Saint-Petersburg State University

Manuscript copyright

Fursov Dmitrii Viktorovich

Intelligent system of management decision support in the task of information dissemination

Scientific speciality: 2.3.1. System analysis, control and information processing, statistics

Thesis submitted in conformity with the requirements for the degree of candidate of technical sciences

Translation from Russian

Scientific supervisor: doctor of physics and mathematics sciences, associate professor
Krylatov Alexander Yurevich

Saint-Petersburg
2024

Contents

Introduction

Research Topic Relevance

In the modern world, information transfer plays an important role and covers all spheres of human activity: from personal correspondence to ensuring the security of entire regions. Today, it is possible to disseminate information using a variety of channels and means, such as the Internet, television, radio, print media and others. Due to the development of technology, the time of information transmission has decreased, and technical means of data exchange have become available, so the volume of transmitted information has increased. According to the analysis of the «Global Digital 2023» [5] report for 2023, the number of Internet users exceeded 64% of the total world population compared to 2010 – 29%, and the number of active users of social networks amounted to more than 59%. These global changes in information technology have significantly affected processes in various subject areas. Companies, organizations and state institutions are forced to adapt to the new digital reality and for this purpose they need modern tools for big data analysis and modeling of processes related to information distribution[1]. Based on the peculiarities of a particular field of science, specialists apply appropriate mathematical methods and types of modeling to implement effective resource allocation, automation of management processes and decision-making support.

Various studies have highlighted the problem of dramatic growth in the number of Internet users worldwide and the multiplied volume of data generated by them, starting from 1990 to the present. To solve this issue, various approaches and methods from such sections of mathematics as: graph theory, simulation modeling, operations research, probability theory, mathematical statistics, machine learning, neural networks and others are used. This dissertation research considers optimization and machine learning methods in solving the problem of determining the sites of information distribution in mass communication media (MCM). Since in this thesis the MCM is understood as a social network[2], and the sites of information

[1] In accordance with Article 2 of the Federal Law of the Russian Federation dated 27.07.2006 № 149-FZ «On Information, Information Technologies and Information Protection»

[2] In accordance with part 1, para. 1 of Art. 10.6 of the Federal Law of the Russian Federation from

dissemination are the communities of the social network, then the results and conclusions obtained for the tasks formulated in the dissertation research can be used not only in the given subject area – information and communication technologies. This suggests that the proposed management decision support tool has the property of scalability.

The peculiarity of using the approach proposed in the thesis is the possibility of analyzing big data and feature space, as well as the formation of various scenarios of information dissemination and recommendations for decision makers (DM). The application of the optimization approach allows to form a set of sites for information dissemination in the MCM under given constraints and preferences, and the methods of cluster analysis allow, firstly, to reduce the dimensionality in the optimization problem with a large number of objects, and secondly, to obtain a well-separable partitioning with compact clusters. The advantage of the proposed approach is that the DM has an opportunity to analyze the scenarios of information dissemination obtained by searching for the optimal solution and constructing the partitioning taking into account the peculiarities of the feature description of objects.

The relevance of the research of problems related to the dissemination of information in the digital environment is due to the applied demand, which allows us to successfully apply the optimization approach and machine learning methods for the analysis of supply and demand in the market of goods and services, the implementation of propaganda activities in the MCM for political purposes, the analysis of behavioral activity of users by territorial affiliation and other tasks where the use of modern communication technologies is of crucial importance for the development of information dissemination in the digital environment. Thus, the development of intelligent tools to support managerial decision-making in the task of information dissemination in MCM with the use of tools for analyzing big data is one of the key tasks in modern technological society.

27.07.2006 № 149-FZ «On information, information technologies and information protection».

Literature review

The analysis of scientific literature on the application of mathematical methods to solve the problem of information dissemination in MCM shows that there are works considering various approaches and methods [11, 27, 35, 58, 92, 106, 107]. A striking example of integration of management decision support systems are tools using machine learning and artificial intelligence algorithms, such as "Albert" - artificial intelligence marketing platform [40] and "MTS Marketer" - advertising platform based on "MTS Big Data" [56] and others. However, many of them have a significant disadvantage – high implementation costs, which can afford only large market participants. In addition, it should be noted that most of the works, where the issue of information dissemination is considered, have an applied value either for marketing or model information contradictions [3, 4, 22, 38, 39, 42, 43, 47, 57].

Studying scientific papers covering various approaches to modeling the process of information dissemination, there are such as - [6, 7, 8, 9, 37, 60, 61, 90], which offer models for optimizing the distribution of resources, including financial resources, when conducting advertising campaigns. The development of such tools for decision support is a good means of analyzing the current situation of supply-demand and trends in the market of goods and services for the DM, but the result of the work of such models is not an answer to the question of where it is better to place an advertisement under given budget constraints. It should be noted that a large percentage of works in this direction is devoted to the construction of models that do not meet the requirements of scalability and do not present the possibility of integration with other systems for convenience and simplification of automation of management processes in the field of communication technologies [14, 15, 41].

The paper [17] considers the problem of mathematical modeling of an advertising company: the analysis of modern methods of assessing the effectiveness of advertising activities is carried out and the mathematical model of optimal distribution of the advertising budget is applied. Note that in this case the problem of information dissemination is formulated in such a way that it becomes possible to apply it in practice only in the context of narrowly focused marketing research, and the model of optimal budget allocation the author applies to identify recommendations for optimizing advertising policy. In addition, a distinctive feature of this work is that when building an optimization model, the type of advertising is taken as a variable, and the target indicator is efficiency. The disadvantage of the proposed approach to solving the problem is that there is no analysis of big data and do not use modern tools and software packages to implement the modeling process.

It is important to note that there are such works as [28, 29, 30, 69, 89, 110, 111], where the issue of choosing platforms for information placement is raised. This indicates that this area of research is quite popular among specialists of various industries. At the same time, the analysis of scientific publications has shown that optimization models using statistical data analysis[103], numerical methods and differential equations[18, 19, 25, 31, 44, 68, 86, 97] have become the main tool for identifying regularities in complex socio-economic systems[23]. However, due to continuous improvement of technical and mathematical applied solutions, there is a possibility of setting new types of tasks in this subject area.

Some researchers choose optimization methods [95] and numerical simulation tools [94] to determine the period of advertising records placement, effective characteristics [53, 54] and metrics for assessing the quality of marketing activities [21], to form an assessment of the influence of information impact on consumers [51], as well as to search for patterns in the media space [93]. It should be noted that the thesis research is a continuation of the research conducted by the author in the undergraduate and graduate programs [72, 73, 84]. Previously, the task of information impact in MCM with the application of methods of simulation modeling, descriptive statistics and construction of a knowledge base to create a prototype of an expert system using actual data was considered. In this regard, we can conclude that the task of modeling information dissemination using modern technical means and applied mathematical apparatus although well studied, it still remains relevant in the research of various models.

Research Purpose and Objectives

The purpose of this work *is to build an intellectual system using methods of system analysis of complex applied objects and modern methods of information processing to optimize the process of management decision support in the task of information dissemination in the MCM by developing new and improving existing methods and tools of analysis of information processing and management of complex systems.* The object of the research is information dissemination in MCM, and the subject of the research is a set of methods for theoretical and experimental analysis of the above process, as well as tools for modeling the results of information promotion using optimization and machine learning methods.

In order to achieve the objective of the research, the following tasks need to be accomplished:

1. *Formulate problem statements for determining the set of information dissemination sites.* This requires analyzing scientific literature sources to identify unique problem statements in the subject area.

2. *Analyze services that provide up-to-date MCM data and implement data import.* For this purpose, it is necessary to formulate criteria for selecting the most appropriate service.

3. *Analyze the existing approaches to building an intelligent decision support system.* It is necessary to analyze the sources of scientific literature devoted to theoretical and applied researches to solve scientific and technical problems related to information dissemination.

4. *Develop a program component that implements the algorithm of preprocessing of statistical data on user activity in MCM.* It is necessary to analyze the structure of files containing statistical data, select the appropriate data types and implement a cyclic algorithm for processing the selected data types.

5. *To develop architecture and realize a software complex for forming scenarios of information dissemination for DM using modern programming tools.* Identify a number of methods of teacherless machine learning and optimization, as well as methods of feature space compression and implement an intelligent system in the form of a software complex consisting of software components for solving the corresponding types of problems.

Scientific Novelty

In the dissertation research the architecture and scheme of the intelligent system of support of managerial decision making for DM in the task of information dissemination in MCM are proposed. The development and implementation of such systems is an urgent task in many spheres of human activity, including communication technologies. An optimization model and a complex model of cluster analysis for the formation of different scenarios of information dissemination with recommendations on record placement for each selected site have been built and programmatically implemented.

The models investigate the influence of seasonality, budget, client preferences and types of goods and services on the dynamics of behavioral activity of the audiences of the sites. Based on the analysis of the obtained results of numerical modeling, it was found that there is a seasonal component of the activity of site participants, as well as differentiation of the sensitivity of criteria to changes in preferences. In this research, the criteria are understood as quantitative characteristics of evaluating the involvement of the site audience in the information placed in it. The practical application of the developed system will allow to adjust the preferences of the DM in accordance with the specified product range, budget and time interval, as well as the selected key quantitative characteristics.

The space of attributes was analyzed and the most significant ones were identified. The influence of sets of significant attributes on obtaining the best scenario, in terms of quantitative characteristics values, depending on the time interval, budget, as well as the range of goods and services is studied. New formulations of the problem of information dissemination in the MCM were formulated. In addition, the methods of cluster analysis are successfully applied in the optimization problem for dimensionality reduction, which allowed to reduce the time of obtaining different scenarios of information dissemination. The dependence of the values of the number of clusters on the budget for the considered input parameters of the system in program blocks with the application of machine learning methods without a teacher is demonstrated. It is shown that when applying clustering methods for dimensionality reduction in the optimization problem the number of clusters changes by some small value, regardless of the used methods of cluster analysis and input parameters of the model.

All the main results presented in the paper were obtained by the author personally and are new.

Theoretical and Practical Significance of Research

This research work is of theoretical importance, as it contributes to the development of a key direction in the analysis of digital space. The research in the dissertation work on the process of information dissemination, as well as the analysis of scientific literature allowed us to formulate new problem statements in the context of automation of management issues in the field of communication technologies. Despite the fact that the work investigates well-known big data analysis and optimization models, we would like to note that the results obtained are of universal applied nature — in the sense of modeling processes occurring outside of marketing. Indeed, taking into account the specificity of the given subject area, the results of the research can be appropriately transferred to such processes as, for example, optimization of enterprise resources (economics and management), analysis of opinions and sentiments by regions (sociology), influence on the public masses to form the image of a political figure (political science). Thus, the importance of this research lies in the development of the theory researching the processes occurring in the MCM, as well as in the possibility of obtaining different scenarios of information placement on certain sites with recommendations.

The program complex developed by the author is of practical importance for DM in the issue of effective allocation of financial resources and can be used to support managerial decision-making in the task of information dissemination. The results obtained in the course of the research can be taken into account in the future when developing and implementing new functionalities of the proposed intelligent system. In addition, we would like to note that the architecture of the software under consideration is built in such a way that it satisfies the properties of scalability and integration with other information and communication systems.

Based on the above, it can be concluded that a successful attempt has been made to propose a universal approach to modeling the information dissemination process and to develop a new applied tool to support managerial decision-making, which can be used in management tasks for the effective allocation of limited resources of the organization when conducting relevant activities in the digital environment, as well as in the economic analysis of the current situation in the market of goods and services, sociology and political science tasks.

Structure of the Thesis and the Main Scientific Results

The structure of the dissertation research includes an introduction, three chapters presented with a division into sections and subsections, a description of the main results and conclusions — in each chapter, conclusion and a reference list containing 111 sources. The total volume of the work is 116 pages of typewritten text, contains 3 tables and 59 figures.

The introduction substantiates the relevance of the topic of the dissertation research, provides a review of literary sources and indicates the degree of development of the problem in the literature, outlines the main goals and objectives, specifies the methodology, and formulates the main results of the research. The scientific novelty, theoretical and practical significance of the work is presented.

In Chapter 1 formulates problem statements for modeling the process of information dissemination in MCM using optimization methods (section 1.1). We analyze the services that provide statistical data of the MCM, based on the structure of files containing data of the selected MCM, we develop and implement the algorithm of statistical data processing, which forms a matrix of object-attributes for the sites of information dissemination. Using the formed matrix, the characteristics allowing to estimate the involvement and feedback of the audience of the sites in relation to the information placed in them were determined. The block of forming recommendations on placement of paid publications on the platforms of MCM is realized. An optimization model for determining the sites of information distribution is proposed and implemented in software (Section 1.2). Numerical simulation with the given input parameters of the model, as well as sensitivity analysis of criteria in the problem of multi-criteria optimization, which will allow to adjust the preferences of users of the IT-product developed by the author. Based on the analysis of the obtained results, a number of conclusions of this research are formulated (Section 1.3).

In Chapter 2, we formulate problem statements for modeling the information dissemination process in the MCM using machine learning and optimization methods with preliminary clustering to reduce the dimensionality and reduce the time of formation of information dissemination scenarios when the number of site objects increases (Section 2.1). The architecture and software implementation of a comprehensive model using optimization and machine learning methods without a teacher is developed and its description is proposed. The methods of feature selection and feature extraction for teacherless learning tasks are considered, the results of applying the mentioned methods are given and observations are formulated. Cluster analysis methods are considered and their hyperparameters are specified, and metrics for evaluating the quality of the resulting partitions are given. The software implementation of blocks realizing the proposed complex model is described (Section 2.2). Training of machine learning models and comparative analysis of modeling results are carried out, and a number of conclusions of this research are formulated (Section 2.3).

In Chapter 3 the developed architecture of the intelligent system of support for management decision-making in the task of information dissemination in the MCM is presented in the form of a scheme (Section 3.1). The features of implementation and application of the system are considered, the scheme of data storage structure is given (Section 3.2), and a number of conclusions of the research based on the comparative analysis of numerical simulation results are formulated (Section 3.3).

The conclusion provides a brief discussion of the results obtained and possible directions for further research.

Research Methodology and Methods

The tools involved in the work are generally recognized rules and approaches to research activities in the field of applied mathematics: mathematical programming (theory and methods of optimization problem solving), machine learning (methods of cluster analysis and feature space compression), mathematical modeling, comparative analysis, numerical simulation in a cross-platform integrated development environment for the Python programming language — PyCharm.

Degree of Credibility and Evaluation of Results

The main results obtained in the course of the research were discussed and presented as reports at the following scientific conferences: International Online Conference «The 6th Computational Methods in Systems and Software 2022 (CoMeSySo2022)», Section: «Data Science and Algorithms in Systems», Prague, Czech Republic. Prague, Czech Republic [101]; All-Russian Conference on Natural Sciences and Humanities with International Participation «Nauka SPbSU 2023», section: «Mathematics, Mechanics, Informatics», St. Petersburg, Russian Federation. St. Petersburg, Russian Federation; «XIII Congress of Young Scientists ITMO 2024», sections: «Big Data and Machine Learning» [75], «Artificial Intelligence and Behavioral Economics» [71], St. Petersburg, Russian Federation; VI Congress of Young Scientists ITMO 2024», sections: «Big Data and Machine Learning» [71], St. Petersburg, Russian Federation. St. Petersburg, Russian Federation; VI All-Russian with international participation scientific and practical conference of students, postgraduates and workers of education and industry — «Management systems, information technologies and mathematical modeling — 2024» within the framework of the I International Forum «IT. Science. Creative» (iFORUM), section: «Applied Mathematics and Informatics in Humanities and Socio-Economic Sciences», Omsk, Russian Federation [83], Scientific Seminar of the Department of Mathematical Theory of Economic Decisions, St. Petersburg State University, St. Petersburg, Russia.

The reliability and validity of the results of this dissertation research is ensured by the correctness of problem statements, arguments and conclusions, as well as by the receipt of positive reviews from members of editorial boards of periodical scientific editions, in which the main results of the work were published.

Publications

The results of the conducted research have been published in 6 scientific publications [71, 74, 75, 76, 77, 101, 83], including, the main results of the dissertation work were published in three scientific journals [74, 76, 77], included in the list of peer-reviewed scientific editions recommended by VAK RF and included in RINC. 5 certificates of registration of the computer program [78, 79, 80, 81, 82] in the Federal Institute of Industrial Property (FIPS) were obtained.

The author's personal contribution consists in independent determination of the aim, tasks and research plan of the thesis work. The author independently conducted numerical experiments and developed program components and complex. Interpretation, statistical processing and final evaluation of the obtained results, analysis of scientific literature, as well as writing the text of the dissertation were carried out by the author independently.

Main scientific results

1. Problem statements for modeling the process of information dissemination in MCM using optimization methods are formulated. The described results were obtained in the first chapter of the research and published in the paper [76].

2. Problem statements for modeling the process of information dissemination in MCM using machine learning methods are formulated. The described results were obtained in the second chapter of the research and published in [77].

3. We have developed a software component that implements a cyclic algorithm for preprocessing statistical data on the user activity of information sites in the task of information dissemination in MCM in the Python programming

language in the cross-platform integrated development environment PyCharm. The described results were obtained in the first chapter of the research and published in [76, 79].

4. A software component with a recommendation block for forming information dissemination scenarios in MCM and solving optimization problems with the ability to transform and visualize information in the Python programming language in the cross-platform integrated development environment PyCharm has been developed. The described results were obtained in the first chapter of the research and published in [76, 81].

5. We have developed software components using machine learning and teacher-less feature selection methods to solve the problem of clustering of information sites and the problem of dimensionality reduction in the optimization problem with the possibility of transformation and visualization of information in the Python programming language in the cross-platform integrated development environment PyCharm. The described results were obtained in the second chapter of the research and published in [77, 80, 82].

6. The architecture has been developed and the intelligent system of management decision support in the task of information dissemination in the MCM has been programmatically implemented, and the scheme of data storage and modeling results with the possibility of transformation and visualization of information in the Python programming language in the cross-platform integrated development environment PyCharm has been proposed. Information dissemination scenarios are compared and the feasibility of forming several sets of information sites is demonstrated. The described results were obtained in the third chapter of the research and published in the paper [74, 78].

7. A tool for numerical simulation of the system under research has been developed, which allows to perform sensitivity analysis of criteria, as well as to analyze the process of forming unique scenarios of information dissemination as a result of changing preferences in the problem of multi-criteria optimiza-

tion on the example of the market of goods and services in the digital environment, taking into account the nomenclature of goods, budget and time interval. Application of the developed software components allows to correct user preferences by adjusting hyperparameters of machine learning methods, as well as to reduce the time of formation of scenarios of information dissemination. The analysis of feature importance allowed to determine the basic set of significant characteristics of objects by the selected methods of feature space compression. The described results were obtained in the first and second chapters of the research and published in [76, 77].

Main results to be Defended

1. Formalization and formulation of problems of information processing and modeling of information dissemination process in MCM using optimization and machine learning methods.

2. Special algorithmic and mathematical support of the intellectual system of management decision support in the field of information dissemination in MCM using modern methods of information processing and analysis, optimization and machine learning.

3. A program complex for realization of the problem-oriented intelligent system of management decision support in the task of information dissemination in MCM.

4. Information visualization, transformation and analysis functions using current computer information processing techniques in a BI tool for the task of information dissemination in the MCM.

5. Methodology of applied research aimed at identifying, measuring and analyzing the emerging market conditions of goods-services in the digital environment, as well as modeling scenarios of information dissemination in the MCM.

Acknowledgements

The author would like to express his gratitude to his supervisor, Doctor of Physical and Mathematical Sciences, Professor of the Department of Mathematical Theory of Economic Decisions, St. Petersburg State University, Alexander Yuryevich Krylatov, for his comprehensive support, responsiveness and assistance in the preparation and execution of the thesis. I would also like to express my gratitude to my first supervisor and mentor, candidate of physical and mathematical sciences, associate professor of St. Petersburg State University Mikhail Vladimirovich Svirkin, for support and consultations during the thesis research, valuable comments and recommendations. Besides, it is impossible not to mention my colleague and not to express gratitude to Vitaly Alekseevich Kochevadov, for evaluation of works and recommendations in preparation and execution of some sections of the dissertation. In conclusion, I would like to thank my high school math teacher Nelly Iskanderovna Amirkhanova, a professional, a person who demonstrated how interesting and fascinatingly the world of mathematics can be.

Chapter 1.
Modeling of site sets in the task of information dissemination based on optimization methods

1.1. Problem statement and description

Application of mathematical models for modeling processes in various applied areas plays a tangible role in making managerial decisions. With their help, specialists can predict or evaluate the behavior of the system, simulating various scenarios when input data changes or when external factors affect the system, predict the results of any processes, thereby minimizing losses, both in monetary and reputational terms, maximize the profit of the enterprise, and much more. Optimization models are an integral part of the mathematical toolkit used both by various government institutions and businesses to help decision makers in complex environments to conduct a complete and objective analysis of the subject activity. In this chapter some possible formulations of optimization problems of information distribution site selection in MCM will be considered.

Since the concepts in question cover a very wide range of means, tools and ways of providing information to an indefinite circle of persons, and also based on the specifics of the experimental part of this research, it is proposed that the dissemination of information should be understood as its promotion by publishing an advertising record[1], and under MCM - social networks[2]. For numerical modeling and ease of interpretation of the results was taken social network «VKontakte» [64]. In this paper, the promotion of information will be carried out in terms of the communicative impact of [26] on its consumer. To demonstrate the performance of the proposed modeling approach, we will consider a number of marketing tasks, such as increasing sales, positioning the brand of the [32, 59] or creating a corporate image of the [20]. Due to the fact that the promotion of information will take place

[1] In accordance with the Federal Law of the Russian Federation dated 13. 03.03.2006 N 38-FZ (ed. of 11.03.2024) "On Advertising"

[2] In accordance with Article 10.6 of the Federal Law of the Russian Federation of 27.07.2006 N 149-FZ (ed. of 12.12.2023) "On Information, Information Technologies and Information Protection"

within the framework of social networks, the site is understood to be a community of the social network [46]. In the community administrators publish records both on a free basis within the theme of the group, and for commercial purposes. Such records are called advertising records. The cost of placing such a record depends on many factors: the activity of the audience, topics, time of year and so on. Accordingly, the person who contacts the community administrators to publish an advertising record is a client.

In order to evaluate the possible results of advertising campaigns, data reflecting the activity in the community is required. In this paper, audience activity in the community is assessed using various indicators and metrics for a selected time period due to the limited possibility of obtaining data containing other complementary information.

So, let's consider possible formulations of the problem of forming a set of communities to place advertising records in them.

Integer Linear Programming Problem

Meaningful formulation of the problem: a client needs to increase the volume of products it sells within a certain budget. The client wants to run an advertising campaign in such a way that as many web users as possible learn about its product. It is required to maximize the total number of views of the published advertising records given the subject, time period, target parameter and budget. Note that the same advertisement record can be published in several communities, and also the advertisement record can be edited separately for each selected community taking into account the peculiarities of its audience.

Mathematical formulation of the problem: let the universal set of communities X, community topics, the desired month for posting information t $\left(t = \overline{1{,}12}\right)$, and the budget of the advertising campaign $(P > 0)$ be defined. Order the set X, i.e., establish a one-to-one correspondence between the sets X and $M = \{1, \ldots, n\} \subset N$, which allows us to specify the cost of placing an advertising record in the i-th community as $b_i > 0$, the value of the selected target indicator in the i-th community for the average statistical record in a given month of the year t – $c_i(t) > 0$, $i \in M$.

The solution of the problem is represented by defining a set of communities $x = (x_1, \ldots, x_m) \subseteq X$, $1 \leq m \leq n$, which satisfies the following requirements:

$$f(x) = \sum_{j=1}^{m} c_j \cdot x_j \to max,$$

$$\sum_{j=1}^{m} b_j \cdot x_j \leq P,$$

$$x_j \in \{0; 1\},\ j = \overline{1,m}.$$

Multi-objective optimization

Meaningful formulation of the problem: the client needs to carry out brand positioning within a certain budget. However, the client favors communities where people leave feedback in comments under posts and actively share community posts on their personal page. The client needs to maximize the number of comments and "share" marks on the published advertising record for a given topic, time period and budget. Thus, for the client two indicators are in equilibrium, the values of which will be maximized. Note that one and the same advertising record can be published in several communities, as well as the advertising record can be edited separately for each selected community, taking into account the characteristics of its audience.

In this research, the criteria are understood as «Likes», «Share», «Comments», «Views». In general, with a greater variety of data, the criteria can be, for example, coverage[3], the number of clicks on links, and so on.

