest place. According to the calculations of the IBS company, only 1,5% of the accumulated data amounts had the information value in 2013. It's clear that to collect all the data one after the other is useless, the priorities are needed and they are defined by the task-setting.

It's interesting that education and medicine, where the relatively small investments into the “big data” begin to give the sensible return already now, are mostly separated by the analysts among all the possible directions of growth. In the opinion of some experts, the massive introduction of Big Data into these branches may increase the quality level of human life in the nearest terms [3].

As it has been already said, the amounts of information grow rapidly (by 50% yearly). The most part of the information, being kept in the world, is non-structuralized, i.e. not suitable for research and use. Big Data – is the tool, which allows adapt to this scope of information.

The classical SCDB, such as Postgres, MySQL, Oracle have no such flexibility in scaling (scoping) at the processing of big data amounts and the further processing becomes impossible at the increase of the amounts.

The spreading of the Big Data technology changes greatly the format of interaction between the consumers in all spheres of life: retail, medicine, transport, insurance, banking sphere. The analysis of big data helps to know more of the preferences of clients, gives an opportunity to make the personalized offer and to

recommend the product or the service, being necessary here and now [5].

2.2 Aim of Research

The big scope of data became to accumulate during the boom of the social nets.

According to the statistics, nearly 2 milliard inquiries are processed by the Facebook every day. Google receives 5,5 milliard inquiries every day. And this is not the limit as the amounts of information grow constantly. The obtained data are analyzed and used for the increase of the contents quality. Due to the thin adjustment of the news ribbon of Facebook, (Facebook has never revealed the operation principle of the news ribbon algorithm, at this, the company restores it regularly), only those posts get into it, which the user will probably like. "Almost everything is taken into account: everything you reacted with your like or the sad like, the fact that you commented, what message you have opened, whether you watched the video to the end, whether you read the post text till the end. The contents, capable to interest the user with the most probability, gets into the ribbon finally. The algorithmic ribbon was included into Instagram, VC and the other social nets after Facebook. There are some working actions in YouTube too: likes, subscriptions to the channel, comments and reviews to the end" [7].

"You may compare any action in the Internet with the traces on the clean snow. Online-purchases, photos, activity in the groups in the social nets, the lists of friends and even the likes "reveal" the user. The history of the searching inquiries also characterizes the user; you may define the sex, the age, the circle of interests. Smartphones send data on the movements. The interested companies scrupulously accumulate the received information and buy it [8]. It gives the powerful push for the development of the Big Data analysis.

According to the above-mentioned, the decision to use the Big Data technologies for the analysis of the users of the Internet sites becomes urgent.

The tasks, which are needed to be solved at putting the Big Data projects into practice, are not only technological. Data collection is not the most complex problem. Every user leaves many "finger-prints" in the Internet that may be used for the analysis. But you may not always receive any benefit from the data, due to the lack of ideas.

You may represent each social net in the view of the mathematical graph, as the totality of the units and ribs between them. It's accepted that the unit is the channel, but the rib, between the two units, is information, by which the channels exchange. The size of the unit depends on the number of subscribers, but the width of the rib – on the scope of information, passing between them.

In order not to "drown" in the sea of information, we made the selection, according to the following criteria:

1. channels in the Ukrainian language;
2. channels with the number of subscribers of more than 500.

We reduced the number of channels by that.

In order to facilitate the further work, it's advisably to divide the channels, according to the categories, but this is not enough too. As Victor Delisov advises, in order to solve this problem, it's necessary to use the cluster analysis which may be conditionally presented in the following formula:

$$Y=F(N,A,C), \quad (1)$$

where Y – type of behavior (cluster);

N – number of recalls in the period t_1 and t_2 ;

A – speed of reviews' selection for $(t_2 - t_1)$;

C – coefficient of the category reposting;

F – function [9].

We would receive the definite number of clusters, which allow us estimate the channels and information in result.

2.3 Practical Realization and Research Results

The research, conducted by us, is the analysis realization of the big data, received from the Ukrainian sites, which use the information agencies (sport.ua, 112ua.tv, pravda.com.ua, rozetka.com.ua, bigmir.net, work.ua, OLX.ua and etc.). In order to fulfill the research, it's supposed to use the possibilities of the programming R language as one of the leading instrumental means of the Big Data technology [4].

It's planned to analyze the attendance frequency dependence of different sites by the young people, aged from 17 to 35 years old, upon their sex, job provision, incomes and age.

We planned to define the other dependencies between those indicators too. Big Data, using the vast information scopes of various data about the net users, find the completely non-evident correlations between the sex, income and the inclination to the Internet use.

The part of the big data, needed for us to make the analysis, are located in the special storehouses DB and KB and are ready for the use. But the access to them is not always granted freely. The special information companies – the suppliers of information – are used. They suggest the definite types of services to the consumers.

The collected data are formed by us as a table of the text information. The table consists of four columns (indications):

- sex of a young man (POL: female = 0 and male = 1);
- occupation: (ZAN: student = 1, working person = 2, not working person = 3);
- incomes in UHA. (DOH)
- age (VOZR):

Collected data on 700,000 people.

The initial data for the analysis have the following view: see Fig.1.

	POL	ZAN	DOH	VOZR
1	0	1	1100	17
2	1	2	2500	26
3	1	3	500	28
4	0	1	1100	18
5	1	2	3000	27
6	0	3	500	29

Fig.1. Part of Data, Collected for Analysis.

```
> str(data)
'data.frame': 51 obs. of 4 variables:
 $ POL : int 0 1 1 0 1 0 1 0 1 0 ...
 $ ZAN : int 1 2 3 1 2 3 1 2 3 1 ...
 $ DOH : int 1100 2500 500 1100 3000 500 1100 4000 500 1100 ...
 $ VOZR: int 17 26 28 18 27 29 19 28 30 20 ...
```

Fig.2. Structure of Data File.

Let's select data only for the female sex, i.e. (SEX = 0):

```
data.f <- data[data$SEX == 0,]
```

```
> data.f <- data[data$POL == 0,]
> head(data.f)
  POL ZAN DOH VOZR
1   0   1 1100   17
4   0   1 1100   18
6   0   3   500   29
8   0   2 4000   28
10  0   1 1100   20
12  0   3   500   31
```

Fig.3 Data Are Only for Female Sex.

As one of the analysis variants, you may imagine the following lines. Let's create the separate object with the data for men, being older than 25 years old:

```
data.m.big <- data [data$SEX == 1 & data$AGE > 25,]
```

```
> data.m.big <- data[data$POL == 1 & data$VOZR > 25,]
> str(data.m.big)
'data.frame': 16 obs. of 4 variables:
 $ POL : int 1 1 1 1 1 1 1 1 1 ...
 $ ZAN : int 2 3 2 3 2 2 2 3 3 2 ...
 $ DOH : int 2500 500 3000 500 5000 6000 7000 500 500 5000 ...
 $ VOZR: int 26 28 27 30 30 31 33 35 30 35 ...
```

Fig.4.Results of Selection: Data Only for Men, Being Older than 25 Years Old.

Let's define the specific weight of a person (the relation of an individual's age to his/her incomes) — WEIG.R:

```
data$WEIG.R <- data$AGE/data$INC
```

```
> data$VES.R <- data$VOZR/data$DOH
> str(data$VES.R)
 num [1:51] 0.0155 0.0104 0.056 0.0164 0.009 ...
```

Fig.5. Specific Weight of a Person.

Our data, with the additional criterion, have the view, represented in Fig.6:

	POL	ZAN	DOH	VOZR	VES.R
1	0	1	1100	17	0.015454545454545455
2	1	2	2500	26	0.0104
3	1	3	500	28	0.056
4	0	1	1100	18	0.016363636363636364
5	1	2	3000	27	0.009
6	0	3	500	29	0.058
7	1	1	1100	19	0.017272727272727273
8	0	2	4000	28	0.007
9	1	3	500	30	0.06
10	0	1	1100	20	0.018181818181818182
11	1	2	5000	30	0.006
12	0	3	500	31	0.062
13	1	1	1100	21	0.019090909090909091
14	1	2	6000	31	0.005166666666666667
15	0	3	500	32	0.064
16	0	1	1100	22	0.02
17	1	2	7000	33	0.004714285714285714
18	0	3	500	34	0.068

Fig.6. Part of the Widened Data.

The programming R language gives the possibility for the fast calculation of the main parameters of the descriptive statistics: Min, Max, Median, Mean, 1st Qu., 3rd Qu., NA's and others. [9].

```
> summary(data$VES.R)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.003625 0.008375 0.018182 0.029150 0.060000 0.070000
```

Fig.7. Data Statistical Indicators of Net Users.

Let's analyze the criterion of incomes (INC). summary (data\$WEIG.R)

```
> summary(data$DOH)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   500    500    1100   2327   4000   8000
```

Fig.8. Statistical Characteristics of INC. Criterion.

Let's calculate the average meanings of incomes separately for all the combinations of the age and sex.

```
> tapply(data$INC, list(data$SEX, data$AGE), mean). The result is in the Fig. 9
```

```
> tapply(data$DOH, list(data$POL, data$VOZR), mean)
 17 18 19 20 21 22 23 24 26 27 28 29 30 31
0 1100 1100 1100 1100 1100 1100 1100 5000 NA 5000 500.000 500 500
1 1100 NA 1100 NA 1100 1100 NA NA 2500 3000 500 4833.333 2000 3250
 32 33 34 35
0 4250 2250 3800 4000
1 500 7000 NA 2000
> |
```

Fig.9. Average Meanings of Incomes Separately for All Combinations of Age and Sex.