Mathematical formulation of the problem: to the notations introduced on page 21, we add 4 criteria defined by functions: $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_4(x)$; we redefine the cost of placing an advertisement record in i- community — $g_i > 0$, and set the values of the corresponding criteria in i- community for an average record in a given month of year t — $a_i(t) > 0$, $b_i(t) > 0$, $b_i(t) > 0$, $c_i(t) > 0$, $d_i(t) > 0$, $i \in M$. The solution of the problem is represented by defining a set of communities $x = (x_1, \ldots, x_m) \subseteq X$, $1 \leq m \leq n$, which satisfies the following requirements:

[3] Unique views of the record by social network users

$$\begin{cases} f_1(x) = \sum_{j=1}^{m} a_j \cdot x_j \to max, f_2(x) = \sum_{j=1}^{m} b_j \cdot x_j \to max, \\ f_3(x) = \sum_{j=1}^{m} c_j \cdot x_j \to max, f_4(x) = \sum_{j=1}^{m} d_j \cdot x_j \to max, \\ \sum_{j=1}^{m} g_j \cdot x_j \leq P, \\ x_j \in \{0; 1\},\ j = \overline{1,m}. \end{cases} \tag{1.1.1}$$

Applying the method of criteria convolution, we reduce the system (1.1.1) to the system (1.1.2) and solve it using known methods of mathematical programming.

$$\begin{cases} f(x) = \sum_{j=1}^{m} w_j \cdot x_j \to max, \\ \sum_{j=1}^{m} g_j \cdot x_j \leq P, \\ x_j \in \{0; 1\},\ j = \overline{1,m}. \end{cases} \tag{1.1.2}$$

where $w_j = \alpha_1 \cdot a_j + \alpha_2 \cdot b_j + \alpha_3 \cdot c_j + \alpha_4 \cdot d_j$ is a measure of total activity in j-community for the «average» record in a given month of year t, $\alpha = \{\alpha_1, \alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ are the criteria weights or customer preferences.

1.2. Optimization model

The architecture of the optimization model is presented as a block diagram in Figure 1.1. The considered software component is implemented by means of four main functional blocks:

1. data preprocessing;
2. data processing;
3. forming recommendations for publishing advertising records;
4. optimization.

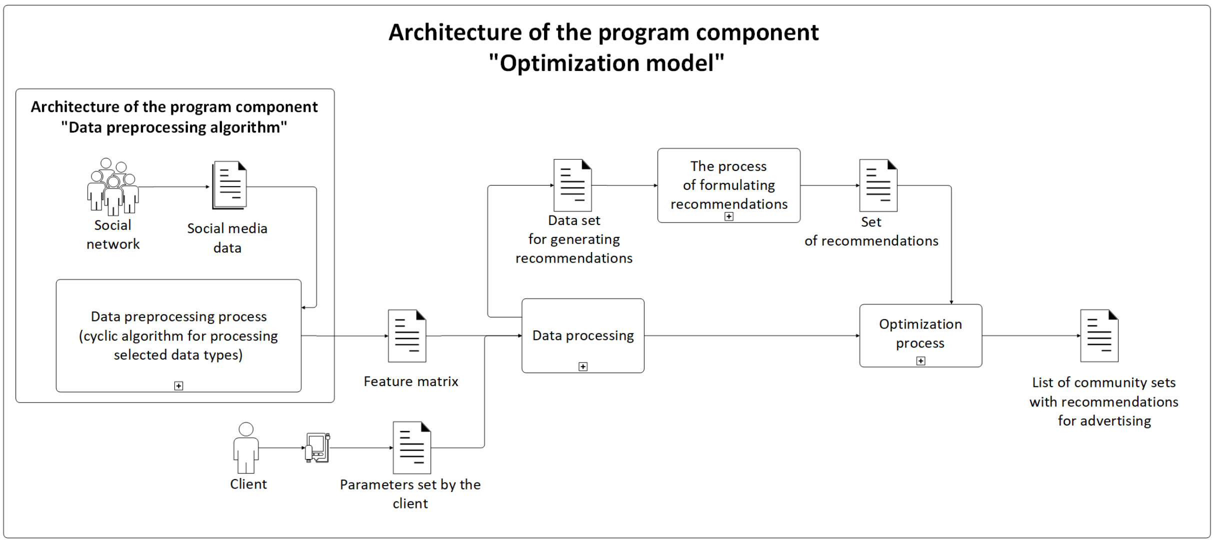


Figure 1.1: Architecture of the program component «Optimization model»

However, before proceeding to the consideration of the structures and features of the implementation of these functions, it is necessary to analyze the sources that provide relevant statistical data of communities of the social network «VKontakte». Note that the development of its own software package that solves the problem of parsing data[4] for various Internet platforms is a separate labor-intensive task, so in this study the author decided to use existing tools to extract statistical data on communities of the social network.

1.2.1. Analyzing sources of relevant statistical data

In order to conduct information dissemination activities in the MCM, it is necessary to analyze the data, access to which is provided by special services. This section will list a number of such resources and select one that meets the following criteria:

1. Versatility of use — the service provides statistics on various social media resources.

2. Ability to do uploads on a selected number of communities.

3. Ability to specify an upload period.

[4] Parsing data is the extraction of structured information from unstructured or semi-structured data.

4. Sufficient number of community characteristics provided by the service (more than 10).

5. Upload format: tabular by characteristics. Desirable file format: xlsx, txt, csv, json, xml.

6. Cost of using the service.

The criteria formulated will enable the researcher to find relevant statistics in a short time period.

Let's take a look at a few of the most popular resources among internet marketing industry professionals:

- «LiveDune» — The service checks accounts for spoofing and provides other various statistics necessary to make a decision on advertising placement [63]. The service also provides hourly statistics on campaign and competitor accounts. Meeting the formulated criteria:

    1. Media resources: «VKontakte», «Odnoklassniki», «Telegram», «TikTok», «YouTube».

    2. There is, but it is severely limited. To get data on 3000 communities, you need to pay more than 9990 rubles.

    3. Available.

    4. More than 50 different metrics.

    5. Download format - pdf, xlsx.

    6. The cost is more than 9990 rubles per month.

- «Pur Ninja» — is a service for delayed posting and analytics in social networks [1]. It knows how to publish videos, add watermarks to media, and display exactly how a post will look once it's published to social media. Meeting the articulated criteria:

    1. Media resources: «Telegram», «VKontakte», «Odnoklassniki».

2. Available, but severely limited. To get data on 3,000 communities, you need to pay for a 30x «Business L» plan.
3. Available.
4. Sufficient metrics.
5. Download format - xlsx, csv.
6. The cost - from 200 thousand rubles per month.

- «Popsters» — is a service of content analytics, statistics and comparison of communities in 12 social networks [62]. It allows you to: evaluate the popularity of different posts, taking into account the content, format, text volume, etc.; allows you to quickly and automatically calculate the effectiveness of posts with different hashtags and attachments; allows you to get statistics, analyze and compare the effectiveness of different campaigns in different communities. Compliance with the formulated criteria:

  1. Media resources: «VKontakte», «Odnoklassniki», «Telegram», «TikTok», «Pinterest» and «YouTube».
  2. Yes, there are, no restrictions.
  3. Available.
  4. More than 20 metrics.
  5. Download format - allows you to generate and download in a convenient format (xlsx, jpg, pdf, pptx, png, csv) reports and graphs based on statistics.
  6. The cost - from 499 rubles per month.

- «JagaJam» — is a company that creates services for working with data from social networks [65]. Statistics of accounts in social networks: the best posts, dynamics of subscribers, engagement, «Views», «My Like», «Comments» and «Share» for any period. The company provides: rating of communities, allowing you to choose the best ones for promotion; data in a convenient form for comparison (up to 10 communities at a time) and much more. Compliance with the formulated criteria:

1. Media resources: «VKontakte», «Odnoklassniki», «Telegram», «TikTok», «YouTube» and others.
2. Yes, there are, no restrictions.
3. Available.
4. Sufficient quantity.
5. Cost - from 9890 rubles per month.
6. Download format - xlsx.

- «AllSocial» — is a service with a lot of useful analytical information of communities of the social network «VKontakte» [66]. The service provides information about the cost of advertising record placement in communities, which corresponds to the exchange price of «Sociate». Compliance with the formulated criteria:

  1. Media resources: «VKontakte».
  2. Yes, there are, no restrictions.
  3. Available.
  4. Sufficient quantity.
  5. The cost is free.
  6. Download format - xlsx.

From the listed services were chosen — «Popsters» and «AllSocial». The first service satisfies most of the criteria and has a rather convenient structure of statistical data upload files for their further conversion and use. The second service contains additional metrics, statistical data, as well as one of the key characteristics for this study — the cost of placing an advertising record in a social network community, which was obtained by the second service from the source[5] called «Sociate», which is an advertising exchange [13].

It should be noted that all data collected are publicly available, non-confidential and non-personal, they can be obtained by any Internet user and belong to its open segment.

The development of an algorithm for preprocessing statistical data of user activity in social networks is a key step to ensure quality analysis of user behavior. Let's proceed to the description of this algorithm.

1.2.2. Algorithm of preprocessing of statistical data

Description of the structure of statistics files and feature space

As a result of the analysis of services that provide analytical information, the data from «Popsters» and «AllSocial» services for the period from 01.06.2021-31.06.2022 was uploaded. In the first service the number of communities amounted to more than 3600, and in the second service an upload of 10000 communities was formed. Each of the uploads has its own structure and its own feature description of the community. Let's analyze the features of each of them in detail. The «AllSocial» service provides statistical data in the form of a feature matrix, hence it does not require the development of an algorithm to build a matrix of this kind. The structure of the generated file is presented in the figure 1.2.

1	Сообщества	id	URL	Прирост за 1 день	Прирост за 7 дней	Прирост за 30 дней	% Offline более месяца	Кол-во постов	ER (100 последних постов)	Цена Sociate	CPM Sociate	Посетители (среднее за 7 дней)
2	MDK	57846937	https://vk.com/mudakoff	127	724	7523	25,5	89745	3,36	55440	155	154856
3	4ch	45745333	https://vk.com/4ch	134	-187	-984	26	89974	11,54	20160	92	61269
4	NR.Music	29573241	https://vk.com/nrmusicru	-51	-117	-3309	23,02	81332	2,91	10080	112	290449
5	БОРЩ	460389	https://vk.com/borsch	-5	200	-1732	22,8	164973	4,85	13090	63	70231
6	Академия Порядочных Парней	45595714	https://vk.com/academyofman	-112	-664	-3985	30,07	136807	2,57	8208	98	42876
7	Science|Наука	29559271	https://vk.com/sci	-477	-2584	-11833	23,11	70937	5,85	11376	91	11095
8	ПОЗОР	71729358	https://vk.com/styd.pozor	354	3199	12973	23,86	43218	5,15	5040	48	117568
9	Лепрозорий	65960786	https://vk.com/leprazo	-33	-768	-4644	21,63	72708	11,58	7128	68	31973
10	MARVEL/DC	32370614	https://vk.com/marvel_dc	4	182	-873	21,84	45720	13,47	10080	86	54926
11	Идеи дизайна интерьера	36184135	https://vk.com/i_des	-44	-805	-3464	21,74	71840	2,62	5775	101	13739
12	Достойные фильмы	33769500	https://vk.com/theworthyfilms	79	1096	4598	26,4	74196	1,86	3600	74	20505
13	Наука и Техника	31976785	https://vk.com/science_technology	80	1489	1036	23,93	57678	2,93	10656	101	15522
14	Лучшие стихи ВП	Литература	38683579	https://vk.com/1poetry	-43	295	699	21,93	112822	4,39	3312	32	13971
15	Cook Good - лучшие рецепты	39009769	https://vk.com/cook_good	-129	-1055	-6457	25,24	100947	1,62	3600	88	9343
16	Киномания - Лучшие фильмы	22798006	https://vk.com/kino_mania	-302	-1365	-8272	26,43	85352	0,67	14400	58	81246
17	Факты	34118551	https://vk.com/scifacts	-162	-1674	-8684	нет данных	8377	1,57	3744	44	5763
18	Лайфхак	40567146	https://vk.com/lhack	6	1016	4042	26,74	101464	1,76	5040	46	4655
19	Begin English. Английский язык для всех	12648877	https://vk.com/beginenglish_ru	-287	-1873	-11756	23,82	69175	2,11	4577	66	8050
20	Дзен	23213239	https://vk.com/dzenpub	-22	226	-2023	20,1	104535	12,22	4320	75	23115
21	Трясни Нормальность	23433159	https://vk.com/trahninormalnost1	115	1169	4124	26,79	104848	2,68	2880	44	12244
22	0% жирности	35486195	https://vk.com/zerofat	-244	-2276	-12449	21,3	111490	2,53	2160	62	7021
23	КФ	75149440	https://vk.com/fact	-120	-975	-5446	24,98	54910	5,01	4284	65	5093
24	AUTO	23783750	https://vk.com/pubauto	-38	-233	-1445	23,22	83000	2,36	2880	65	4878
25	МАМА™	20249656	https://vk.com/love_mama	-46	-313	-2646	22,61	125393	2,2	2434	0	3637
26	Арт Бот	147720339	https://vk.com/artibot	-141	-1048	-6197	27,29	13007	9,59	5184	51	3747
27	ЯЖЕБАТЯ	158484774	https://vk.com/imfather	-90	-655	-3227	16,03	20501	4,7	5760	65	8387
28	Я хочу...	36008740	https://vk.com/i_want_love_dream	-265	-1054	-4938	26,24	70211	2,37	5040	66	6439
29	КБ	67580761	https://vk.com/countryballs_re	-20	-153	-3274	15,81	43163	61,67	5760	0	109276
30	Опасная Земля	150802579	https://vk.com/world_danger	-116	-308	-3275	23,87	8303	4,21	3600	60	3339
31	Интересная планета - путешествия, туризм	14897324	https://vk.com/interestplanet_ru	-68	-762	-4329	22,65	69722	6,5	3051	103	3839
32	Бумажный самолётик	52537634	https://vk.com/paper.comics	-45	-471	-778	19,34	51648	13,24	5040	64	17836
33	Combo Vine	78996568	https://vk.com/combovine	-32	-503	-334143	27,08	31684	3,94	3540	93	20728

Figure 1.2: Structure of the file with statistics from «AllSocial» service. Data set 1

[5] All the presented sources, including their operability, are up to date as of 01.06.2022

The «Popsters» service allows you to select communities and generate an upload with statistics on communities for the selected period. The file with summary statistics for all selected communities (service limitation - no more than 10 communities) has the following structure (see Fig. 1.3).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Сравнительная таблица												
2	Название	Подписчиков	Всего публикаций	Мне нравится	Поделиться	Комментариев	Просмотров	ER Day	ER Post	ER View	LR	TR	Период
3	Мужские мысли	3599375	6037	1560012	306625	250043	689570645	0.161	0.010	0.282	0.007	0.001	01.06.2021 - 01.06.2022
4	Без кота и жизнь не та ツ	3240495	6539	16789163	2630864	363280	549388382	1.668	0.093	3.416	0.079	0.002	01.06.2021 - 01.06.2022
5	Фабрика идей	3689080	2770	880987	427385	45832	213739151	0.100	0.013	0.606	0.009	0.000	01.06.2021 - 01.06.2022
6	Красиво сказано . . .	7018600	3435	3831040	1093445	131924	491390259	0.197	0.021	0.948	0.016	0.001	01.06.2021 - 01.06.2022
7	Хитрости жизни	3310591	5789	1291720	424675	181786	654955100	0.157	0.010	0.271	0.007	0.001	01.06.2021 - 01.06.2022
8	Моя квартира	2216635	3651	977846	467770	41379	252955222	0.183	0.018	0.557	0.012	0.001	01.06.2021 - 01.06.2022
9	Между нами девочками...	3886822	4767	889558	298348	51503	239136915	0.087	0.007	0.477	0.005	0.000	01.06.2021 - 01.06.2022
10	Корпорация Юмора ツ	3456891	1025	527780	303668	41883	171299400	0.069	0.025	0.535	0.015	0.001	01.06.2021 - 01.06.2022
11	Маникюр 2022 | Дизайн ногтей	4306133	781	794582	257897	43268	200280228	0.070	0.033	0.524	0.024	0.001	01.06.2021 - 01.06.2022
12	Just Art	1717702	6984	9544201	5456163	248671	612767572	2.426	0.127	2.446	0.080	0.002	01.06.2021 - 01.06.2022
13													
14	Дни недели												
15		Мужские мысли	Без кота и жизнь не та ツ	Фабрика идей	Красиво сказано . . .	Хитрости жизни	Моя квартира	Между нами девочками...	Корпорация Юмора ツ	Маникюр 2022 | Дизайн ногтей	Just Art		
16	Пн	13,917	14,1799	12,6476	14,3112	15,7197	13,0133	15,343	15,5342	16,2118	14,1566		
17	Вт	14,6694	14,4105	13,6202	13,4347	13,9706	14,6642	14,5558	14,8213	13,1396	14,4326		
18	Ср	13,8506	13,645	13,3977	16,3721	13,9964	13,5135	14,7402	15,6295	14,7466	14,4051		
19	Чт	13,924	14,2958	14,5441	14,2092	14,1887	14,5813	15,2358	12,9978	13,574	14,6087		
20	Пт	14,6284	14,5495	14,3933	14,4493	13,6812	15,2118	13,0126	12,1306	14,1048	14,2397		
21	Сб	14,3737	13,9069	15,1615	13,58	14,0497	14,1881	13,4545	13,2675	12,9962	13,6703		
22	Вс	14,6369	15,0125	16,2357	13,6434	14,3938	14,8278	13,6581	15,6191	15,227	14,4871		
23													
24	Время суток												
25		Мужские мысли	Без кота и жизнь не та ツ	Фабрика идей	Красиво сказано . . .	Хитрости жизни	Моя квартира	Между нами девочками...	Корпорация Юмора ツ	Маникюр 2022 | Дизайн ногтей	Just Art		
26	0:00	3,1234	2,809	1,4038	3,4072	3,1044	2,9706	0	0	5,6053	4,2105		
27	1:00	2,8336	2,7288	4,3404	2,2956	2,5602	1,9978	0	0	6,124	1,2415		
28	2:00	1,7223	2,613	9,0511	4,1978	3,0399	0	0	5,1902	3,8045	2,8717		
29	3:00	4,1189	1,7917	0	4,2211	0,1357	1,4837	0	0	2,8401	0		
30	4:00	10,4128	3,3267	9,5502	3,1323	2,6011	0	0	0	7,0868	1,9431		
31	5:00	5,8309	4,3168	0	17,1024	8,0295	1,3246	0	3,0613	1,8627	5,5474		
32	6:00	5,418	4,3515	5,1994	5,9515	6,6843	13,8265	10,0798	8,1761	4,6899	5,6759		
33	7:00	5,2501	4,1922	6,6202	2,6006	5,5279	5,628	3,2907	0	5,7353	5,9615		
34	8:00	4,3701	4,6259	7,503	5,3783	4,8259	6,1321	6,0537	6,5047	3,2963	6,0841		
35	9:00	5,1139	6,0063	4,9607	4,1894	4,5973	5,6315	4,6652	7,4791	4,3112	4,9892		
36	10:00	4,1119	3,9098	5,0407	3,5454	4,5868	5,1291	4,2499	0,2826	3,4513	6,1139		
37	11:00	3,5824	5,4714	3,7135	3,1789	3,9826	5,0632	3,8109	6,5993	3,0379	5,0541		
38	12:00	3,6178	5,0065	3,427	3,7063	4,4493	4,4181	5,4042	5,8724	2,874	5,2979		
39	13:00	4,0683	3,4375	3,0037	3,5611	4,6448	5,0736	4,1294	7,3604	4,508	4,84		
40	14:00	3,527	5,9092	3,455	2,7548	4,8456	4,5193	5,8313	7,6683	3,3308	4,4288		

Figure 1.3: File structure with summary statistics on all selected communities for the set period from «Popsters» service. Data set 2

To apply an optimization approach or machine learning methodology, it is necessary that the data be represented in the form of a matrix of feature objects. In addition, the statistical data should reflect the activity of social network users as a feedback on the content posted in the community. The data from the first set do not need to be transformed, but for the data from the second set it is necessary to develop an algorithm for forming the matrix of feature objects.

The features presented in the upload from the «AllSocial» service are additional, and the main features are selected from the «Popsters» service. To learn more about the features from the «Popsters» service, you should follow the link [87]. Using the data in Set 2, new features were introduced, such as:

- The Engagement Rate per day for a month per number of community subscribers:

$$ER_{day}\ month_j = \frac{\sum_{i=1}^{n_j}(L_i + R_i + C_i)}{Subscribers_j \times n_j} \times 100,$$

where L_i, R_i, C_i is the sum of «Like»/«Share»/«Comments» from all released posts on the i day in j-month; $i = \overline{1,n_j}$, n_j is the number of days in j-month, $Subscribers_j$ is the number of community subscribers in j-month.

- The Engagement Rate per publication per month per number of community subscribers:

$$ER_{post}\ month_j = \frac{\sum_{i=1}^{n_j}(L_i + R_i + C_i)}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

where L_i, R_i, C_i is the sum of «My Likes»/«Share»/«Comments» from all released publications on the i-th day in j-month, $i = \overline{1,n}$; n_j is the number of days in j-month, $Subscribers_j$ is the number of community subscribers in j-month; $Publication_i$ is the number of publications in i day.

- The Visibility Rate of audience per day for a month per number of community subscribers:

$$VR_{day}\ month_j = \frac{\sum_{i=1}^{n_j} V_i}{Subscribers_j \times n_j} \times 100,$$

where V_i is the sum of «views» on the i day from all released publications in the j-month; $i = \overline{1,n_j}$, n_j is the number of days in the j-month; $Subscribers_j$ is the number of community subscribers in the j-month.

- The Visibility Rate per publication per month per number of community subscribers:

$$VR_{post}\ month_j = \frac{\sum_{i=1}^{n_j} V_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

where V_i is the sum of «views» on the i day from all released publications in the j-month, $Subscribers_j$ is the number of community subscribers in the j-month; $Publication_i$ is the number of publications in the i day; $i = \overline{1,n_j}$, n_j is the number of days in the j-month.

- The Love Rate of an audience per day for a month per number of community subscribers:

$$LR_{day}\ month_j = \frac{\sum_{i=1}^{n_j} L_i}{Subscribers_j \times n_j} \times 100,$$

where L_i is the sum of «Like» on the ith day from all released publications in j-month; $i = \overline{1, n_j}$, n_j is the number of days in j-month ; $Subscribers_j$ is the number of community subscribers in j-month.

- The Love Rate per publication per month per number of community subscribers:

$$LR_{post}\, month_j = \frac{\sum_{i=1}^{n_j} L_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

where L_i is the sum of «My Likes» on the ith day from all released publications in the jth month, $Subscribers_j$ is the number of community subscribers in the jth month; $Publication_i$ is the number of publications on the ith day; $i = \overline{1, n_j}$, n_j is the number of days in the jth month.

- The Talk Rate of the audience per day for a month per number of community subscribers:

$$TR_{day}\, month_j = \frac{\sum_{i=1}^{n_j} C_i}{Subscribers_j \times n_j} \times 100,$$

where C_i is the sum of «Comments» on the ith day from all released publications in j-month; $i = \overline{1, n_j}$, n_j is the number of days in j-month; $Subscribers_j$ is the number of community subscribers in j-month.

- The Talk Rate per publication per month per number of community subscribers:

$$TR_{post}\, month_j = \frac{\sum_{i=1}^{n_j} C_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

where C_i is the sum of «Comments» on the ith day from all released publications in the jth month, $Subscribers_j$ is the number of community subscribers in the jth month; $Publication_i$ is the number of publications in the ith day; $i = \overline{1, n_j}$, n_j is the number of days in the jth month.

- The Amplification Rate among the audience per day for a month per number of community subscribers:

$$AR_{day}\ month_j = \frac{\sum_{i=1}^{n_j} R_i}{Subscribers_j \times n_j} \times 100,$$

where R_i is the sum of «Share» on the ith day from all released publications in j-month; $i = \overline{1,n_j}$, n_j is the number of days in j-month; $Subscribers_j$ is the number of community subscribers in j-month.

- The Amplification Rate to the audience per publication per month per number of community subscribers:

$$AR_{post}\ month_j = \frac{\sum_{i=1}^{n_j} R_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

where R_i is the sum of «Share» on the ith day from all released publications in the jth month, $Subscribers_j$ is the number of community subscribers in the jth month; $Publication_i$ is the number of publications on the ith day; $i = \overline{1,n_j}$, n_j is the number of days in the jth month.

- Total publications during the month:

$$Publication_j = \sum_{i=1}^{n_j} Publication_i,$$

where $Publication_j$ is the number of publications in j-month, $j = \overline{1,m}$; m is the number of months, $i = \overline{1,n_j}$, n_j is the number of days in j-month.

- Intensity of publication activity per day during the month (hereinafter «IPA per day (for the year)»):

$$IPA_{month_j} = \frac{\sum_{i=1}^{n_j} Publication_i}{n_j},$$

where $\sum_{i=1}^{n_j} Publication_i$ is the number of publications in j-month, $j = \overline{1,m}$; m is the number of months, $i = \overline{1,n_j}$, n_j is the number of days in j-month.

- Average value of marks («Like», «Share», «Comments», «Views») of a publication during a month:

$$L_{post}\ month_j = \frac{\sum_{i=1}^{n_j} L_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

$$R_{post}\ month_j = \frac{\sum_{i=1}^{n_j} R_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

$$C_{post}\ month_j = \frac{\sum_{i=1}^{n_j} C_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

$$V_{post}\ month_j = \frac{\sum_{i=1}^{n_j} V_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

  where $\sum_{i=1}^{n_j} L_i, \sum_{i=1}^{n_j} R_i, \sum_{i=1}^{n_j} C_i, \sum_{i=1}^{n_j} V_i$ is the sum of «Like», «Share», «Comments», «Views» under all publications for j-month; $\sum_{i=1}^{n_j} Publication\_day_i$ - number of publications in i-th day, $j = \overline{1,m}$; m - number of months, $i = \overline{1,n_j}$, n_j - number of days in j-th month.