Let's define the average arithmetical number for the occupation, separately for men and women.

```
tapply(data$OCC, data$SEX, mean)
```

```
> tapply(data$ZAN, data$POL, mean)
      0      1
1.870968 2.200000
> |
```

Fig.10. Results of Mean Function for Occupation, Being Differentiated, according to the Sex.

Let's define the average arithmetical number for the occupation separately, according to the age categories.

```
tapply(data$OCC, data$AGE, mean)
```

```
> tapply(data$ZAN, data$VOZR, mean)
 17 18 19 20 21 22 23 24
1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
 26 27 28 29 30 31 32 33
2.000000 2.000000 2.333333 2.500000 2.833333 2.666667 2.666667 2.333333
 34 35
2.400000 2.500000
> |
```

Fig.11. Average Arithmetical Number for Occupation Separately, according to Age Categories.

Let's calculate the standard deviation: (standard deviation) > sd(data\$WEIG)

```
> sd(data$VES)
[1] 0.02465665
> |
```

Fig.12. Indicator of Standard Deviation.

Let's calculate the dimensionless coefficient of the variation (CV):

```
> 100*sd(data$WEIG)/mean(data$WEIG)
```

```
> 100*sd(data$VES)/mean(data$VES)
[1] 84.58492
> |
```

Fig.13. Coefficient of Variation.

We calculated the correlation coefficients of Pearson, Spearman, Kendall, the coefficient of variation, the

standard deviation and the other statistical characteristics, defined the correlation degree between the variables [10]. We used 13 methods of the analysis in the process of the research. Due to the use of the programming R language as one of the leading instrumental means of the Big Data technology, we received the curious results, serving as the sufficient basis for the conclusions. Therefore, basing on the final results of more than 85 calculations, you may come to the following conclusions.

3 Conclusions

Our expectations of the fact that incomes depend significantly on the age, were not confirmed: there is some dependence, but the weak one. There is no link between the incomes and the sex of the young people. It's seen, according to the results of the calculations, the men visit sites more seldom.

We received the following results in the process of the scrupulous analysis of the data, taking part in the research:

1. Sites are visited by the men, being older than women;
2. There are more working ones and not working ones among men, there are more students among women;
3. Women have the higher incomes after they are 30 years old;
4. It is seen, according to the results of the calculations, that men visit sites more seldom than women;
5. You may come to the conclusion that the young people after 28 years old visit sites more often;
6. 39,2% of men and 60,8% of women take part in the experiment;
7. The older the individual is (at the age from 17 to 35), the higher are his/her incomes;
8. There is a weak link between the sex and occupation. Many other conclusions were received at the use of the Big Data technology that may be used for the following:

- ✓planning of the advertising production, designed for the definite auditorium of the net users;
- ✓at the study and the decision of social problems;
- ✓at the elaboration of the thematic (subjects') sites;
- ✓for the career-guidance work in the net;
- ✓for the employers;
- ✓at the decision of the gender problems;
- ✓at the organization of leisure-time of the young people and others.

The analysis of the big data allows get the exact information of the net users. The application of the new technological possibilities, being revealed by the Big Data, helps the corporations to become still more client-oriented. The analysis of big data helps to know more of the preferences of clients, gives an opportunity to make the personalized offer and to recommend the product or the service, being necessary here and now.

References

1. A. Dobrychenko. *How to Study Big Data — Experts Answer*.
<https://tproger.ru/experts/learn-big-data/>
Access - 10.01.2021.
2. O. Petrov. *Specialist on Big Dat*.
<https://rb.ru/opinion/resheniya-data-science/>
Access - 08.12.2020.
3. T. Susan. *Big Data: All You Need to Know*. PC Week/RE. № **25**. (2012).
4. I.N. Vdovychenko. *Big Data Technologies and Their Use for Analysis of Net Users*. III International Scientific-Practical Conference “Information Security and Computer Technologies”. Kropivnytsky. P.189-191. (2018).
5. P.D. Ivanov, V. Vampilov. *Big Data Technologies and Their Use at Modern Industrial Enterprise*. Engineering Journal: Science and Innovations. Iss. **8**. (2014).
6. I. Savchuk. *Big Data – Technology, Giving Birth to New Type of Business*. Magazine “Bit”. **3**. (2014).
7. I. Popov. *How Media Corporations Use Big Data*. <https://rb.ru/opinion/media-big-data/>
Access - 11.10.20.
8. A. Rychkov. *How FinTech Uses Big Data*. <https://rb.ru/opinion/big-data-v-fintech/> Access - 20.12.20.
9. V. Delisov. *How we learned to track emerging trends with mathematical precision*. <https://rb.ru/opinion/martech-advlab/>
10. R. I. Kabakov. *R in Action. Analysis and Visualization of Data in R Language*. Moscow. DMK Press. (2014).
11. G. Groulmund, K. Uikem. *R Language in Tasks of Science on Data: Import, Preparation, Processing, Visualization and Modelling of Data*. Dialectics - Williams. Moscow. (2019).