- Average value of marks («Like», «Share», «Comments», «Views») of the publication during the year:

$$L_{post}\ year = \frac{\sum_{j=1}^{m} L_j}{\sum_{j=1}^{m} Publication\_month_j},$$

$$R_{post}\ year = \frac{\sum_{j=1}^{m} R_j}{\sum_{j=1}^{m} Publication\_month_j},$$

$$C_{post}\ year = \frac{\sum_{j=1}^{m} C_j}{\sum_{j=1}^{m} Publication\_month_j},$$

$$V_{post}\ year = \frac{\sum_{j=1}^{m} V_j}{\sum_{j=1}^{m} Publication\_month_j},$$

  where $\sum_{j=1}^{m} L_j, \sum_{j=1}^{m} R_j, \sum_{j=1}^{m} C_j, \sum_{j=1}^{m} V_j$ - sum of «Like», «Share», «Comments», «Views» marks under all publications for j-month; $Publication\_month_j$ - number of publications in month j, $j = \overline{1,m}$, m - number of months.

Using the data in set 1, new features were introduced, such as:

- Average age of the target audience of the community (hereinafter referred to as «Average age of TA»):

$$Age_{mean} = \frac{\sum_{i=1}^{n} Age\_interval\_mean_i \times Percent\_value\_interval_i}{100},$$

where $Age\_interval\_mean_i$ - average age in i-interval, for example, for «% 27-30 years» the average age will be 28,5 years (the extreme values: «% under 18 years» and «% from 45 years», were taken as follows: 16.5 and 50 years old, respectively); $Percent\_value\_interval_i$ - what percentage of the total audience belongs to the given age range.

- Ratio of men and women in the community (hereinafter «Gender (M/W)»):

$$Sex\ (male/female) = \frac{Percent\_male}{Percent\_female},$$

where $Percent\_male$ / $Percent\_female$ - audience share of men and women respectively.

Some of the features introduced above will be used to implement clustering algorithms, some will be used to form recommendations for content placement and for complex display of activity within the community.

In addition, a characteristic - «Thematic» was introduced, which was filled in manually, as there is no such characteristic in the selected services.

As a result, from the «AllSocial» service we have the following characteristics of communities: «Communities»$>$, «id», «URL», «Growth in 1 day», «Growth in 7 days», «Growth in 30 days», «% Offline more than a month», «ER (100 last posts)», «Sociate price», «CPM Sociate», «Customers (average for 7 days)», «% Mobile», «% Computer», «% Male», «% Female», «% under 18 years», «% 18-21 years», «% 21-24 years», «% 24-27 years», «% 27-30 years», «% 30-35 years», «% 35-45 years», «% from 45 years», «$Age_{mean}>$, «$Sex\ (male/female)$», «% Russia», «% Belarus», «% Other countries».

From the «Popsters» service we get the following characteristics: «Days of week», «Time of day», «Days of week/text volume», «Time of day/text volume», «Relative activity days of week/text volume», «Relative activity time of day/text volume», «Like», «Share», «Comments»>, «Views», «Subscribers», «Number of publications», «Number by text length», «ER by text length», «Number by content type», «ER by content type», «Relative activity by content type», «$ER_{day}\ month_j$», «$ER_{post}\ month_j$», «$VR_{day}\ month_j$», «$VR_{post}\ month_j$», «$LR_{day}\ month_j$», «$LR_{post}\ month_j$», «$LR_{post}\ month_j$», «$TR_{post}\ month_j$», «$AR_{day}\ month_j$», «$AR_{post}\ month_j$», «$Publication_j$», «IPA_{month_j}», «$L_{post}\ month_j$», «$R_{post}\ month_j$», «$C_{post}\ month_j$», «$V_{post}\ month_j$», «$L_{post}\ year$», «$R_{post}\ year$», «$C_{post}\ year$», «$V_{post}\ year$».

For modeling and implementation of machine learning methods, it is required to build a feature matrix. In accordance with this requirement, a data preprocessing algorithm was developed. Let us consider it in more detail.

Algorithm for forming a feature matrix

It is necessary to transform the data set 2 into a feature matrix, where the object is the community and the features are its characteristics. To realize the required transformation, we developed an algorithm (see Fig. 1.4) consisting of the following parts:

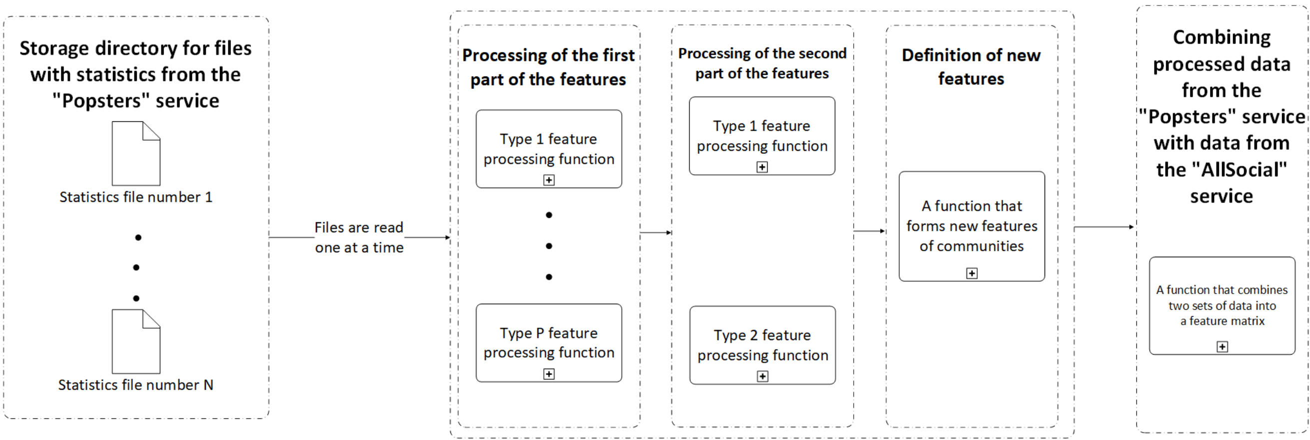


Figure 1.4: Data preprocessing algorithm scheme

1. A cycle is implemented in which files are read in turn and their preprocessing begins according to the principle - one file is divided into tables, the number of which is equal to the number of attributes. It should be noted that the attributes are united by type, for which the processing functions will be the same.

1.1 A nested loop that removes empty rows and puts in the table feature names;

1.2 Further functions of processing of all types of attributes and functions of checking of identical names of communities in one file, and also in those cases when the same community is in different files are realized:

- If such communities exist, their line number is added to the community name, for example: «My dacha», «My dacha_6», «My dacha_7». This is necessary for correct operation of the merge and concat methods of the pandas library;
- If the same community exists in different files, the function of checking and deleting such repetitions is implemented for this case, and all but the first one are deleted.

1.3 Thus, matrices are formed according to the number of traits, each of which stores data for all communities.

2. A function is implemented that defines and calculates new features using the formulas described earlier from existing data.

3. Functions for processing two types of tables containing object attributes and forming a feature matrix are implemented.

4. Next, the transformed data is merged with dataset 1 to form a single feature matrix.

As a result of the algorithm's work, a feature matrix with 3604 communities, more than 20 target features and more than 30 characteristics is created to form recommendations for placing records in each community (see Fig. 1.5). The software component implementing this algorithm is programmatically implemented and registered in the FIPS [79].

1		0	1	2	3	4	5	6	7	8	9	10	11	12	13	
2	0	Сообщества	Подписчиков	Всего публикаций	Мне нравится	Поделиться	Комментариев	Просмотров	ER Day_year	ER Post_year	ER View_year	LR Post_year	TR Post_year	Период	VR Post_year	A
3	1	MDK	11822715	14585	37259544	27224948	1565814	4800829108	1.526	0.038	1.296	0.022	0.001	01.06.2021 - 01.06.2022	2,784	
4	2	4ch	5071206	2253	13336557	3486182	305833	763120438	0.923	0.150	2.277	0.117	0.003	01.06.2021 - 01.06.2022	6,679	
5	3	NR.Music	5291361	8814	9844436	4496675	655541	1569389740	0.774	0.032	0.869	0.021	0.001	01.06.2021 - 01.06.2022	3,365	
6	4	БОРЩ	7057083	9608	22405135	15947605	919675	2185478037	1.520	0.058	1.768	0.033	0.001	01.06.2021 - 01.06.2022	3,223	
7	5	Академия Порядочных Парней	5393647	5216	4622969	2922419	611221	1200581400	0.413	0.029	0.689	0.016	0.002	01.06.2021 - 01.06.2022	4,267	
8	6	Science|Наука	5230844	5557	8935943	6435616	2333322	2221480677	0.925	0.061	0.737	0.031	0.008	01.06.2021 - 01.06.2022	7,642	
9	7	ПОЗОР	4557846	13979	43874125	4751586	88569	2396305273	3.686	0.076	2.100	0.069	0.000	01.06.2021 - 01.06.2022	3,761	
10	8	Лепрозорий	3481261	1775	4228431	4031156	1047800	882556824	0.730	0.151	1.118	0.068	0.017	01.06.2021 - 01.06.2022	14,283	
11	9	MARVEL/DC	3703678	6352	40593633	4880416	919600	1566532984	3.423	0.197	2.801	0.173	0.004	01.06.2021 - 01.06.2022	6,659	
12	10	Идеи дизайна интерьера	5831257	3389	2630301	1324220	166410	618186253	0.193	0.021	0.644	0.013	0.001	01.06.2021 - 01.06.2022	3,128	
13	11	Книги	2159267	8980	5804354	1773459	316706	533217485	0.999	0.041	1.341	0.030	0.002	01.06.2021 - 01.06.2022	2,75	
14	12	Цитаты и статусы	3585503	4400	666863	230101	34196	162473930	0.071	0.006	0.492	0.004	0.000	01.06.2021 - 01.06.2022	1,03	
15	13	В приколе (18+)	5258065	12095	3414599	2540972	355613	669237223	0.328	0.010	0.976	0.005	0.001	01.06.2021 - 01.06.2022	1,052	
16	14	• Неприличные Анекдоты	3688681	2677	1471915	911649	35834	186808899	0.179	0.025	1.154	0.015	0.000	01.06.2021 - 01.06.2022	1,892	
17	15	Экспериментатор | Наука	1641143	8471	2638863	726776	270150	435153324	0.605	0.026	0.883	0.019	0.002	01.06.2021 - 01.06.2022	3,13	
18	16	Кулинарное искусство	3115565	1829	1511781	779355	25462	136355640	0.203	0.041	2.019	0.027	0.000	01.06.2021 - 01.06.2022	2,393	
19	17	Простые рецепты	3251767	2385	1454300	1255197	19068	193581958	0.229	0.035	1.235	0.019	0.000	01.06.2021 - 01.06.2022	2,496	
20	18	КАЕФ	1599307	9554	2309271	1166501	32319	233995670	0.599	0.023	1.420	0.015	0.000	01.06.2021 - 01.06.2022	1,531	
21	19	Боги Смеха	1507737	1502	3595186	254900	50695	66028494	0.707	0.172	5.860	0.159	0.002	01.06.2021 - 01.06.2022	2,916	
22	20	PINK PARADISE	445362	3282	1143357	1036158	11314	72786601	1.344	0.150	3.005	0.078	0.001	01.06.2021 - 01.06.2022	4,98	
23	21	Деревня РУКОДЕЛКИНО рукоделие, вязание, дизайн	588480	3440	352518	239732	5768	52215744	0.278	0.030	1.020	0.017	0.000	01.06.2021 - 01.06.2022	2,579	
24	22	Сама Себе Тренер	473304	4076	685022	589163	9599	62371222	0.741	0.067	1.971	0.036	0.000	01.06.2021 - 01.06.2022	3,233	
25	23	Питер Live	303433	5417	722949	161466	40833	61124528	0.833	0.056	1.400	0.044	0.002	01.06.2021 - 01.06.2022	3,719	
26	24	Знать и познавать	421557	996	821587	329895	24884	42479634	0.762	0.280	2.471	0.196	0.006	01.06.2021 - 01.06.2022	10,117	
27	25	Queen	644965	6173	2376196	557917	9868	115287131	1.247	0.074	2.083	0.060	0.000	01.06.2021 - 01.06.2022	2,896	
28	26	Новинки Музыки 2022 | Свежая Музыка	345278	657	71387	27570	13755	8604158	0.089	0.050	1.226	0.031	0.006	01.06.2021 - 01.06.2022	3,793	

Figure 1.5: Feature matrix

In this paragraph, the most popular services for providing statistical data from the «VKontakte» social network were considered, the characteristics of communities that will be used in optimization and clustering tasks were determined, and an algorithm that forms a feature matrix was developed.

Algorithm of statistical data processing

Data processing algorithm, is a series of transformations with a feature matrix, such as:

- Deletion of objects that have an empty «Sociate price» column. This is explained by the necessity to use for modeling objects that have a known cost of advertising record placement;
- Deletion of objects whose value in the «Total publications» column is equal to 0. This is explained by the need to use for modeling a set of objects with non-zero activity during the time period: 01.06.2021 - 31.05.2022;
- Convert categorical attributes with string values, such as «Topic», «Community Name», «URL», to a numeric data type.

In addition, this block uses data from the client's terms of reference, such as: «Topic», «Target Indicator», «Month». Based on this, the following is formed: a subset of objects, specified topics, with non-zero values of target indicators in the month specified by the client. If there are several months, then for each month it is necessary to perform modeling separately, because the subsets of objects with non-zero values of the target indicator, in general, may differ from month to month. These transformations are explained by the necessity to allocate a certain subset of objects for a specific technical task of the client in order to perform numerical modeling and form a solution.

The output is two sets of data, one designed to apply optimization and machine learning methods, the other to generate recommendations for advertising record placement for each object. In the first case, the set includes the attributes: «Subscribers», «ER Post_year», «ER View_year», «LR Post_year», «TR Post_year», «VR Post_year», «AR Post_year», «Likes post average (per year)», «Reposts post average (per year)», «Comments post average (per year)», «Views post average (per year)», «Visitors (average for 7 days)», «IPA per day (for a year)», «% Offline more than a month», «Price of Sociate», «CPM of Sociate», «Gender (M/W)», «Topic_id», «ER (100 last posts)», «Average age of TA», «Growth for 30 days». Note that «Like» and «Reposts» are understood as «My Likes» and «Share», respectively.

1.2.3. Peculiarities of optimization model implementation

Implementation of the recommendation block

Support of managerial decision-making for the DM in the task of information dissemination in the MCM is carried out by analyzing the behavioral activity of the audience of the sites. Consequently, in addition to the characteristics introduced in Section 1.2.2., it is necessary to develop functions for making recommendations for posting records in social network communities.

More than 10 functions have been written to realize the block of recommendations formation (see Fig. 1.6).

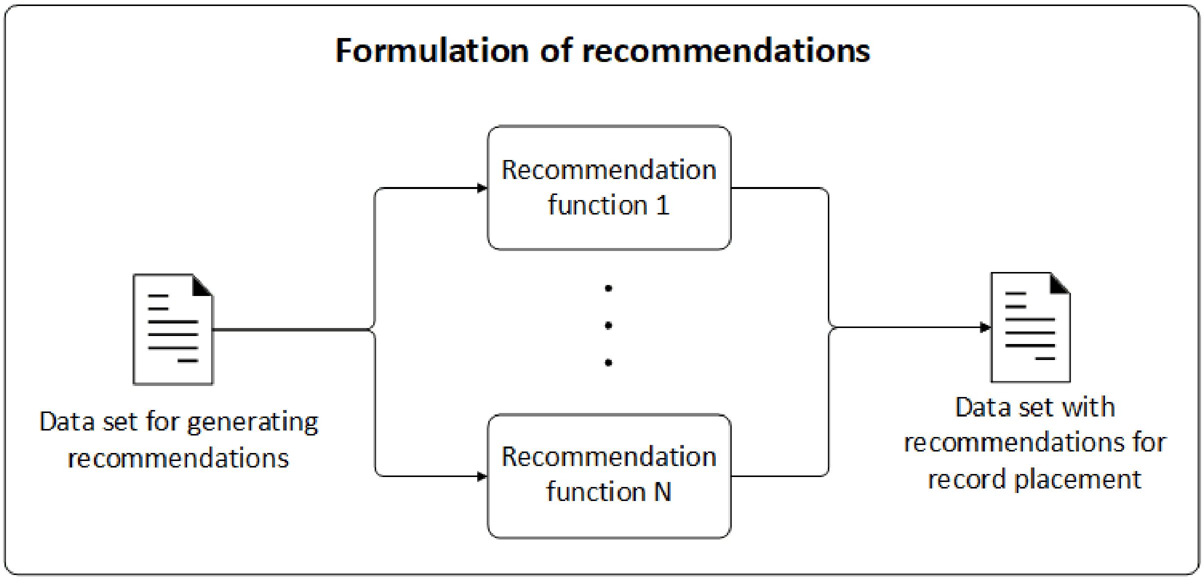


Figure 1.6: Scheme of implementation of the developed library for forming recommendations on publication placement

Thanks to the formation of recommendations, it is possible to determine the best time period for publishing a record, the amount of text in it, the need to attach photo/audio/video materials to it, as well as links, in such a way that the audience response to the posted record was maximized in a given community, i.e. to make it so that the published record has more «Like», «Share», «Comments», «Views» (see fig. 1.7).

```python
def best_dayweek_func(data: pd.DataFrame) -> pd.DataFrame:
    """..."""
    data_copied = data.copy()
    list = ['Дни недели Пн', 'Дни недели Вт', 'Дни недели Ср', 'Дни недели Чт', 'Дни недели Пт', 'Дни недели Сб',
            'Дни недели Вс']
    data_copied = data_copied[list]
    best_res = pd.DataFrame(
        [data_copied.columns[i].tolist() for i in (data_copied.values == data_copied.max(axis=1)[:, None])])
    best_res.rename(columns={0: 'День недели для публикации', 1: 'Значение (День недели)'}, inplace=True)
    best_res['Значение (День недели)'] = data_copied.max(axis=1)
    best_res = best_res.replace(list,
                                ['Понедельник', 'Вторник', 'Среда', 'Четверг', 'Пятница', 'Суббота', 'Воскресенье'])
    return best_res
```

Figure 1.7: Example of a recommendation function

In addition to the basic recommendations, the client is provided with all available metrics reflecting the activity in the community, for example, not «Average age of TA», but the percentage of the number of subscribers by age intervals.

Implementation of the optimization model

For software implementation of the optimization model it is necessary to formulate a mathematical statement of the problem, form a matrix of objects-attributes and functions of recommendations for each object-community on record placement. The set tasks will be solved with the help of known methods of linear programming, methods of multicriteria optimization, and programmatically implemented algorithms for solving problems will be in the programming language «Python» with appropriate libraries in a cross-platform integrated development environment PyCharm. Due to the optimization of the running time of algorithms solving the set tasks, it was decided to implement one function, which solves both the problem of multi-criteria optimization and the problem of integer linear programming.

Let us proceed to the description of the architecture of the function implementing the proposed optimization approach to solving the problem of information dissemination in the MCM:

1. It is set: lists of topics, month of the year, client's budget - P, list of criteria: «Like», «Share», «Comments», «Views» - f_1, f_2, f_3, f_4, matrix of weight coefficients in the form of a list with nested lists - A, table for fixing the results of modeling. It should be noted that the client's budget does not exceed the maximum budget for the given topics.

2. 3 nested cycles are defined: the first by the list of topics, the second by the client's budget, and the third by the timing of the weighting matrix.

3. The function of data processing is called, which selects objects from the given list of topics with non-zero values of criteria in the given month of the year. Thus, a subset of objects on which modeling will be performed is formed.

4. The function that implements optimization methods is called, where recommendations are formed for objects of a given subset, selecting objects where the cost of advertising record placement is less than or equal to P.

5. We specify: an array of prices for advertising record placement - g_j, where $j = \overline{1,m}$; arrays of values for criteria f_1, f_2, f_3, f_4.

6. A single criterion is specified using the criterion convolution method:

$$\hat{f}(x) = \alpha_1 \cdot f_1(x) + \alpha_2 \cdot f_2(x) + \alpha_3 \cdot f_3(x) + \alpha_4 \cdot f_4(x) =$$
$$= \alpha_1 \cdot \sum_{j=1}^{m} a_j \cdot x_j + \alpha_2 \cdot \sum_{j=1}^{m} b_j \cdot x_j + \alpha_3 \cdot \sum_{j=1}^{m} c_j \cdot x_j + \alpha_4 \cdot \sum_{j=1}^{m} d_j \cdot x_j =$$
$$= \sum_{j=1}^{m} (\alpha_1 \cdot a_j + \alpha_2 \cdot b_j + \alpha_3 \cdot c_j + \alpha_4 \cdot d_j) \cdot x_j = \sum_{j=1}^{m} w_j \cdot x_j$$

7. Then, using the «scipy.optimize» [12] library, the optimization problem is solved and the resulting vector x is determined.

8. The values of the corresponding criteria for the optimal sets of communities at given significance coefficients are calculated.

9. The simulation results are recorded in comparison tables (see Figure 1.8 for an example).

20 rows 20 rows × 16 columns CSV

	alpha_0	alpha_1	alpha_2	alpha_3	Всего Лайки поста Июль	Всего Репосты поста Июль	Всего Комментарии поста Июль	Всего Просмотры поста Июль	F_value	Obj_fun	Бюджет клиента	Общая стоимость	Количество сообществ	Выбранные тематики	Время работы	Месяц
1	1.00	0.00	0.00	0.00	4346	469	140	134267	139222	-4346.00	1000	995	8	Образование	0 days 00:00:00.003001	Июль
2	0.00	1.00	0.00	0.00	2332	1455	45	144614	148446	-1455.00	1000	974	6	Образование	0 days 00:00:00.003001	Июль
3	0.00	0.00	1.00	0.00	3782	453	168	131835	136238	-168.00	1000	995	7	Образование	0 days 00:00:00.003001	Июль
4	0.00	0.00	0.00	1.00	3486	664	110	168863	173123	-168863.00	1000	989	8	Образование	0 days 00:00:00.015004	Июль
5	0.25	0.25	0.25	0.25	3486	664	110	168863	173123	-43280.75	1000	989	8	Образование	0 days 00:00:00.011002	Июль
6	0.97	0.01	0.01	0.01	4249	556	116	148354	153275	-5611.79	1000	995	8	Образование	0 days 00:00:00.007002	Июль
7	0.01	0.97	0.01	0.01	2763	1320	62	162099	166244	-2929.64	1000	988	7	Образование	0 days 00:00:00.001999	Июль
8	0.01	0.01	0.97	0.01	3523	618	121	167930	172192	-1838.08	1000	997	9	Образование	0 days 00:00:00.012002	Июль
9	0.01	0.01	0.01	0.97	3486	664	110	168863	173123	-163839.71	1000	989	8	Образование	0 days 00:00:00.013003	Июль
10	0.94	0.02	0.02	0.02	3960	646	112	167236	171954	-7082.28	1000	974	9	Образование	0 days 00:00:00.005002	Июль
11	0.02	0.94	0.02	0.02	2763	1320	62	162099	166244	-4539.28	1000	988	7	Образование	0 days 00:00:00.002001	Июль
12	0.02	0.02	0.94	0.02	3486	664	110	168863	173123	-3563.66	1000	989	8	Образование	0 days 00:00:00.011986	Июль
13	0.02	0.02	0.02	0.94	3486	664	110	168863	173123	-158816.42	1000	989	8	Образование	0 days 00:00:00.015004	Июль
14	0.91	0.03	0.03	0.03	3960	646	112	167236	171954	-8643.42	1000	974	9	Образование	0 days 00:00:00.000999	Июль
15	0.03	0.91	0.03	0.03	2763	1320	62	162099	166244	-6148.92	1000	988	7	Образование	0 days 00:00:00.003001	Июль
16	0.03	0.03	0.91	0.03	3486	664	110	168863	173123	-5290.49	1000	989	8	Образование	0 days 00:00:00.014003	Июль
17	0.03	0.03	0.03	0.91	3486	664	110	168863	173123	-153793.13	1000	989	8	Образование	0 days 00:00:00.014003	Июль
18	0.88	0.04	0.04	0.04	3960	646	112	167236	171954	-10204.56	1000	974	9	Образование	0 days 00:00:00.000999	Июль
19	0.04	0.88	0.04	0.04	2763	1320	62	162099	166244	-7758.56	1000	988	7	Образование	0 days 00:00:00.007001	Июль
20	0.04	0.04	0.88	0.04	3486	664	110	168863	173123	-7017.32	1000	989	8	Образование	0 days 00:00:00.012002	Июль

Figure 1.8: Example of a comparison table of modeling results

1.3. Numerical modeling and analysis of results

To demonstrate the operation of the proposed approach and the possibility of comparative analysis of the obtained results, we set the following input parameters:

1. Topics - «Automobiles, car owners», «Culinary, recipes», «Education». Number of objects in each theme: 124, 126, 103;

2. The time intervals are monthly, January through December;

3. Budget - from 1000 rubles to the maximum possible budget within the given topics and months of the year with a step of 5000 rubles;

4. Customer preferences are given in vector form $\alpha = \{\alpha 1, \alpha 2, \alpha 3, \alpha 4\}$ and are represented in table 1.1:

Table 1.1: Weighting coefficients for conducting sensitivity analysis for changes in customer preferences

Part 1				**Part 2**				**Part 3**			
$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$
1.0	0.0	0.0	0.0	0.05	0.85	0.05	0.05	0.1	0.1	0.1	0.7
0.0	1.0	0.0	0.0	0.05	0.05	0.85	0.05	0.67	0.11	0.11	0.11
0.0	0.0	1.0	0.0	0.05	0.05	0.05	0.85	0.11	0.67	0.11	0.11
0.0	0.0	0.0	1.0	0.82	0.06	0.06	0.06	0.11	0.11	0.67	0.11
0.25	0.25	0.25	0.25	0.06	0.82	0.06	0.06	0.11	0.11	0.11	0.67
0.97	0.01	0.01	0.01	0.06	0.06	0.82	0.06	0.64	0.12	0.12	0.12
0.01	0.97	0.01	0.01	0.06	0.82	0.06	0.82	0.12	0.64	0.12	0.12
0.01	0.01	0.97	0.01	0.79	0.07	0.07	0.07	0.12	0.12	0.64	0.12
0.01	0.01	0.01	0.97	0.07	0.79	0.07	0.07	0.12	0.12	0.12	0.64
0.94	0.02	0.02	0.02	0.07	0.07	0.79	0.07	0.61	0.13	0.13	0.13
0.02	0.94	0.02	0.02	0.07	0.07	0.07	0.79	0.13	0.61	0.13	0.13
0.02	0.02	0.94	0.02	0.76	0.08	0.08	0.08	0.13	0.13	0.61	0.13
0.02	0.02	0.02	0.94	0.08	0.76	0.08	0.08	0.13	0.13	0.13	0.61

Continuation of table 1.1

0.91	0.03	0.03	0.03	0.08	0.08	0.76	0.08	0.58	0.14	0.14	0.14
0.03	0.91	0.03	0.03	0.08	0.08	0.08	0.76	0.14	0.58	0.14	0.14
0.03	0.03	0.91	0.03	0.73	0.09	0.09	0.09	0.14	0.14	0.58	0.14
0.03	0.03	0.03	0.91	0.09	0.73	0.09	0.09	0.14	0.14	0.14	0.58
0.88	0.04	0.04	0.04	0.09	0.09	0.73	0.09	0.55	0.15	0.15	0.15
0.04	0.88	0.04	0.04	0.09	0.09	0.09	0.73	0.15	0.55	0.15	0.15
0.04	0.04	0.88	0.04	0.7	0.1	0.1	0.1	0.15	0.15	0.55	0.15
0.04	0.04	0.04	0.88	0.1	0.7	0.1	0.1	0.15	0.15	0.15	0.55
0.85	0.05	0.05	0.05	0.1	0.1	0.7	0.1	–	–	–	–

Such a number of variants of the vectors of weight coefficients is due to the fact that a sensitivity analysis of the given system will be carried out in order to form recommendations for decision makers depending on the budget, seasonality and nomenclature of goods or services. Note that to assess the impact of changing the values of criteria weights on the final result, the principle of selecting the main criterion and further even distribution of the balance among the others is used. Let's move on to analyzing the simulation results.

Observation 1.1. *The number of unique solutions does not exceed 30 percent regardless of changes in the values of weighting coefficients and budget for the specified topics and months of the year (see fig. 1.9, 1.10, 1.11, 1.12).* The table in Figure 1.9 shows the maximum values of the percentage of unique solutions among all possible budgets.

Topics	January	February	March	April	May	June	July	August	September	October	November	December
Culinary, recipes	25	26,7	28,3	26,7	21,7	25	26,7	28,3	30	26,7	26,7	28,3
Automobiles,car owners	16,7	16,7	20	25	18,3	21,7	21,7	15	15	16,7	16,7	16,7
Education	23,3	23,3	23,3	23,3	26,7	18,3	25	23,3	30	25	23,3	23,3

Figure 1.9: Maximum percentage of unique solutions among all budgets

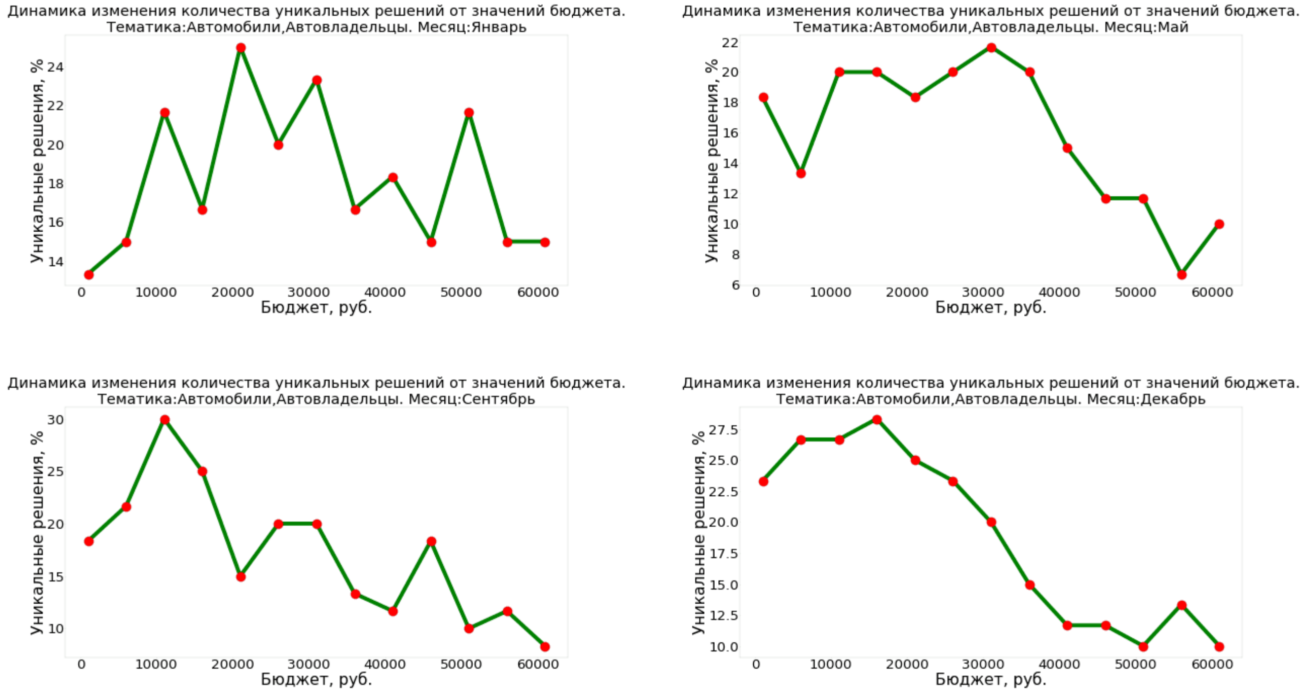


Figure 1.10: Dynamics of change in the percentage of unique solutions depending on the budget for the topic: «Automobiles, Car Owners»

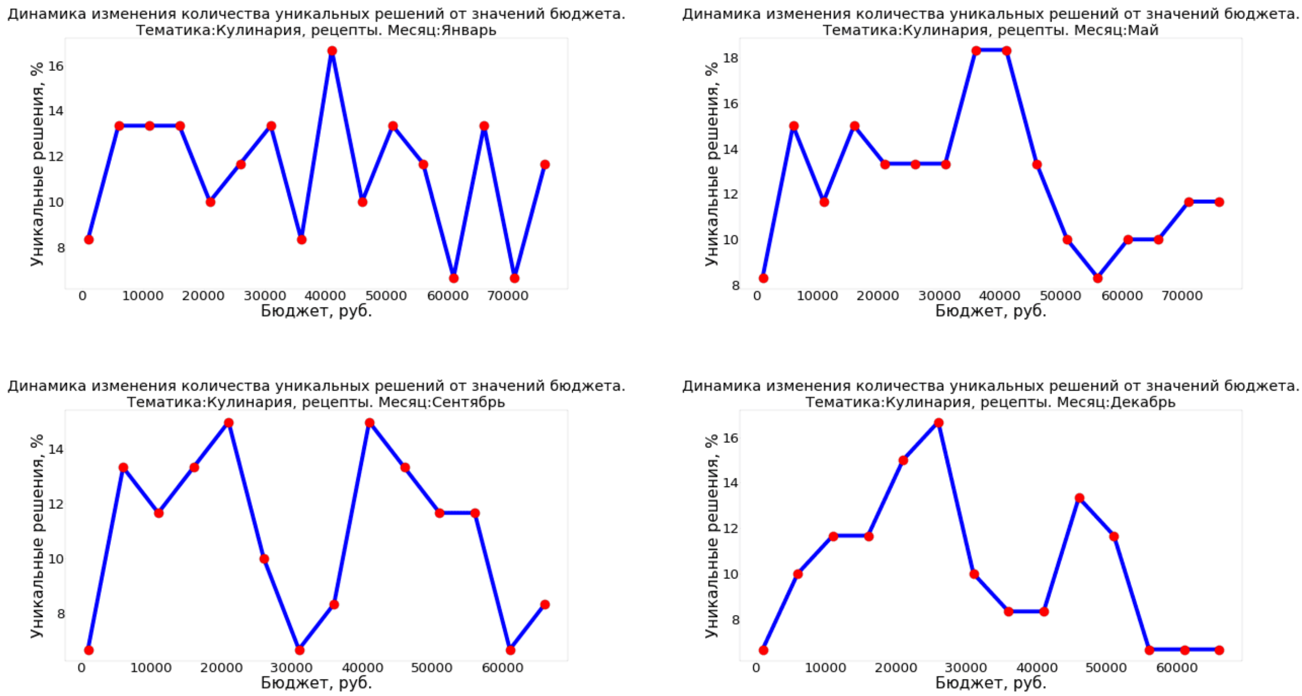


Figure 1.11: Dynamics of change in the percentage of unique solutions depending on the budget for the theme: «Culinary, recipes»

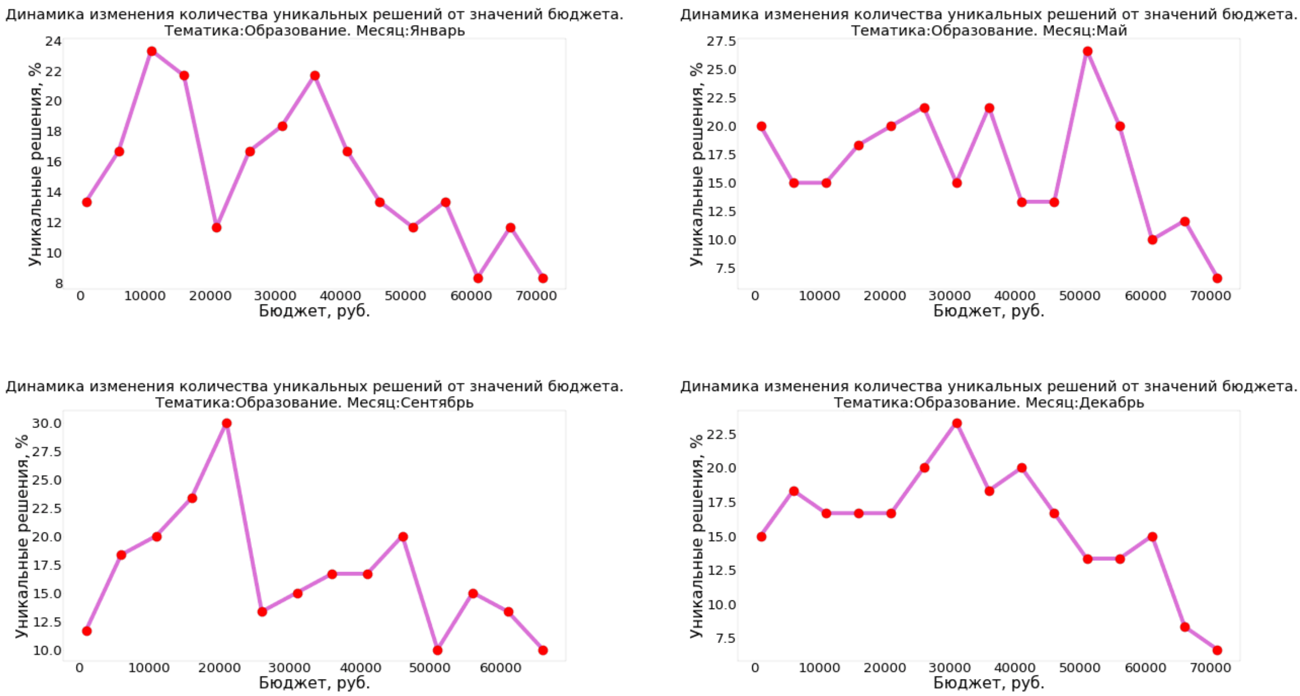


Figure 1.12: Dynamics of change in the percentage of unique solutions depending on the budget for the subject: «Education»

Conclusion 1.1. *The specifics of behavioral activity of social network users are such that changing preferences by a third affects the process of forming a unique solution.*

Remark 1.1. *The dynamics of changes in the percentage of unique decisions from changes in the values of weighting coefficients by topics and months of the year shows how active and different from each other the audience behaves in different groups of the same topic, regardless of the selected criteria.*

Indeed, on the presented graph (see Fig. 1.13) we can see that in different time intervals as a result of the increased interest of network users in certain areas of their life, due to the seasonality of demand for goods (or services) and the manifestation of other external factors, the probability of obtaining a unique set of sites increases. This remark allows us to draw the following conclusion to substantiate the recommendations on the formation of preferences.

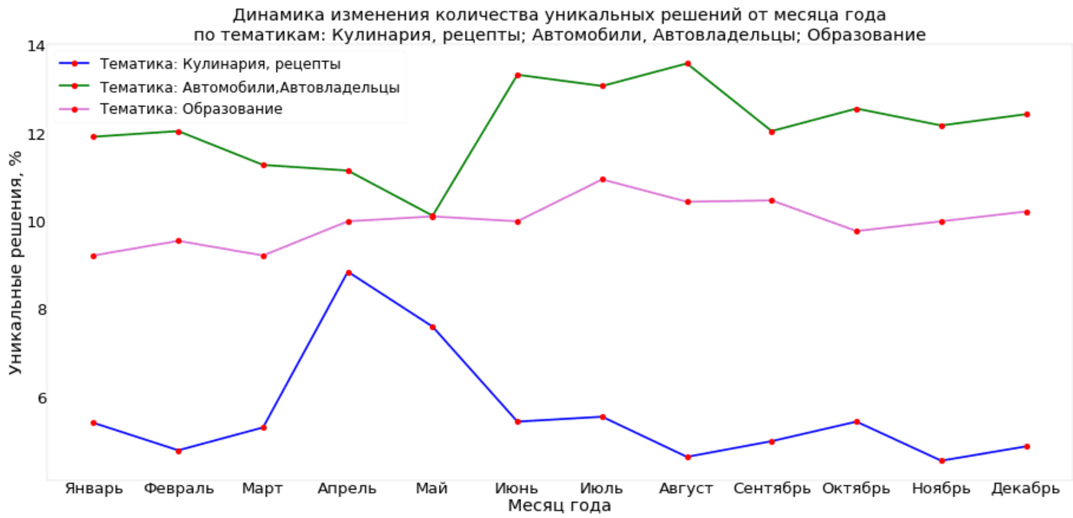


Figure 1.13: Dynamics of change in the number of unique solutions by month of the year by topic

Conclusion 1.2. *The more diverse the community audience, the higher the percentage of unique results with small changes in preferences.*

Remark 1.2. *Based on the data for 12 months, 3 themes and set budget values, we notice that there is a differentiation in the sensitivity of the criteria.*

This remark indicates that it is possible to adjust the preferences of the client depending on the input parameters set by him, as well as to analyze the degree of activity of the audience of sites.

Conclusion 1.3. *The current market trends are such that different criteria have different sensitivity to changes in preferences (weighting values) regardless of topics, budget and month of the year (see fig. 1.14, 1.15, 1.16, 1.17).*

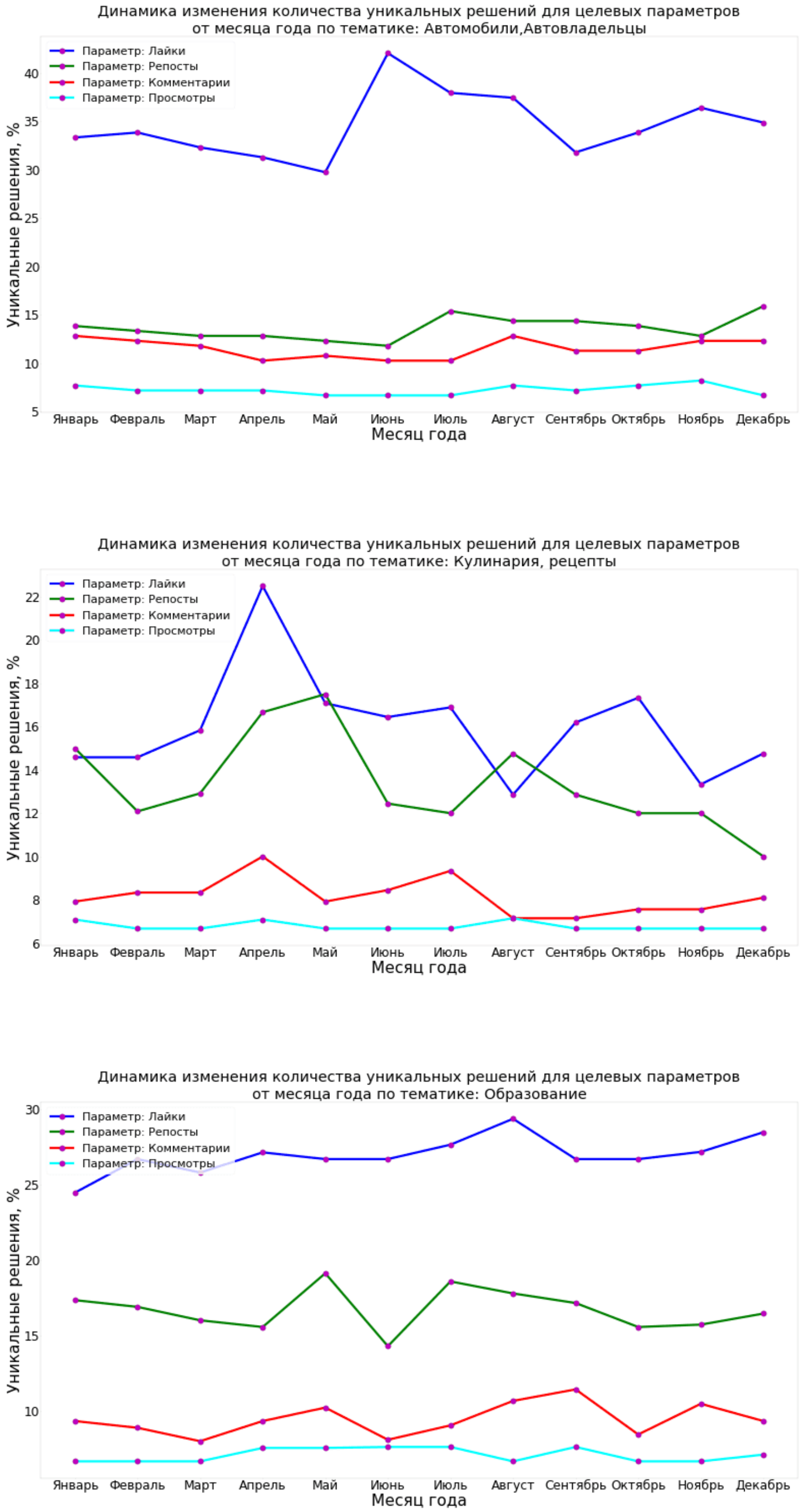


Figure 1.14: Dynamics of the percentage of unique solutions by themes and months of the year depending on the criterion

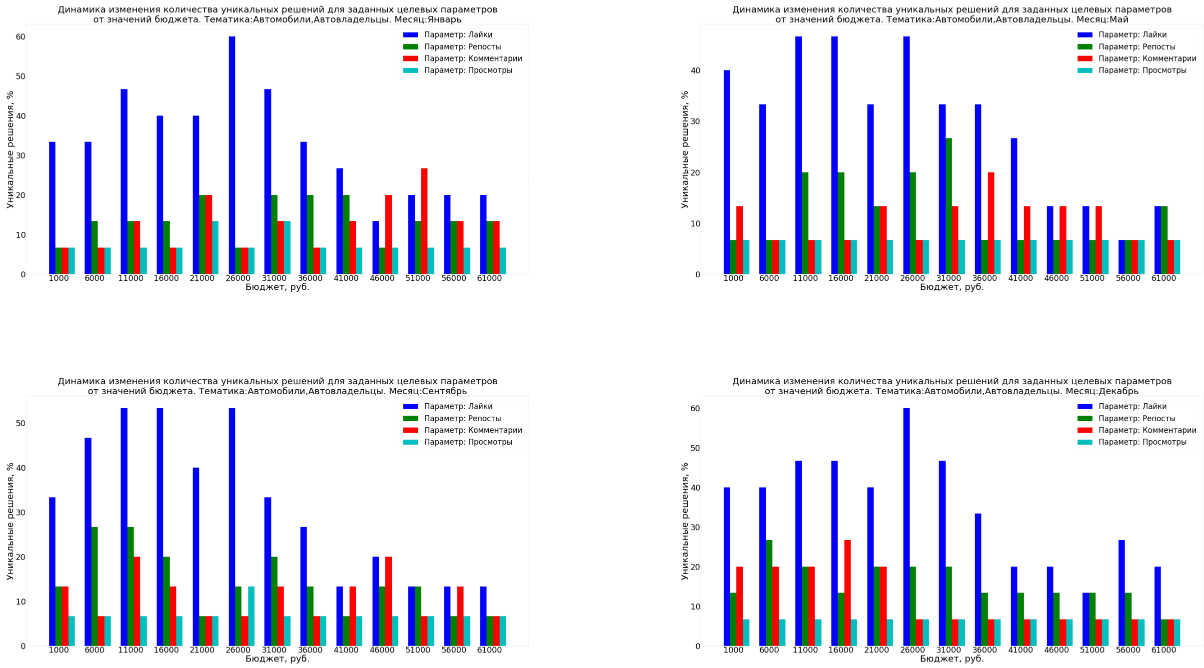


Figure 1.15: Dynamics of change in the number of unique solutions by criteria depending on the budget. Subject: «Automobiles, car owners»

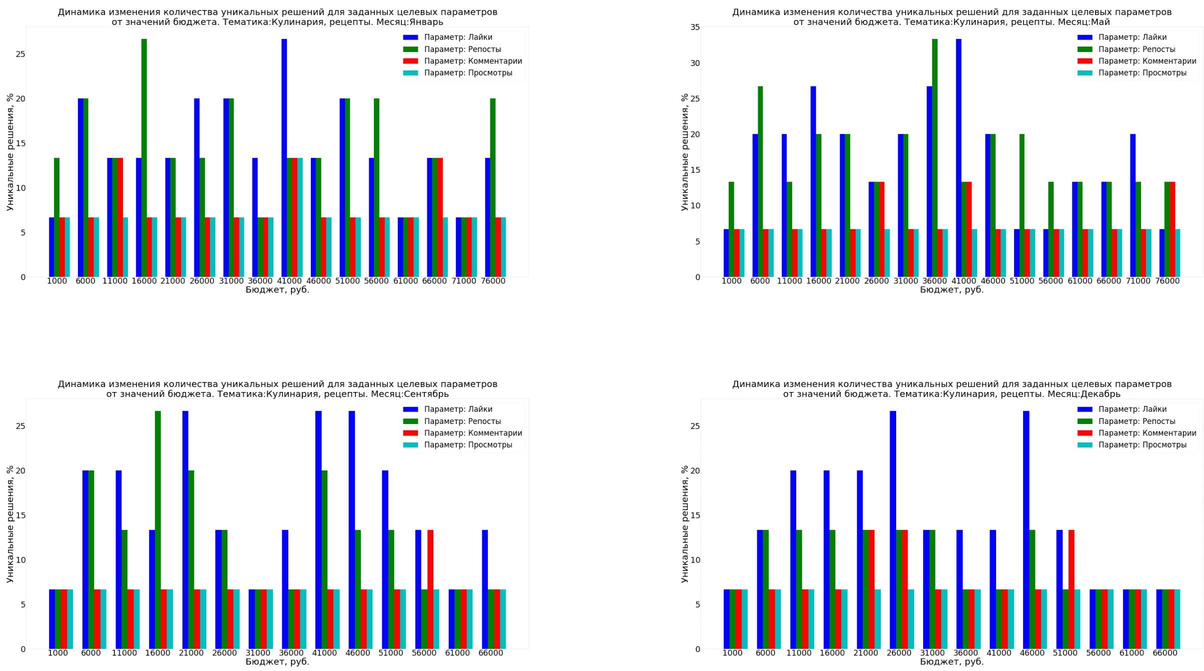


Figure 1.16: Dynamics of change in the number of unique solutions by criteria depending on the budget. Subject matter: «Culinary, recipes»

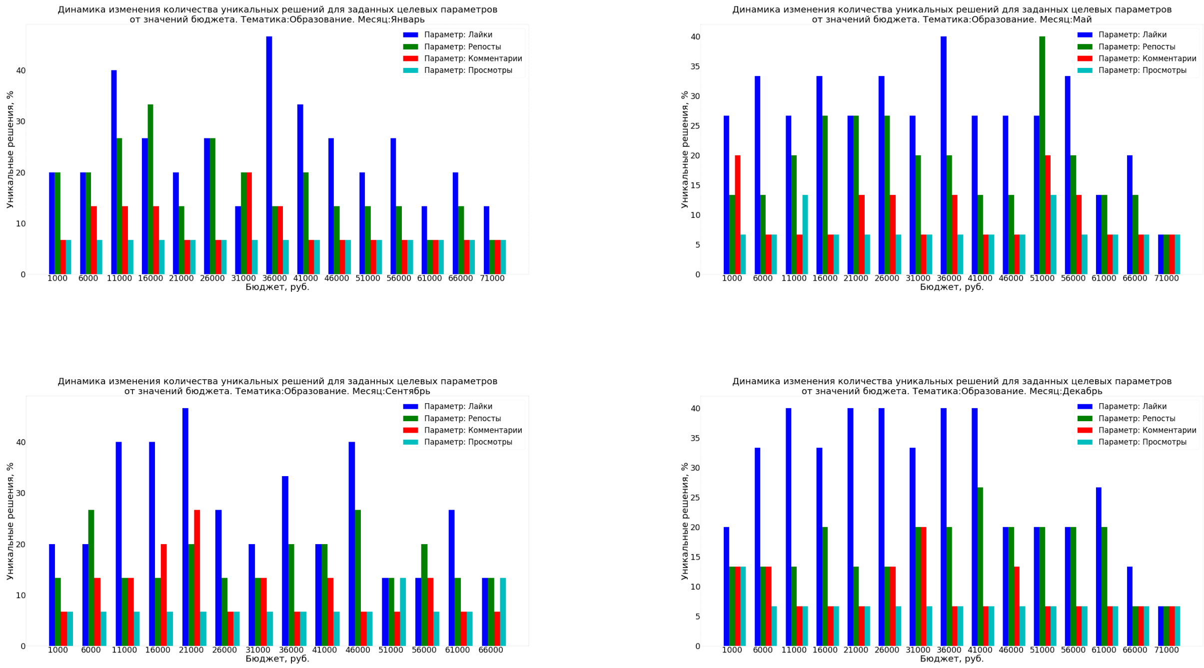


Figure 1.17: Dynamics of change in the number of unique solutions by criteria depending on the budget. Subject: «Education»

1.4. Conclusions to the first chapter

This chapter is devoted to the development of an optimization model, which is a tool for consulting on the correction of the initial preferences of the customer, taking into account the current dynamics of changes in the key characteristics of the selected product range on the market.

The chapter formulates problem statements for modeling the process of information dissemination in MCM using optimization methods. The services providing social network data are analyzed. Algorithms for processing statistical data on user activity of information sites in the task of information dissemination in MCM are developed and implanted. An architecture is proposed and an optimization model with visualization is implemented programmatically, which allows to form a set of social network communities with recommendations on information placement in them. Numerical simulation and sensitivity analysis of criteria in the problem of multi-criteria optimization are carried out, which showed that there is a differentiation of criteria sensitivity regardless of the nomenclature of goods, budget and time period on the example of the market of goods-services in the digital environment. It is shown that the specifics of behavioral activity of social network users are such that the change of preferences in the multicriteria optimization problem by one third affects the process of forming a unique scenario of information dissemination.

Chapter 2.
Modeling of site sets in the task of information dissemination based on machine learning methods

2.1. Problem statement and description

Cluster analysis methods are referred to the section of machine learning without a teacher, because the system under test is trained to perform the task without the experimenter's intervention, which is a type of cybernetic experiment. This approach allows solving problems when descriptions of a set of objects are known and it is necessary to discover internal interrelationships or regularities between objects (see Fig. 2.1).

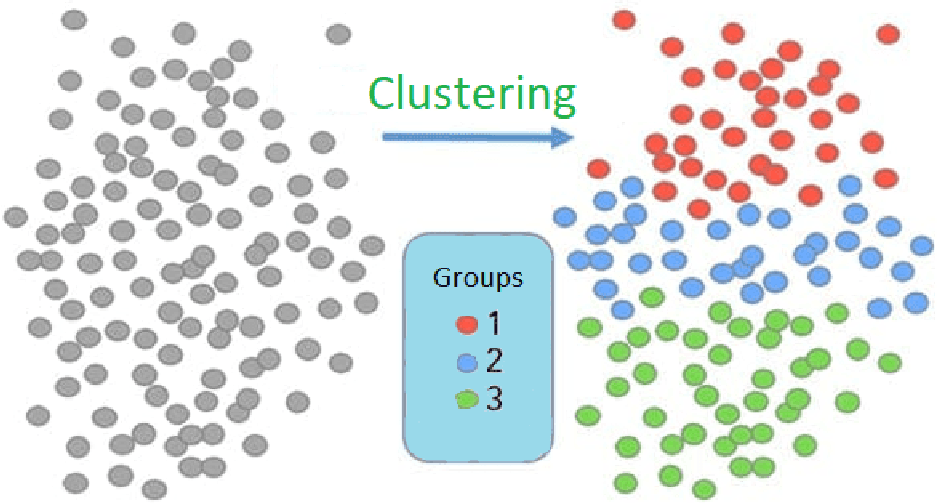


Figure 2.1: An example of how cluster analysis methods work

In the modern world, machine learning and artificial intelligence methods are becoming more and more widespread and in demand. The practical significance of the application of these methods lies in the possibility of creating more effective and innovative solutions in various fields. This chapter considers the problem of information dissemination in QMS using cluster analysis and optimization methods. To demonstrate the work of the proposed approach to modeling the process of information dissemination, one of such areas is taken — marketing, namely the task of dividing the set of all customers into clusters to identify typical preferences. In this research, the problem will be formulated differently, but its essence remains unchanged [70].

The chapter proposes a comprehensive model of information dissemination in the MCM, which forms a set of social network communities based on the given input parameters. The application of this approach will automate the process of management decision-making in this task, as well as reduce the time and financial costs of conducting relevant activities on the Internet. The proposed model takes into account the peculiarities of behavioral activity of the audience of each community, which contributes to maximizing the key indicators of feedback when placing paid ads. The use of mathematical statistics methods allows the DM to comprehensively evaluate the modeled scenarios of information promotion and effectively make operational and strategic decisions on the allocation of a limited amount of resources.

The solution presented in this paper allows us to model a set of communities with small budgets, taking into account the importance of the criteria formulated by the customer and the peculiarities of the behavioral activity of the audience. The relevance of using the proposed approach to automate the processes occurring in complex organizational systems is determined by the need to optimize the process of making managerial decisions in the field of communication technologies in the conditions of increasing volume of disseminated information. It should be noted that the development of such applied tools will make it possible to model processes in the field of economics and management, sociology and political science.

This chapter will discuss heuristic methods for solving the problem of determining the set of information dissemination sites in a MCM, as well as for dimensionality reduction in an optimization problem.

Cluster Analysis Task

Meaningful formulation of the problem: A client needs to create an image of his company in social media within a certain budget. The client wants to conduct advertising activities in such a way that as many web users as possible learn about his company. It is required to maximize the number of views of the published advertising record for a given topic, time period and budget. Note that one and the same advertising record can be published in several communities, and also the advertising record can be edited separately for each selected community taking into account the peculiarities of its audience.

Formal formulation of the problem: we use the notations introduced on page 21. We need to define such a cluster (set of communities) U_j from the partition $\{U_\alpha\}_{\alpha \in \Theta}$ of the set X, where $\Theta \subseteq M$ — some set of cluster indices, using the given clustering methods and their hyperparameters, so that it maximizes a given target index of j-th cluster in a given month of year t — $C_j(t) = \sum_{k=1}^{r_j} c_k$, $c_k(t) > 0$ — the value of the target index in k-th community, under the budget constraint $B_j = \sum_{k=1}^{r_j} b_k \leq P$, $b_k > 0$ — the cost of posting an entry in k-th community, $j \in \Theta$, r_j — the number of communities in j cluster.

Optimization problem with preliminary clustering

Meaningful formulation of the integer linear programming problem: a client needs to increase the volume of products it sells within a certain budget. The client wants to run an advertising campaign in such a way that as many web users as possible learn about his product. It is required to maximize the total number of views of the published advertising records for a given topic, time period, target parameter and budget. Note that the same advertisement record can be published in several communities, and also the advertisement record can be edited separately for each selected community taking into account the peculiarities of its audience.

Mathematical formulation of the integer linear programming problem: we use the notations introduced on pages 21, 53. We need to determine such a combination of clusters (community sets) $\{U_j\}_{j=1}^{s}$, $1 \leq s \leq |\Theta|$, from the partitioning $\{U_\alpha\}_{\alpha \in \Theta}$ of the set X, where $\Theta \subseteq M$ — some set of cluster indices, using the given clustering methods and their hyperparameters, which will achieve the maximum value of the given target index $\sum_{j=1}^{s} C_j$ for a given combination of clusters, under the constraint $\sum_{j=1}^{s} B_j \leq P$, $j \in \Theta$. The solution of the problem is represented by defining a set of clusters $U = (U_1, \ldots, U_{|\Theta|})$ that satisfies the following requirements:

$$f(U_\alpha) = \sum_{\alpha\in\Theta} C_\alpha \cdot U_\alpha \to max,$$

$$\sum_{\alpha\in\Theta} B_\alpha \cdot U_\alpha \leq P,$$

$$U_\alpha \in \{0; 1\},\ \alpha \in \Theta.$$

Meaningful formulation of the multi-objective optimization problem: a client needs to carry out brand positioning within a certain budget. The client wants to conduct promotional activities in such a way that as many online users as possible learn about their product. However, the client prefers communities in which people leave feedback in comments under the entries, as well as actively share community entries on their personal page. Therefore, it is required to maximize the number of comments and "share" marks on the published advertising record for a given topic, time period and budget, and possibly the presence of a portrait of the target audience. Thus, for the client two indicators are in equilibrium, the values of which we will maximize. Note that one and the same advertising record can be published in several communities, as well as the advertising record can be edited separately for each selected community, taking into account the characteristics of its audience.

Mathematical formulation of the multi-objective optimization problem: To the notations introduced on pages 21, 53, 54 we add 4 criteria defined by functions: $f_1(U_\alpha), f_2(U_\alpha),\ f_3(U_\alpha), f_4(U_\alpha)$ and set the values of the corresponding criteria in the j-th cluster for the average record in a given month of year t — $A_j(t)$, $B_j(t)$, $C_j(t)$, $D_j(t)$, where $A_j(t) = \sum_{k=1}^{r_j} a_k$, $B_j(t) = \sum_{k=1}^{r_j} b_k$, $C_j(t) = \sum_{k=1}^{r_j} c_k$, $D_j(t) = \sum_{k=1}^{r_j} d_k$, and $a_k(t) > 0$, $b_k(t) > 0$, $c_k(t) > 0$, $d_k(t) > 0$; let us redefine the cost of placing an advertising record in j-th cluster (set of communities) — $G_j = \sum_{j=1}^{r_j} g_k$, where $g_k > 0$ — the cost of an advertising record in k-th community. The solution to the problem is represented by defining a set of clusters $U = (U_1, \ldots, U_{|\Theta|})$ that satisfies the following requirements:

$$\begin{cases} f_1(U_\alpha) = \sum_{\alpha\in\Theta} A_\alpha \cdot U_\alpha \to max, \\ f_2(U_\alpha) = \sum_{\alpha\in\Theta} B_\alpha \cdot U_\alpha \to max, \\ f_3(U_\alpha) = \sum_{\alpha\in\Theta} C_\alpha \cdot U_\alpha \to max, \\ f_4(U_\alpha) = \sum_{\alpha\in\Theta} D_\alpha \cdot U_\alpha \to max, \\ \sum_{\alpha\in\Theta} G_\alpha \cdot U_\alpha \le P, \\ U_\alpha \in \{0;1\},\ \alpha \in \Theta. \end{cases} \tag{2.1.1}$$

Using the method of convolution of criteria, transform the system (2.1.1):

$$\hat{f}(U_\alpha) = \beta_1 \cdot f_1(U_\alpha) + \beta_2 \cdot f_2(U_\alpha) + \beta_3 \cdot f_3(U_\alpha) + \beta_4 \cdot f_4(U_\alpha) =$$
$$= \beta_1 \cdot \sum_{\alpha\in\Theta} A_\alpha \cdot U_\alpha + \beta_2 \cdot \sum_{\alpha\in\Theta} B_\alpha \cdot U_\alpha + \beta_3 \cdot \sum_{\alpha\in\Theta} C_\alpha \cdot U_\alpha +$$
$$+\beta_4 \cdot \sum_{\alpha\in\Theta} D_\alpha \cdot U_\alpha = \sum_{\alpha\in\Theta} (\beta_1 \cdot A_\alpha + \beta_2 \cdot B_\alpha + \beta_3 \cdot C_\alpha + \beta_4 \cdot D_\alpha) \cdot U_\alpha =$$
$$= \sum_{\alpha\in\Theta} W_\alpha \cdot U_\alpha.$$

Applying the above linear transformations to (2.1.1), we obtain a system of the form:

$$f(U_\alpha) = \sum_{\alpha\in\Theta} W_\alpha \cdot U_\alpha \to max,$$
$$\sum_{\alpha\in\Theta} G_\alpha \cdot U_\alpha \le P,$$
$$U_\alpha \in \{0;1\}, \alpha \in \Theta.$$

where $W_\alpha = \beta_1 \cdot A_\alpha + \beta_2 \cdot B_\alpha + \beta_3 \cdot C_\alpha + \beta_4 \cdot D_\alpha$ – this is a measure of total publication activity for the selected time period; $\beta = \{\beta_1,\ \beta_2,\ \beta_3,\ \beta_4\}$ – the criteria weights or customer preferences.

2.2. Complex model using cluster analysis methods

The architecture of the model is represented as a block diagram in Figure 2.2 and is implemented using five main functional blocks:

1. Data preprocessing;
2. Data processing;
3. Generating recommendations for the publication of advertising records;
4. Partitioning construction (object clustering);
5. Optimization.

The clustering problem formulated in paragraph 2.1 differs from the dimensionality reduction problem by the presence or absence of an optimization block. This can be verified by studying the scheme shown in Figure 2.2.

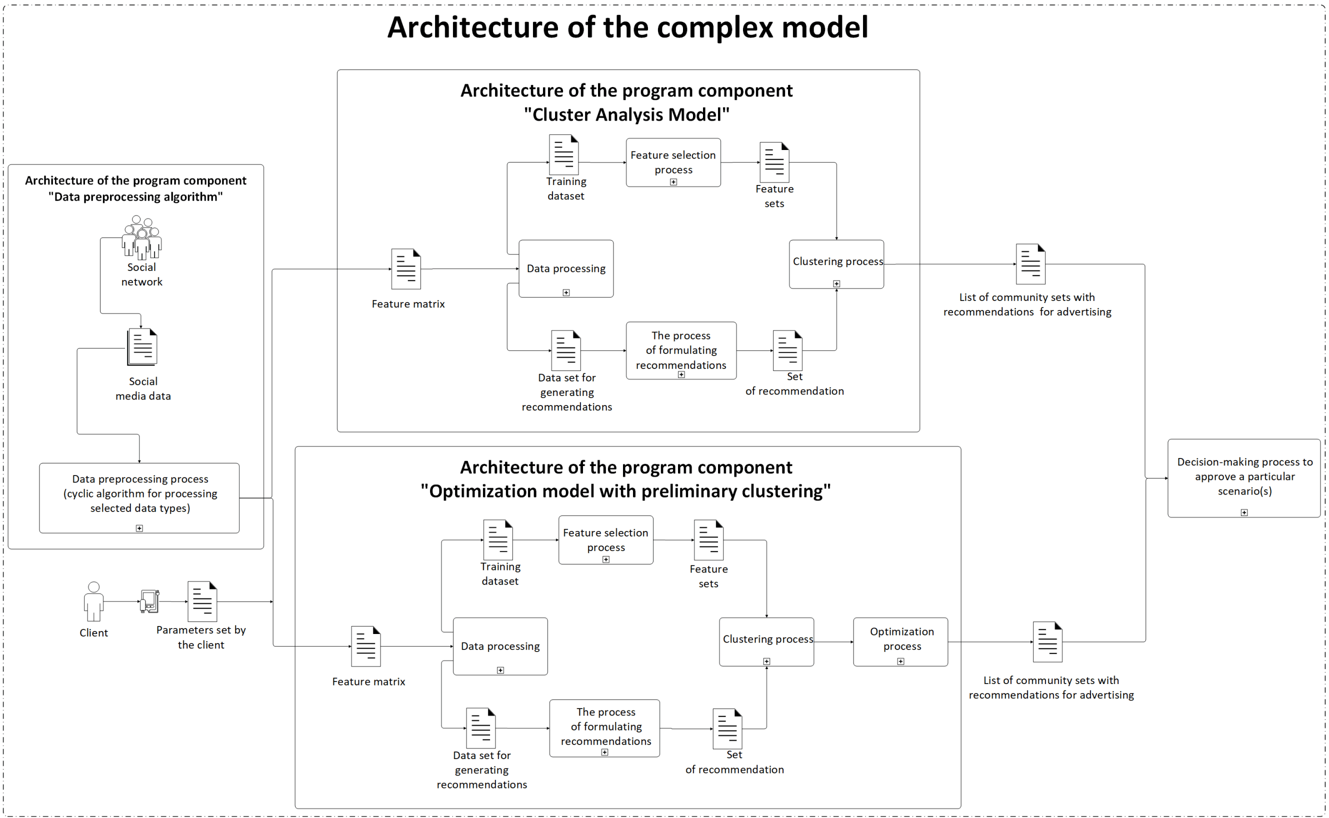


Figure 2.2: Model architecture using cluster analysis and optimization techniques

Let's consider each of the stages in more detail. Note that «Data preprocessing process (cyclic algorithm for processing selected data types)», «Data processing process», «Recommendations formation process» were described in Chapter 1.

2.2.1. Methods of selection and feature extraction

In this paper, both feature selection and feature extraction methods are applied to implement the feature space compression procedure.

Feature selection is a procedure for evaluating the significance of certain features using statistical methods and, among others, machine learning algorithms in order to form a feature space of lower dimensionality. Feature selection is used for four reasons:

- to «simplify» the model and make the results more interpretable;
- to shorten the learning curve;
- to avoid the «curse» of dimensionality;
- to improve the generalizability of the model and combat overfitting.

The purpose of applying feature selection techniques is to remove redundant data or insignificant features without significant loss of information. «Insignificance and redundancy are different concepts, this is because one significant feature may be redundant in the presence of another significant feature that is highly correlated with it. The choice of an appropriate feature selection method depends on the specifics of the problem and the available data.

Feature extraction, unlike the selection procedure, aims to generate new features as a function of the original features, while the result of the selection procedure is a subset of features.

There are sufficient number of algorithms suitable for various application problems and data analysis. In this study, feature selection and feature extraction methods will be used for the task of teacherless learning. In the developed model, feature selection methods are represented as functions (see Figure 2.3). It should also be noted that some algorithms return a single feature vector, while others return one or more. Multi-cluster feature selection methods in teacherless machine learning tasks are an effective approach to reduce the dimensionality of the data and extract the most informative features. These methods are based on the idea of grouping features into several clusters, which allows to reveal the structure of the data and identify the most significant features for each cluster.

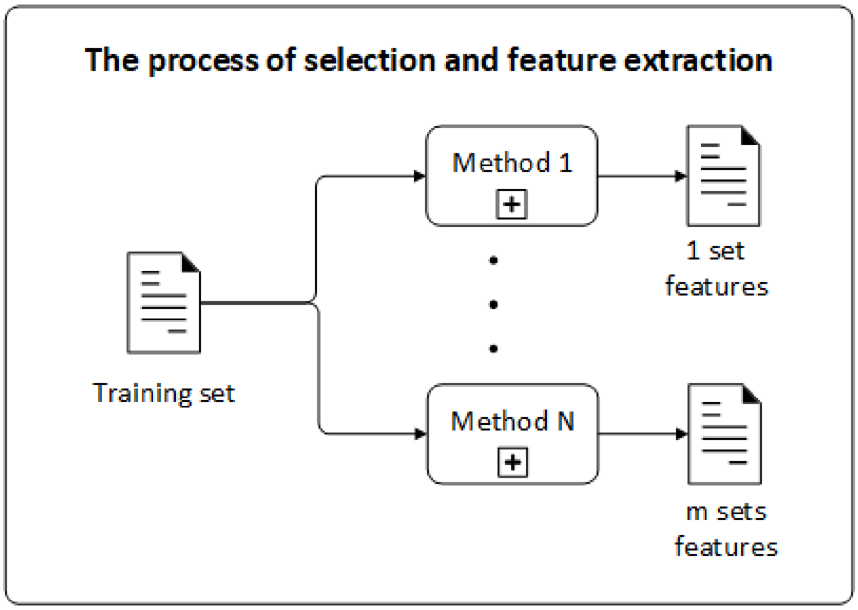


Figure 2.3: Scheme of implementation of selection and feature extraction

Let's take a closer look at the methods used:

1. Dispersion: It has been shown that estimating the variance of a feature can be an effective way of selecting features. Typically, features with near-zero variance are not significant and can be removed [50].

2. Mean absolute difference: the average absolute difference between the values of a trait and its mean value is calculated. Higher values tend to have higher predictive power [85].

$$MAD_i = \frac{1}{n} \times \sum_{j=1}^{n} |X_{ij} - \overline{X_i}|$$

3. Ratio of dispersions: arithmetic mean divided by geometric mean. Higher dispersion corresponds to more relevant features [85].

$$AM_i = \overline{X} = \frac{1}{n} \times \sum_{j=1}^{n} X_{ij}$$

$$GM_i = \prod_{j=1}^{n} X_{ij}$$

Since $AM_i \geq GM_i$ is satisfied if and only if $X_{i1} = X_{i2} = ... = X_{in}$, consequently:

$$R_i = \frac{AM_i}{GM_i} \in [1, +\infty)$$

4. Laplace estimation: based on the observation that data from the same class are often closer together, so it is possible to assess the importance of a feature by its ability to reflect this proximity. The method consists of embedding the data in a nearest neighbor graph by measuring an arbitrary distance and then computing a matrix of weights. Then for each feature we calculate Laplace criterion and obtain the property that the smallest values correspond to the most important dimensions [102].

5. Multi-cluster feature selection: features are selected to best preserve the multi-cluster structure of the data. The number of clusters was set from 2 to the maximum number of features in increments of 5 percent of the total number of features. Out of all the resulting variants of feature sets, three unique ones were selected [98].

6. Principal component method: is an algebraic method of transforming a set of observations, possibly correlated variables, into a set of values of linear uncorrelated variables [55].

Thus, feature selection methods in teacherless machine learning tasks play an important role in data processing, allowing us to focus on the most relevant aspects and improve the quality of subsequent analysis without the need for expert data partitioning.

Let us present the results of the first five feature selection methods (see Fig. 2.4, 2.5, 2.6 , 2.7 , 2.8).

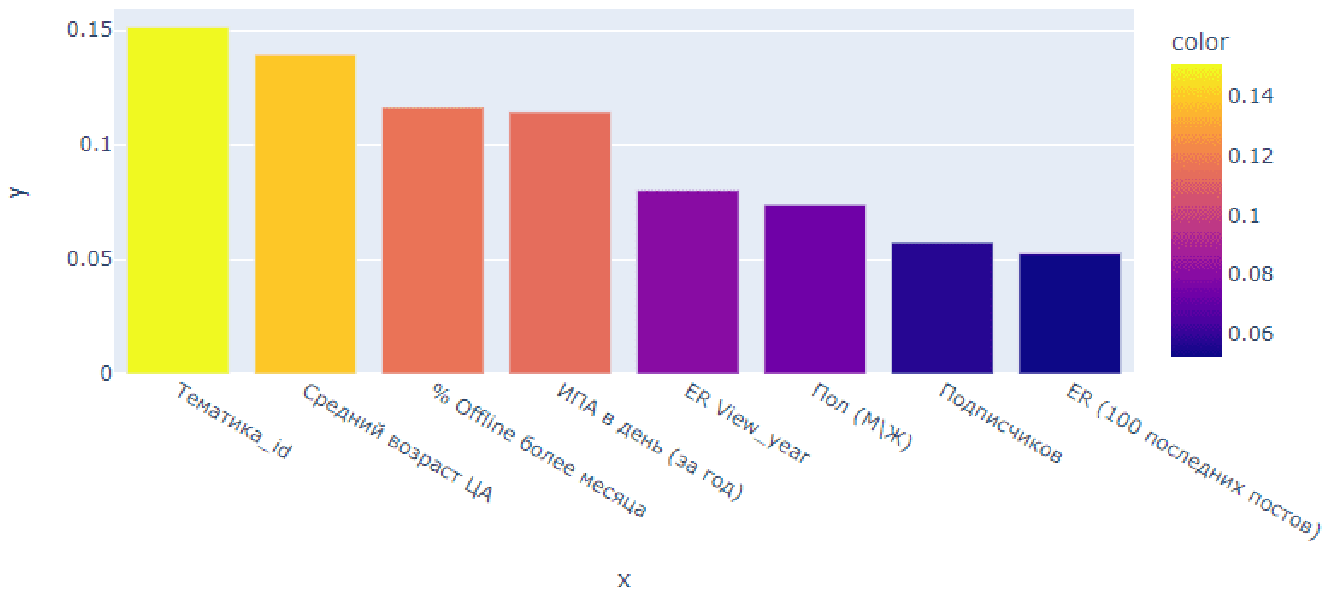


Figure 2.4: Result of applying the method: «Dispersion»

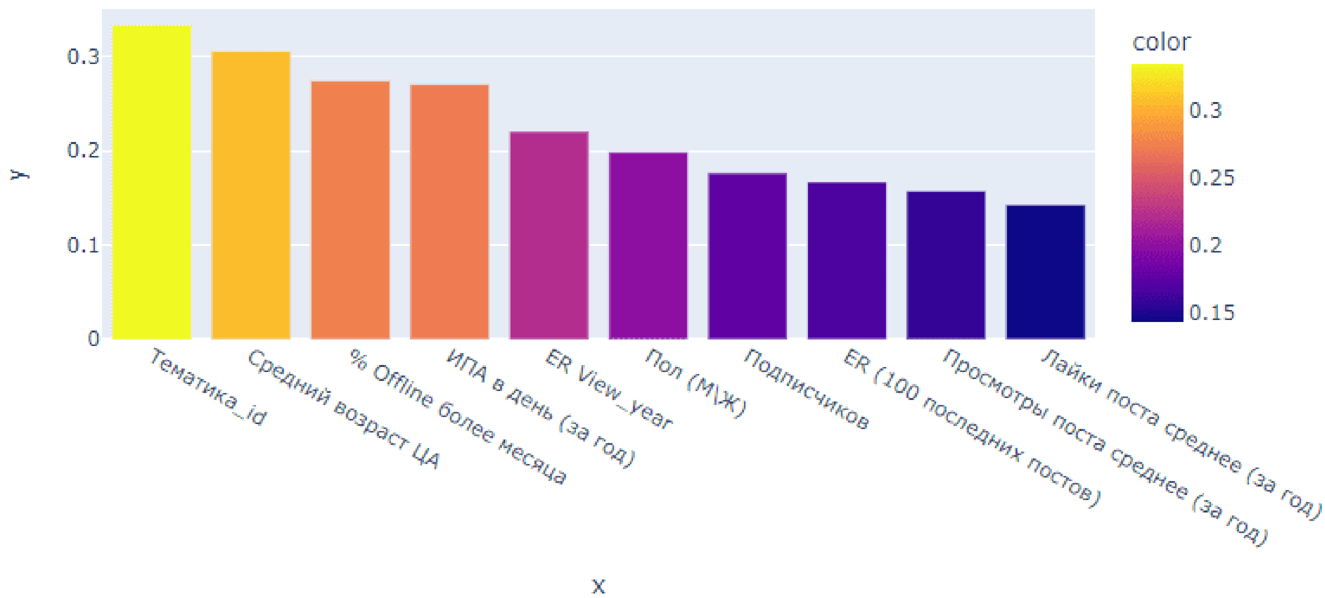


Figure 2.5: Result of the method application: «Average absolute difference»

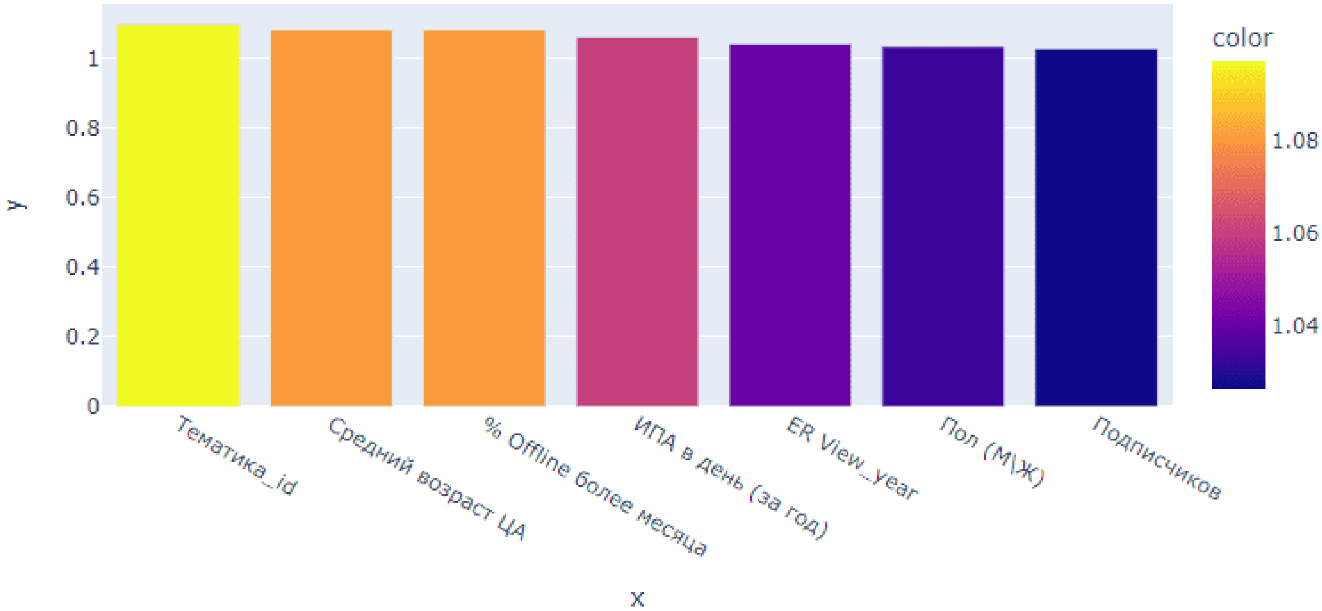


Figure 2.6: Result of the method: «Ratio of dispersions»

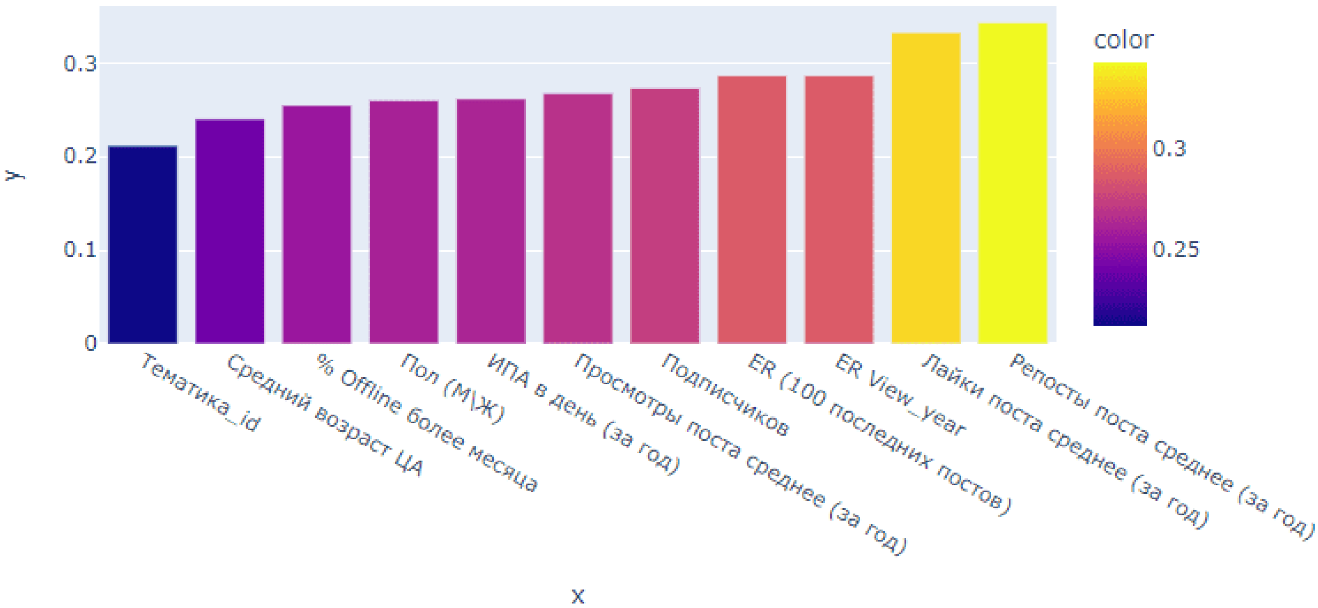
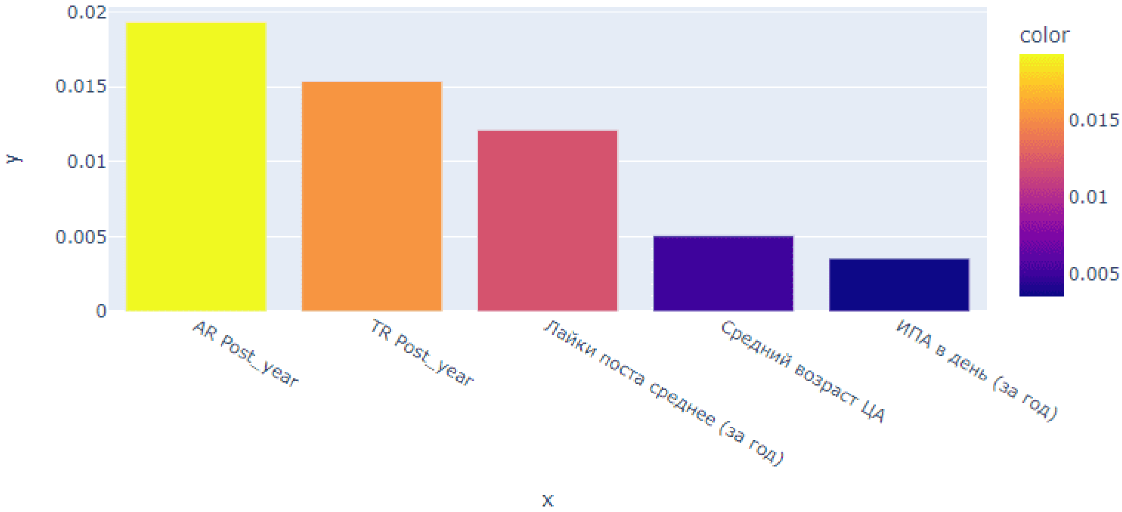
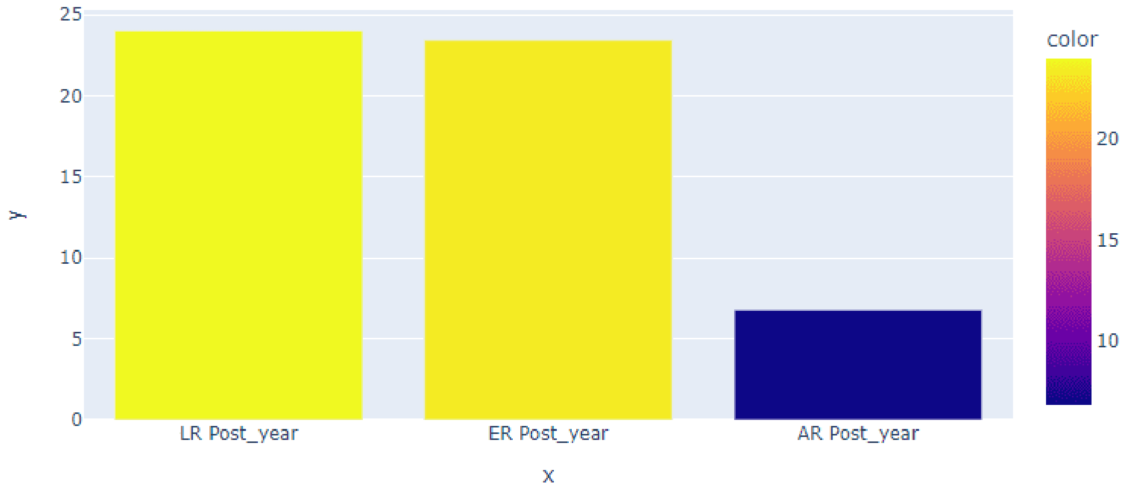


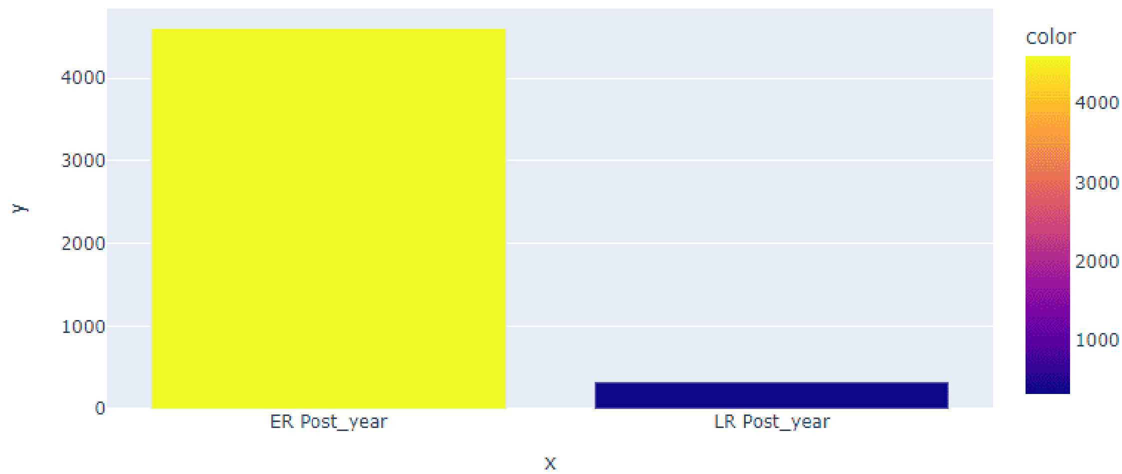
Figure 2.7: Result of the method application: «Laplace estimation»



(a) For the parameter n_clusters = 2



(b) For the parameter n_clusters = 171



(c) For the parameter n_clusters = 678

Figure 2.8: Method result: «Multicluster feature selection»

The peculiarity of the implementation of the first five methods is that the control values for each feature were calculated, and then the mean was calculated and the features that took control values above the mean were selected. Thus, we obtained 7 sets of the most significant features for methods 1-5. Then these sets are fed into the clustering model, training takes place, selection of the best hyperparameters of the clustering models according to the external criterion of the quality of the obtained partitions.

It should be noted that the basic set of significant characteristics are: «Topic_id»>, «Age of TA», «% Offline more than a month», «IPA per day (for a year)», «ER View_year», «Gender (M/W)», «Subscribers». Removing redundant attributes allows for a better understanding of the data, as well as reducing model setup time, improving model accuracy, and making the results easier to interpret. Sometimes this task may even be the most significant one, for example, finding the optimal set of features can help decipher the mechanisms underlying the problem under study.

Principal Component Method — one of the main ways to reduce the dimensionality of data while losing the least amount of information. Mathematically, the principal component method is an orthogonal linear transformation that maps data from the original feature space into a new feature space of lower dimensionality. The goal of the principal component method is to construct a new feature space of lower dimensionality, with the variance between the axes redistributed in a way that maximizes the variance of each axis. To do this, a sequence of the following steps is performed:

- The total variance of the original feature space is calculated. This cannot be done by simply summarizing the variance of each variable, because in most cases they are not independent. Therefore, it is necessary to summarize the mutual dispersions of the variables, which are determined from the covariance matrix.

- The eigenvectors and eigenvalues of the covariance matrix, which determine the directions of the principal components and the magnitude of the associated variance, are calculated.

- Dimensionality reduction is performed. The diagonal elements of the covariance matrix show the variance in the original coordinate system, while its eigenvalues show the variance in the new coordinate system. Then, dividing the variance associated with each principal component by the sum of the variances of all components, we obtain the fraction of variance associated with each component. After that, we discard so many principal components that the share of the remaining ones is 80-90%.

A significant disadvantage of using this method to implement the feature extraction procedure is the impossibility of meaningful interpretation of the components, since they "absorb" the variance from several original variables. In addition, it is necessary to check all or most of the principal components, i.e., if in the previous methods we obtained a set of features, which we fed into the clustering model and then determined the best hyperparameters and some set of communities according to an external criterion, then here, it is necessary to analyze the clustering results for 2,3,4,5,...,n principal components, where n is the total number of features, and select the best value to obtain the optimal result according to the external criterion. As a consequence, this method requires a separate software implementation architecture from the previous ones.

2.2.2. Cluster analysis methods and partitioning quality metrics

Data clustering is one of the key tools of machine learning whose main goal is to group objects based on their similarity, so that objects in one group (cluster) are more similar to each other than to objects in other groups. This method finds wide application in various fields such as bioinformatics, financial analytics, image processing and many others. There are several major data clustering methods including hierarchical methods (agglomerative and divisional clustering), k-means method, DBSCAN method, spectral clustering methods and many others. Each of these methods has its own advantages and disadvantages, and the choice of a particular method depends on the characteristics of the original data, the objectives set and the required interpretability of the results. The results of cluster analysis can be used to identify groups of similar objects, create audience segmentation, identify behavior patterns, and many other applied tasks. To conduct this research, several clustering methods were considered (see Fig. 2.9), and their applicability for solving the formulated tasks was analyzed using different methods of feature selection.

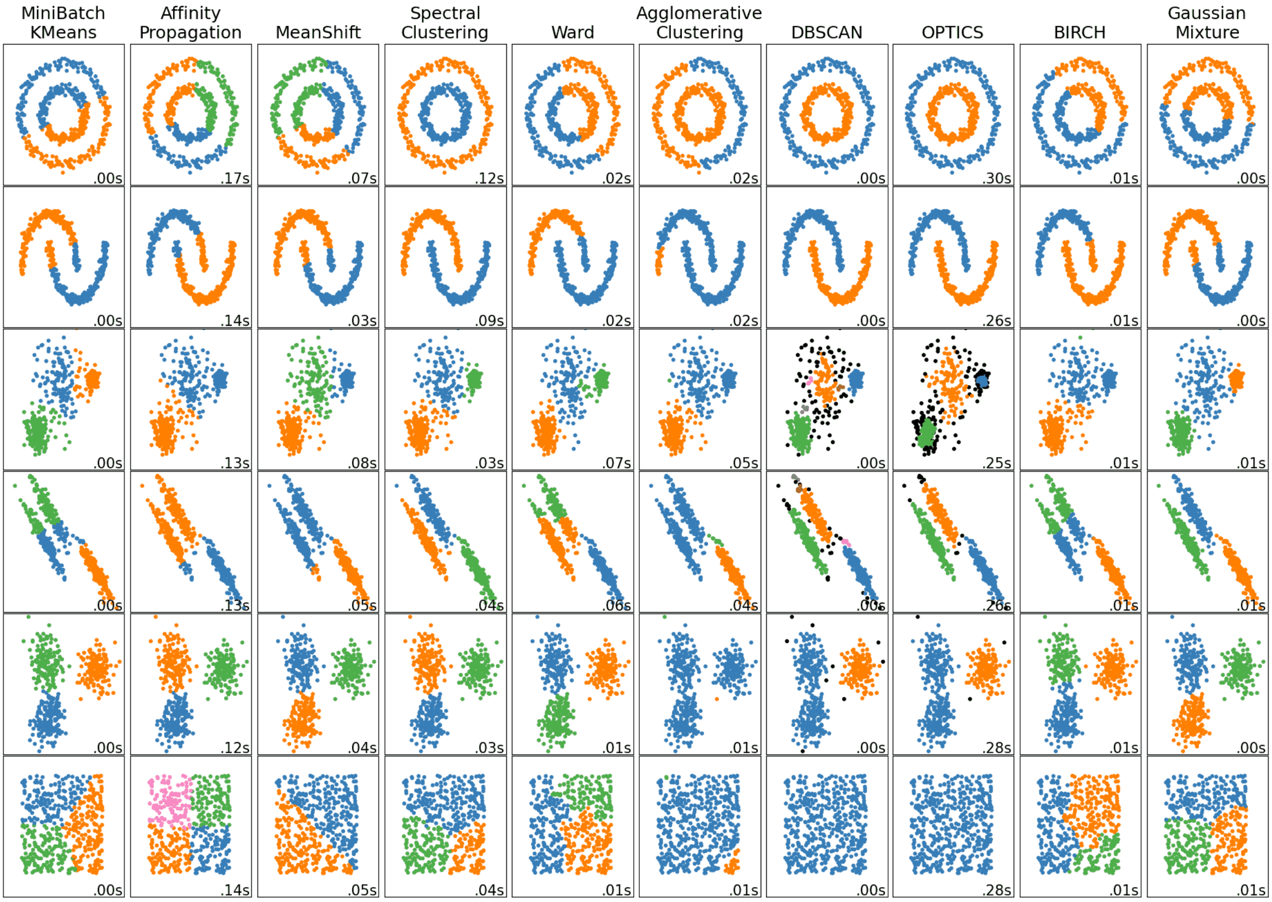


Figure 2.9: Cluster analysis methods

In this section, the cluster analysis methods will be presented in the form of a table 2.1 with a description of the specified hyperparameters and intervals of their values. Note that such a number of selected clustering methods is justified by the comparative analysis in the considered applied problem. Let us proceed to the consideration of cluster analysis methods [48].

Table 2.1: Cluster analysis methods

№	Clustering method	Hyperparameters	Hyperparameter values
1	«Agglomerative clustering»	«n_clusters»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
		«linkage»	[«complete»; «average», «single»]
2	«Ward»	«n_clusters»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
3	«K-Means»	«n_clusters»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
4	«MiniBatch K-Means»	«n_clusters»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
5	«Affinity propagation»	«damping»	from 0.5 to 0.95 with step 0.05
6	«Mean-shift»	-	-
7	«Spectral clustering»	«n_clusters»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
		«gamma»	$[10^{-2}; 10^{-1}; 1; 10^{1}; 10^{2}]$

Continuation of table 2.1

8	«DBSCAN»	«eps»	from 0.1 to 0.9 with step 0.2
		«min_samples»	from 5 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
9	«OPTICS»	«min_samples»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
10	«Gaussian mixtures»	«n_components»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
		«covariance_type»	[«full»; «tied»; «diag»; «spherical»]
11	«BIRCH»	«n_clusters»	from 2 to $(n-1)$ with step $(0.1 \times n)$, where n - is the number of objects in a given set
		«threshold»	from 0.1 to 0.9 with step 0.2

In the following, we present measures for assessing the quality of the resulting partitions, relying only directly on the structure of the clusters, without using external information. The problem of quality assessment in the clustering problem is intractable for at least two reasons: 1) Kleinberg's impossibility theorem [105] – there is no optimal clustering algorithm; 2) Many clustering algorithms are unable to determine the true number of clusters in the data, most often the number of clusters is given as an input to the algorithm and is picked up by several runs of the algorithm. Therefore, this paper considers several quality metrics of the resulting partitions [49] as a result of clustering algorithms for comparative analysis of numerical simulation results:

- Index «Silhouette».

  The silhouette value indicates how similar an object is to its cluster compared to other clusters [108]. Score for the entire cluster structure:

  $$Sil(C) = \frac{1}{N} \cdot \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{b(x_i,c_k) - a(x_i,c_k)}{max\{a(x_i,c_k); b(x_i,c_k)\}},$$

  where: $a(x_i,c_k) = \frac{1}{|c_k|} \cdot \sum_{x_i \in c_k} \|x_i - x_j\|$ - average distance from $x_i \in c_k$ to other objects in the cluster c_k (compactness);
  $b(x_i,c_k) = min_{c_l \in C \setminus c_k}\{\frac{1}{|c_l|} \cdot \sum_{x_j \in c_l} \|x_i - x_j\|\}$ - average distance from $x_i \in c_k$ to objects from another cluster $c_l : k \neq l$ (separability).

  The closer this score is to 1, the better it is.

- Index «Calinski–Harabasz».

  Compactness is based on the distance from cluster points to their centroids, and separability is based on the distance from the centroids of clusters to the global centroid [99]. Represents the ratio of the mean of the variance between clusters to the variance within a cluster. For a dataset E of size n_E that has been partitioned into k clusters:

  $$CH = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1},$$

  where: $W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - c_q) \cdot (x - c_q)^T$ - inter-cluster variance matrix; $B_k = \sum_{q=1}^{k} n_q \cdot (c_q - c_E) \cdot (c_q - c_E)^T$ - intracluster dispersion matrix; C_q - set of cluster objects q; c_q - cluster center q; c_E - center E; n_q - number of objects in the cluster q.

  The higher the value, the better the clustering is done.

- Index «Davies–Bouldin Index».

  This is perhaps one of the most used measures of clustering quality [100]. It computes compactness as the distance from cluster objects to their centroids, and separability as the distance between centroids. That is, it shows the average "similarity" of clusters: the distance between them is compared to their size.

$$DB = \frac{1}{k} \cdot \sum_{i=1}^{k} max_{i \neq j}\{R_{ij}\},$$

  where: $R_{ij} = \frac{s_i + s_j}{d_{ij}}$ - cluster similarity measure i и j, s_i - average distance between each point in the cluster i and the centroid of this cluster, i.e. the diameter of i-th cluster; d_{ij} - distance between cluster centroids i and j.

  The smaller the value, the better the clustering is done.

2.2.3. Description and implementation of a complex model

Complex mathematical models are widely used in science, engineering, economics and other fields to analyze, predict and optimize various systems. They allow researchers and specialists to gain a deep understanding of system behavior under different conditions, conduct numerical experiments and simulate scenarios. The use of complex mathematical models allows the DM to make informed decisions, optimize processes and improve the overall system performance. Using a complex model with machine learning techniques, managers can analyze large amounts of data, identify hidden dependencies and optimize decision-making strategies. Such models are able to take into account multiple variables and factors, allowing for more accurate and informed management decisions.

The developed complex model is implemented as follows: the input of the function implementing the clustering method is given data: for training, for recommendations, topics, target, client's budget, month of the year. Then the following items are performed:

1. Formation of a subset of objects of the specified topics.

2. The starting budget, step and maximum possible budget are defined for it. Since the data is limited, it is necessary to determine the largest possible client budget for each set of topics.

3. Tables are created to record the results of the simulation.

4. Nested loops are executed for a given budget, target, topics, month of the year, and one of the normalized training datasets.

5. In the loop body, the function that implements cluster analysis methods is called. It calls functions from the «Formation of recommendations» block.

6. Vectors of hyperparameter values of the clustering method are set and nested loops are run.

7. A clustering method is performed and some partitioning $\{U_\alpha\}_{\alpha\in\Theta}$ of the set X is obtained.

8. The quality metrics of the partitioning are computed.

9. For each cluster, the total cost of record placement is calculated and those clusters whose cost does not exceed the client's specified budget are selected: $\{G_\alpha \mid G_\alpha \leq P, \alpha \in \Theta\}$.

10. Next, depending on the approach is performed:

    - if it is the task of determining the best cluster among the obtained partitions, then for each cluster the total value of the target indicator $C_j(t)$ is calculated, the maximum of $C_j(t)$ is calculated and the best cluster is determined in this way;

    - If it is a multicriteria optimization problem with preliminary clustering, then, as in Chapter 1, in order to optimize the running time of algorithms solving the set problems, it was decided to implement a single function that solves both the multicriteria optimization problem and the integer linear programming problem.

- then using «scipy.optimize» [12] library the optimization problem is solved and the resulting vector is determined U.
- values of corresponding criteria for optimal combinations of clusters at a given matrix of weighting coefficients are calculated.

11. The simulation results are recorded in comparative tables.

Thus, the proposed comprehensive model based on machine learning and optimization techniques is an integrated approach to data analysis and modeling of information dissemination scenarios in MCM for management decision making. This model combines machine learning algorithms with an integrated view of a system or process to provide a deep understanding of the relationships and patterns in the data. Applying complex models with machine learning and optimization techniques in management activities can improve decision making, minimize risk, and enhance strategic planning based on data and analytics.

2.3. Numerical modeling and comparative analysis of results

To demonstrate the operation of the proposed approach and the possibility of comparative analysis of the obtained results, we set the following input parameters:

1. Topics – «Automobiles, car owners», «Culinary, recipes», «Education». Number of objects in each theme: 124, 126, 103;

2. Time intervals – monthly, January, May, September, December;

3. Budget – from 1000 rubles to the maximum possible budget within the given topics and months of the year with a step of 5000 rubles;

4. Customer preferences are given in vector form $\beta = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ and are presented in the table below 2.2:

Table 2.2: Weighting factors for modeling

β_1	β_2	β_3	β_4
1.0	0.0	0.0	0.0
0.0	1.0	0.0	0.0
0.0	0.0	1.0	0.0
0.0	0.0	0.0	1.0
0.25	0.25	0.25	0.25

Note that the parameters specified in item 4 will be used only when solving the optimization problem with preliminary markup of objects, and such a number of variants of vectors of weight coefficients is explained by the need to demonstrate the possibility of obtaining different solutions when changing preferences.

For program realization we used Python programming language, PyCharm development environment, as well as corresponding libraries: Pandas, Numpy, Datetime, Scikit-learn, Math, SciPy, Matplotlib, Glob. It is important to note that the components of the complex model have been programmatically implemented and the corresponding computer programs have been registered with FIPS [80, 82].

It should also be pointed out that this complex model has been designed in the form of separate blocks and functions in such a way that it allows the researcher, if external failures occur during the simulation, to seamlessly continue it from the point where it was interrupted. There are four such blocks in total:

- Block 1: clustering using interpretable feature selection methods.
- Block 2: clustering using principal component method to compress the feature space.
- Block 3: Pre-clustering using interpretable feature selection methods in an optimization problem.
- Block 4: Pre-clustering using principal component method to compress the feature space in an optimization problem.

As a consequence, the formulation of conclusions based on the modeling results will be based on the type of observations. Let us proceed to their consideration and analysis.

Conclusion 2.1. *The obtained solutions show a decrease in the number of clusters with increasing budget when applying cluster analysis methods regardless of the selected topics, target parameters and months of the year (see Fig. 2.10, 2.11).*

This conclusion allows us to conclude that when the budget increases, it is necessary to decrease the values of the «n_clusters» hyperparameter in order to reduce the training time in the corresponding clustering methods. This is explained by the procedure of selecting the best cluster among the resulting partitioning at given hyperparameters of the method described in paragraph 2.2.

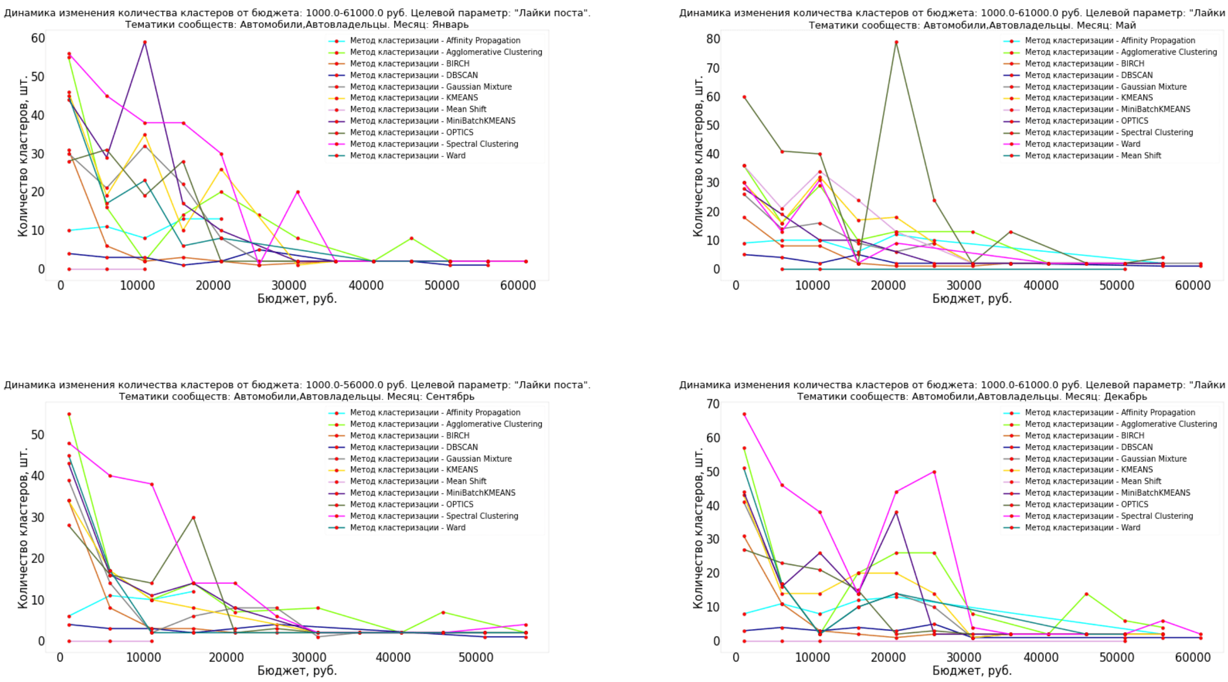


Figure 2.10: Dynamics of change in the number of clusters depending on the budget. Block 1

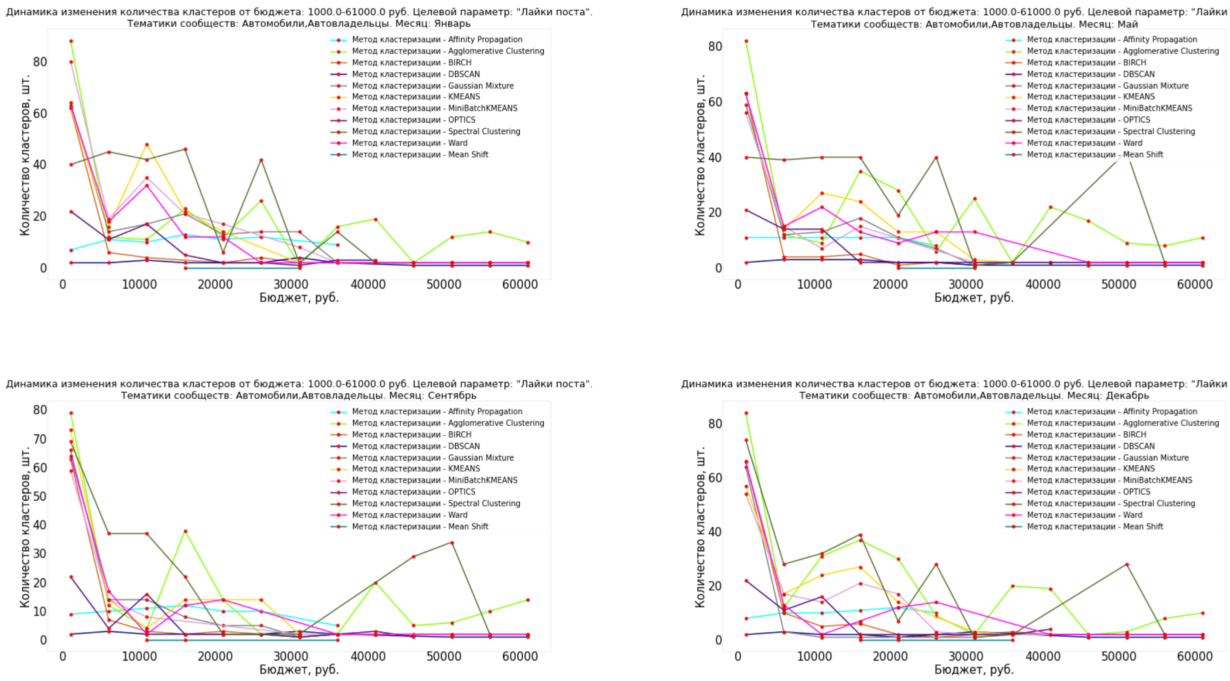


Figure 2.11: Dynamics of change in the number of clusters depending on the budget. Block 2

Conclusion 2.2. *When applying optimization methods with preliminary clustering, the change in the number of clusters with increasing budget occurs, if at all, then by some small value regardless of the selected topics, target parameters, budget and months of the year (see Fig. 2.12, 2.13).*

It should be noted that when performing the procedure of preliminary object partitioning, some cluster analysis methods work in such a way that the dimensionality reduction in the linear programming problem is up to $25-30\%$, while others reduce the number of variables by 70% and more. This can be explained by different principles of set partitioning in the considered clustering methods. In addition, this conclusion allows us to conclude that there is a possibility to reduce the training time by setting limits for the values of the «n_clusters» hyperparameter in the corresponding algorithms.

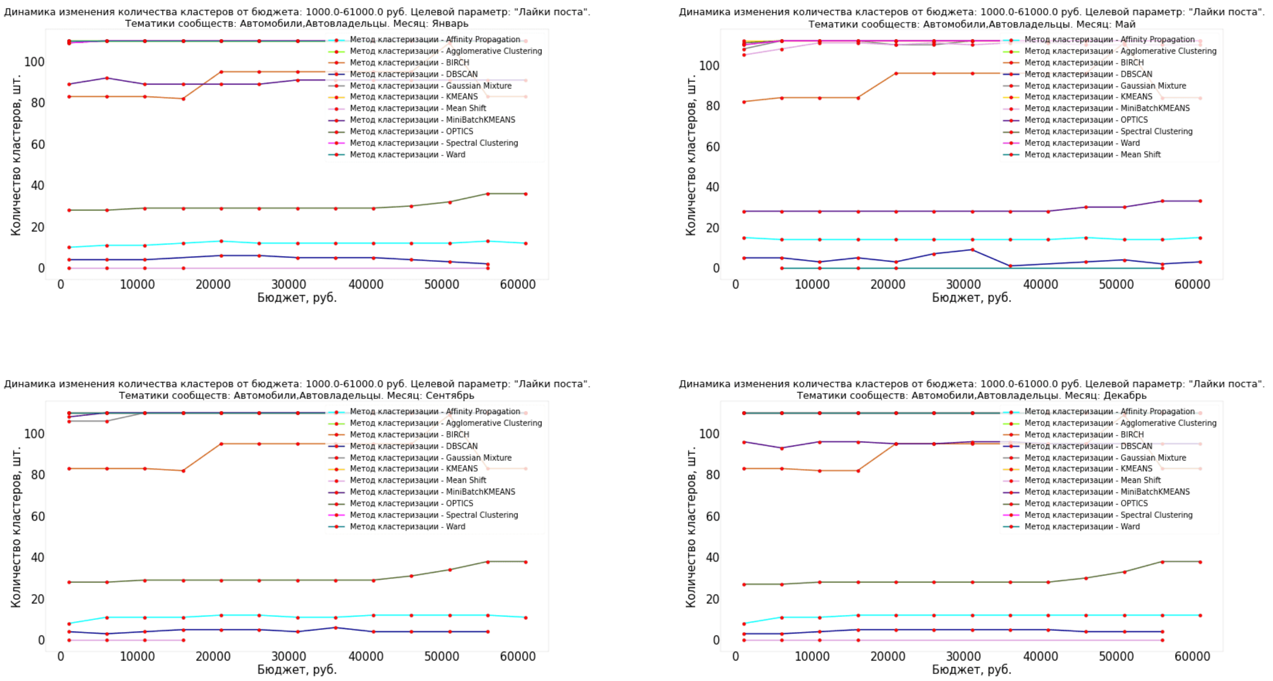


Figure 2.12: Dynamics of change in the number of clusters depending on the budget. Block 3

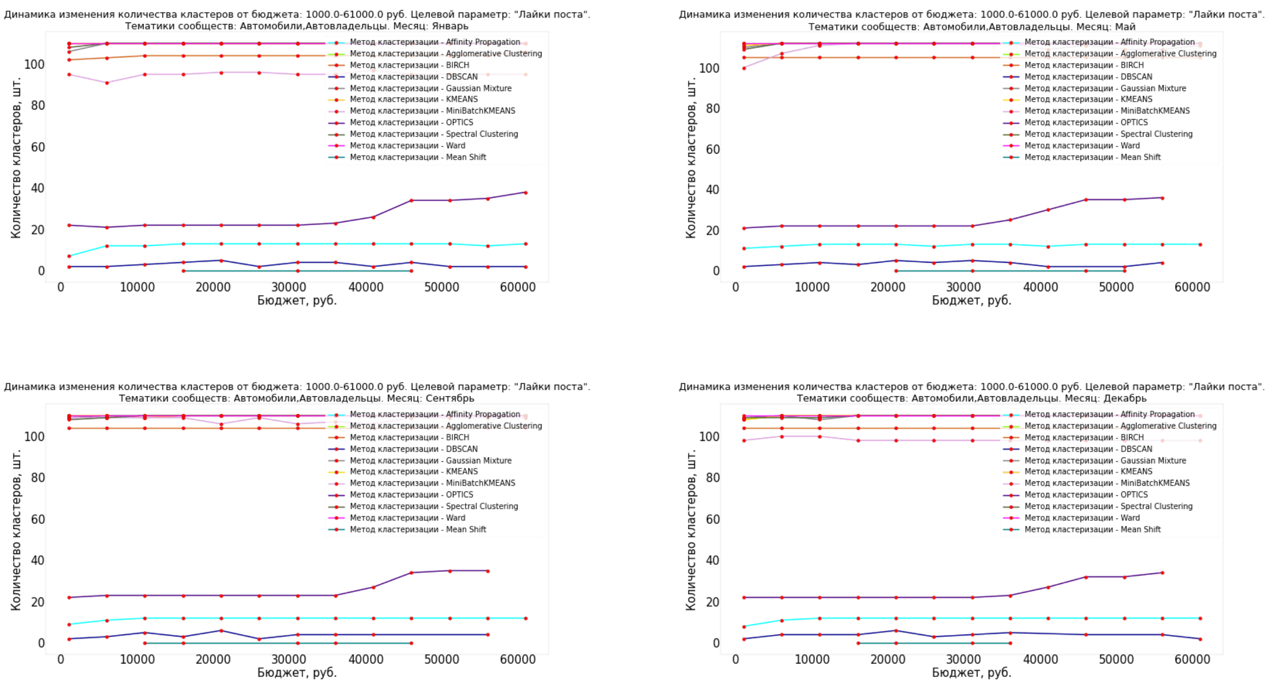


Figure 2.13: Dynamics of change in the number of clusters depending on the budget. Block 4

Conclusion 2.3. *The dynamics of change in the number of objects in the obtained solutions is such that, other things being equal, it is increasing with* *increasing budget, regardless of the applied mathematical toolkit, selected topics, target parameters and months of the year (see Fig. 2.14, 2.15, 2.16, 2.17).*

This conclusion is logical, because with an increase in budget there is an opportunity to place the publication in a greater number of social network sites. However, when applying the methods of cluster analysis and the presence of communities in a given topic, which have a ratio of the unit of invested money to the quantitative indicators of audience feedback higher than others, we can get that when increasing the budget, from some of its values, the number of objects in the resulting solutions will be a non-increasing value. This is rare when using the optimization approach and is also a consequence of the procedure of selecting the best clusters among the resulting partitioning.

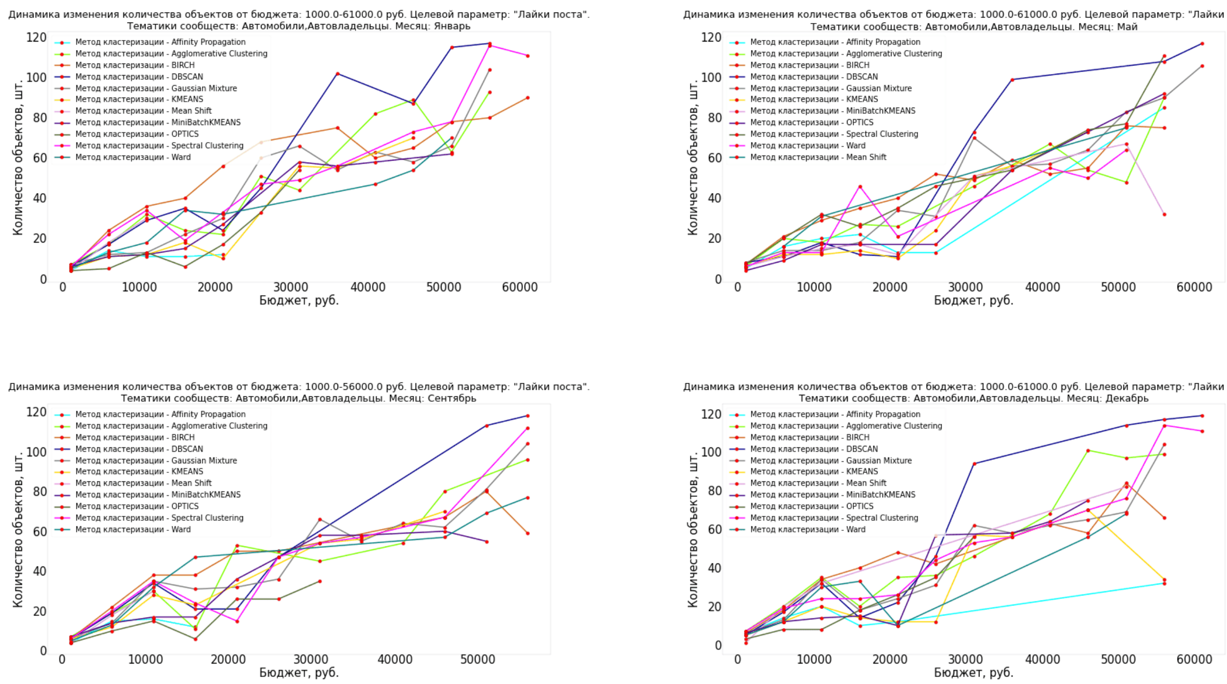


Figure 2.14: Dynamics of change in the number of objects depending on the budget. Block 1

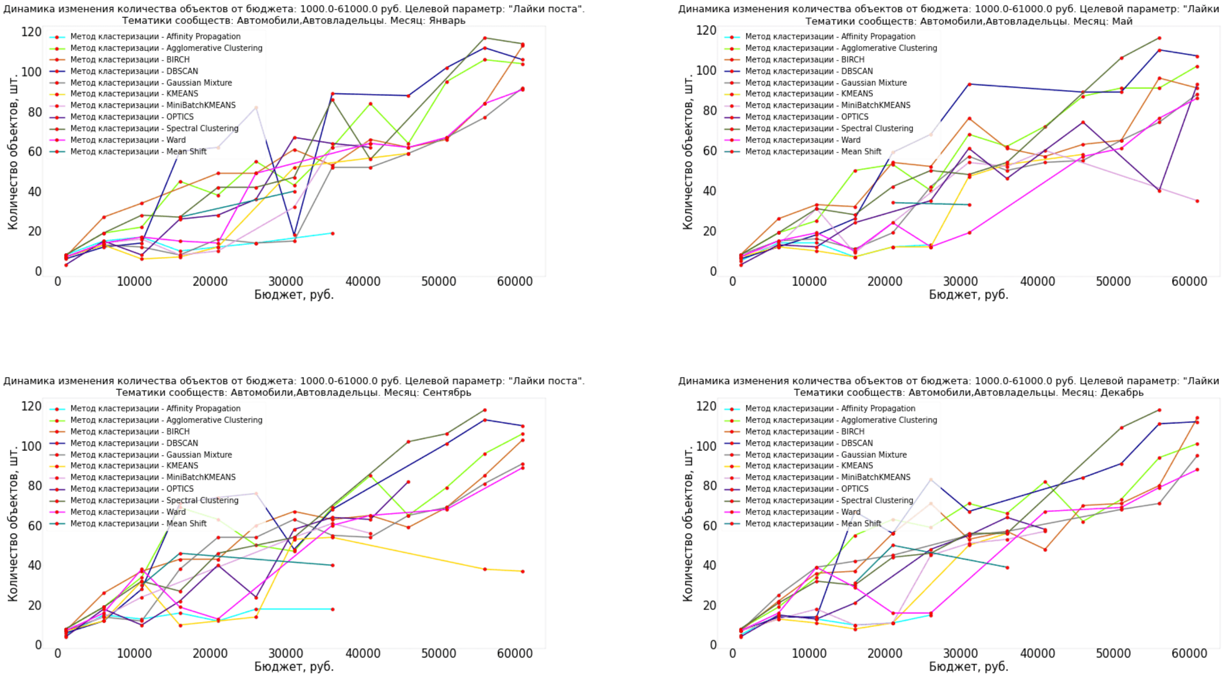


Figure 2.15: Dynamics of change in the number of objects depending on the budget. Block 2

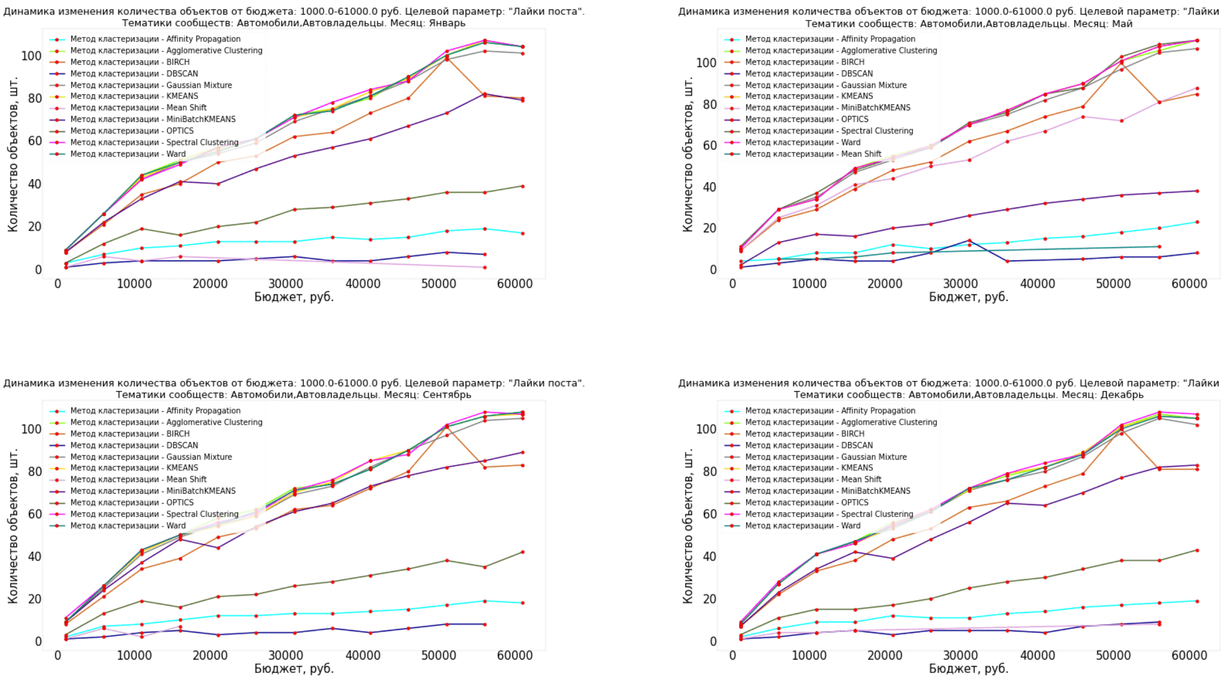


Figure 2.16: Dynamics of change in the number of objects depending on the budget. Block 3

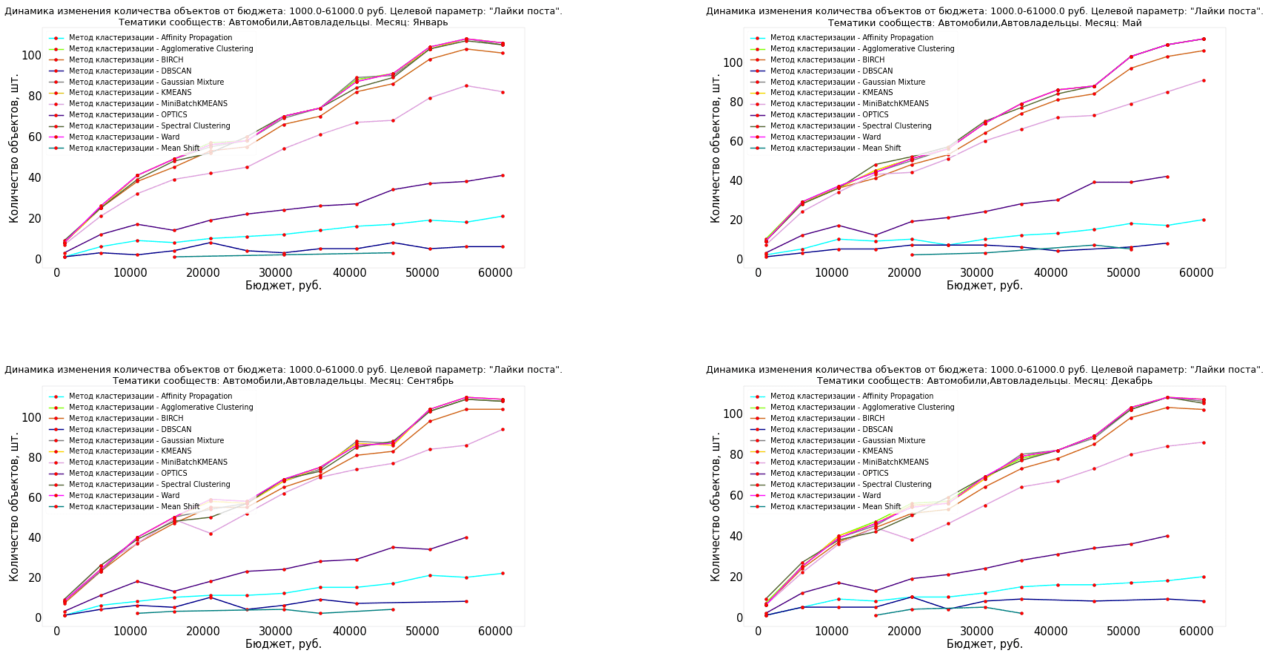


Figure 2.17: Dynamics of change in the number of objects depending on the budget. Block 4

In addition, the developed tool allows providing the client with the right to choose scenarios with different number of communities to publish an advertising record, and researchers to determine causal relationships between the obtained results and the application of appropriate cluster analysis methods.

Remark 2.1. *Analyzing the dynamics of changes in the maximum values of the target parameter, we can identify a number of clustering methods that give the best solution regardless of the applied mathematical toolkit, selected topics, target parameters, budget and months of the year (see Fig. 2.18, 2.19, 2.20, 2.21).*

This remark indicates that there is an opportunity to determine which methods of cluster analysis should be used to obtain the maximum values of the target parameter in different scenarios of information promotion, to reduce the time of their formation, as well as to formulate recommendations for the client on the feasibility of setting a certain value of the budget.

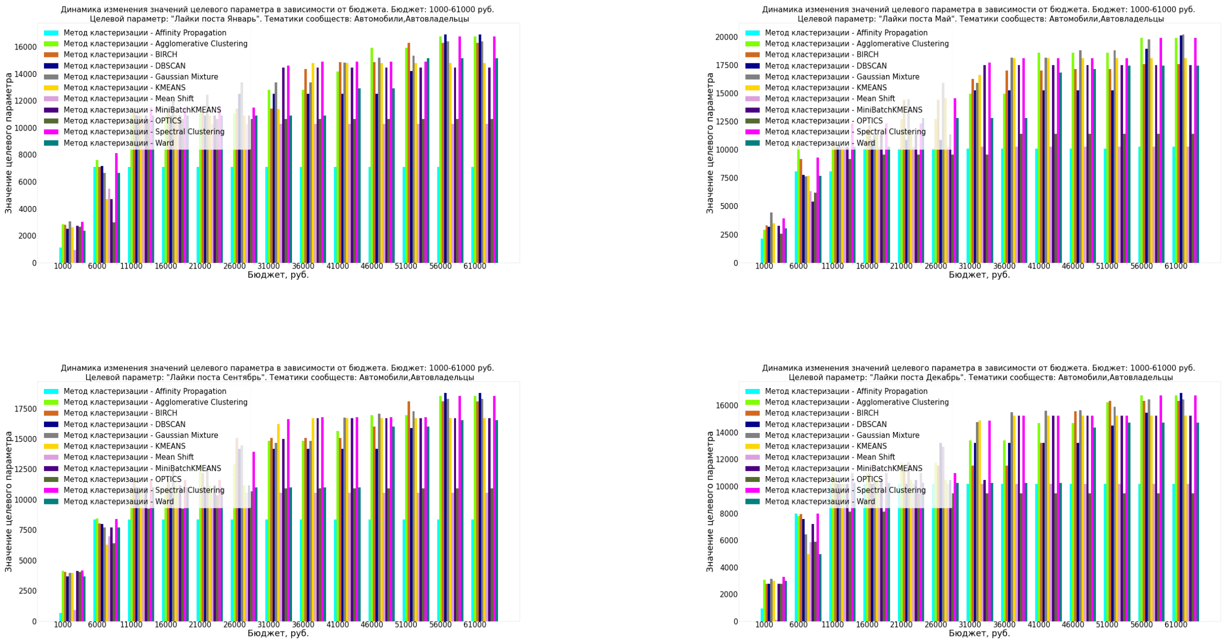


Figure 2.18: Dynamics of changes in the values of the target parameter depending on the budget. Block 1

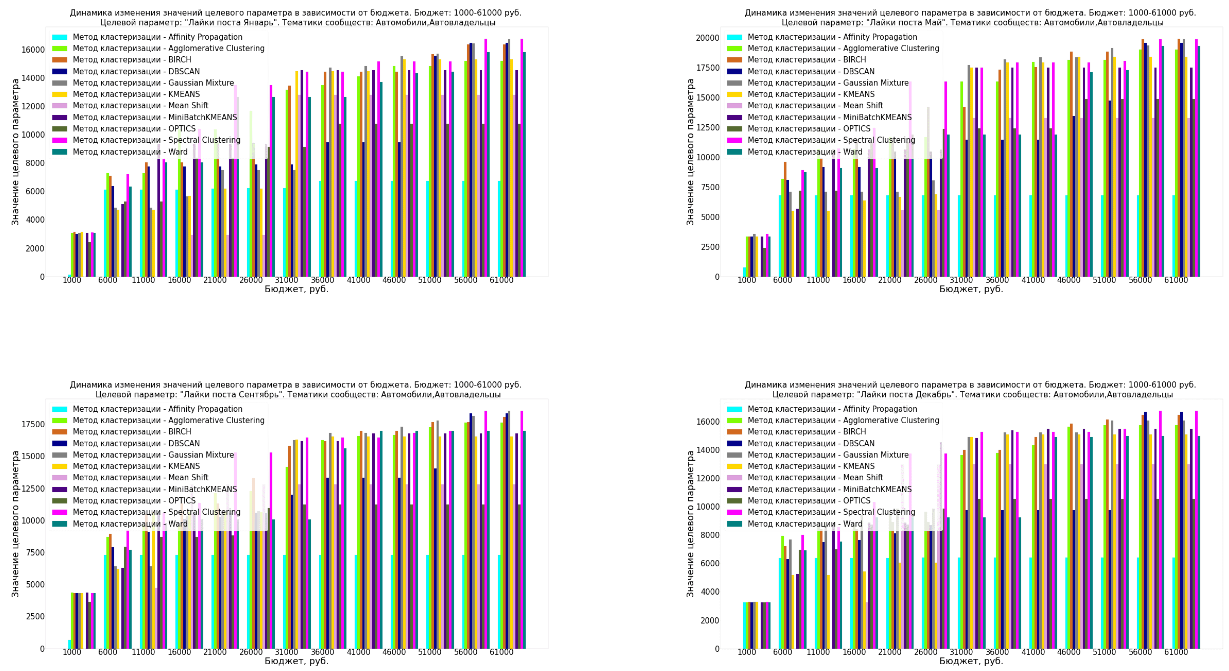


Figure 2.19: Dynamics of changes in the values of the target parameter depending on the budget. Block 2

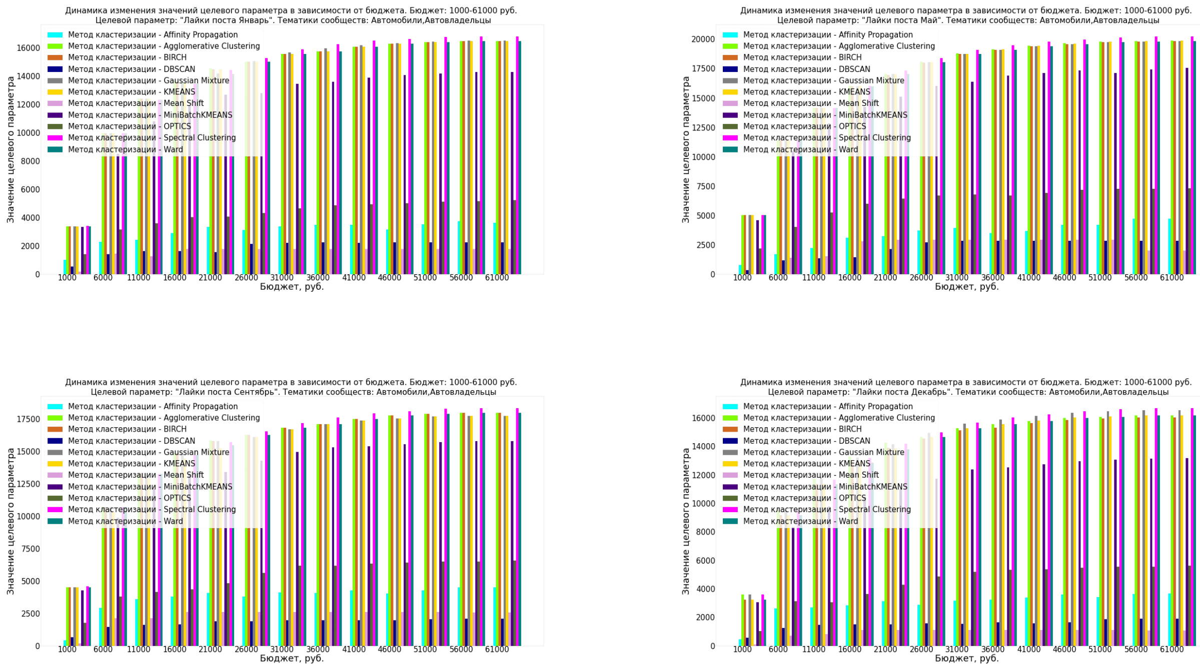


Figure 2.20: Dynamics of changes in the values of the target parameter depending on the budget. Block 3

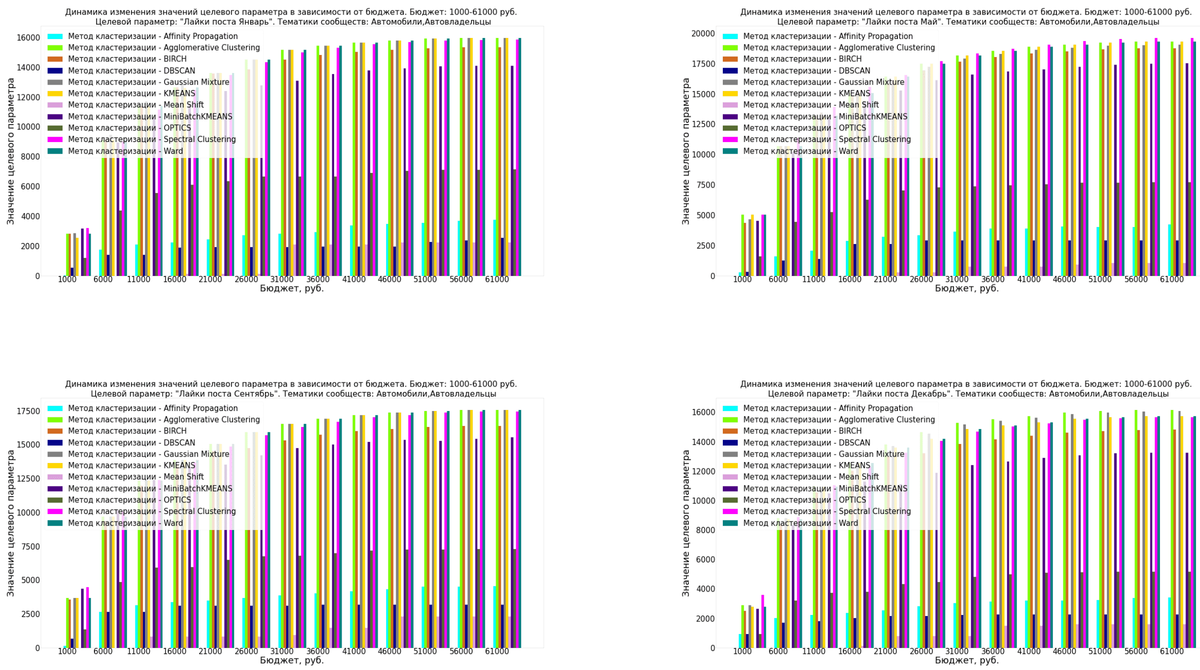
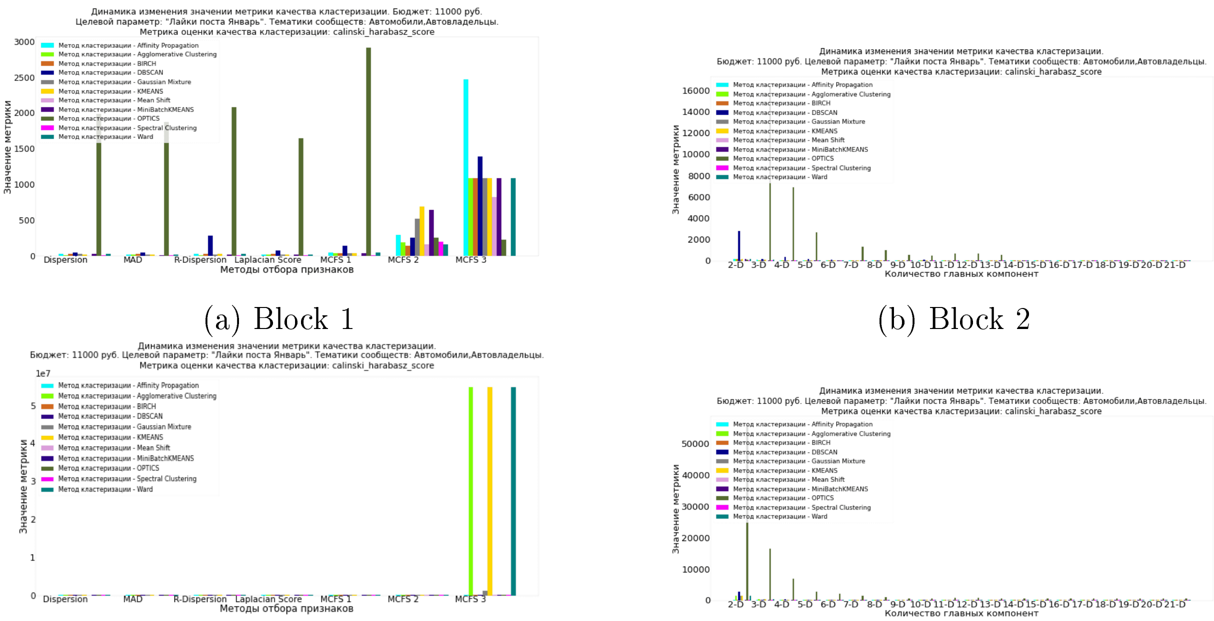


Figure 2.21: Dynamics of changes in the values of the target parameter depending on the budget. Block 4

Remark 2.2. *The dynamics of the values of the quality metrics shows that it is possible to select the best methods to reduce the dimensionality of the feature* *space for the corresponding clustering methods and input parameters (see Fig. 2.22, 2.23, 2.24).*

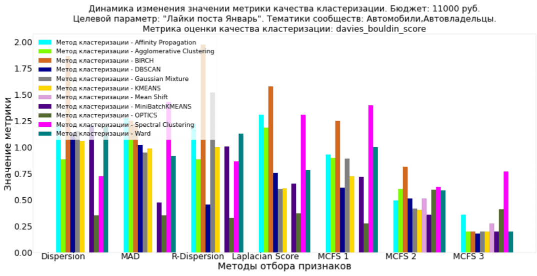
Since there are several ways to assess the quality of partitioning, this study has considered the main ones for the teacherless learning partition. It should be noted that the proposed approach for the solution of the set tasks will allow us to formulate conclusions about separability and density of the obtained partitions. In addition, it will be possible to determine in the first case, the best interpretable method of feature selection, in the second case, the necessary number of principal components for the used methods of cluster analysis, which may have a positive impact on reducing the time of obtaining different scenarios of information promotion in mass communication.



(a) Block 1 (b) Block 2

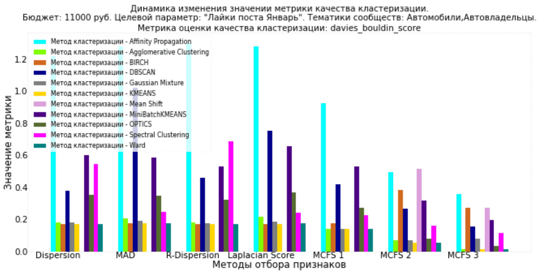
(c) Block 3 (d) Block 4

Figure 2.22: Dynamics of change in the values of the «Kalinski-Harabash» index



(a) Block 1

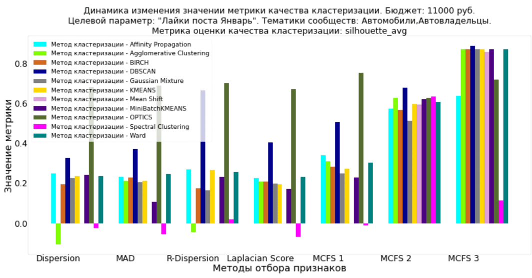
(b) Block 2



(c) Block 3

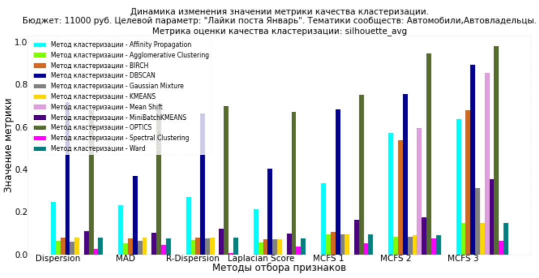
(d) Block 4.

Figure 2.23: Dynamics of change in the values of the «Davis-Bouldin» index



(a) Block 1

(b) Block 2



(c) Block 3

(d) Block 4

Figure 2.24: Dynamics of change in values of the «Silhouette» index

Remark 2.3. *The dynamics of changing the values of the target parameter depending on the budget shows that it is possible to determine the best clustering methods for appropriate ways of feature selection and input parameters (see Fig. 2.25, 2.26).*

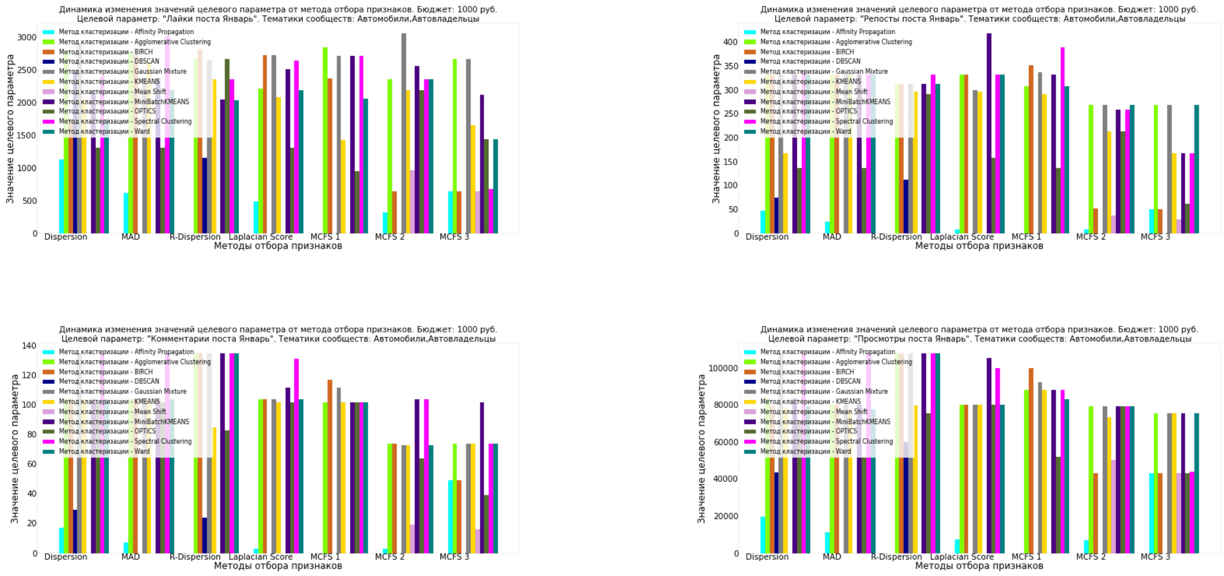


Figure 2.25: Dynamics of changes in the values of the target parameter depending on the budget. Block 1

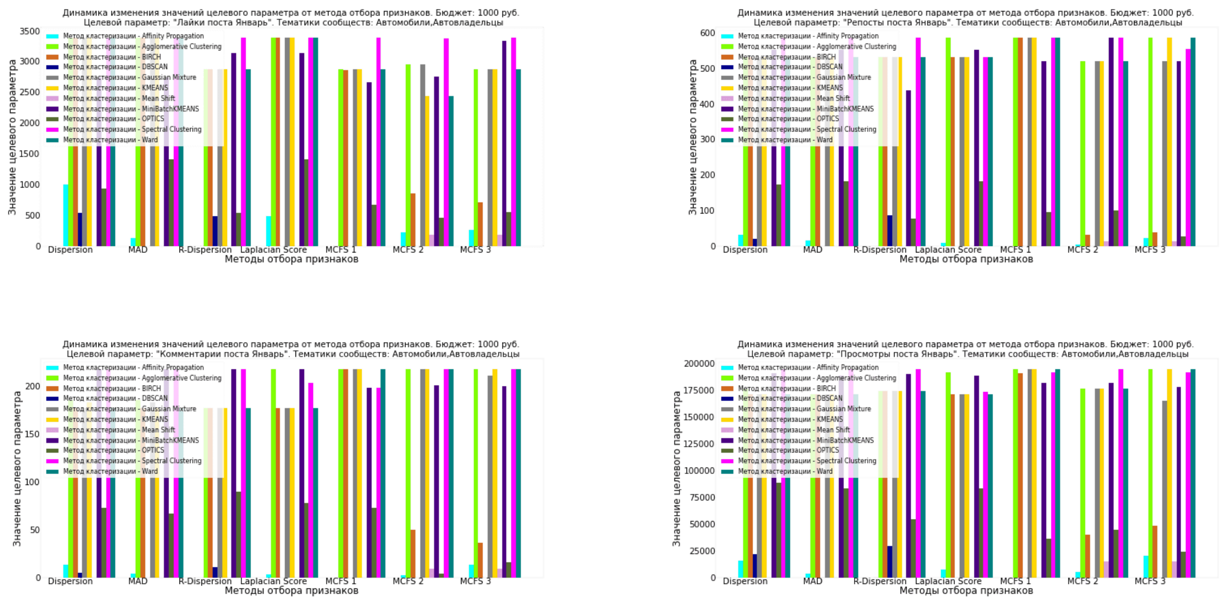


Figure 2.26: Dynamics of changes in the values of the target parameter depending on the budget. Block 3

Note that this remark will reduce the time for model training by sampling certain cluster analysis methods under different task conditions.

Remark 2.4. *The dynamics of changing the number of principal components depending on the budget is such that it allows us to limit the array of values for* *compressing the feature space with appropriate input parameters(see Fig. 2.27, 2.28).*

Each column in the above diagrams is the number of principal components, which corresponds to the best, in quantitative terms, numerical result for the given parameters and used methods of cluster analysis. Thus, the analysis of these visual representations of the simulation results will allow us to hypothesize the limitation of the values of the principal components to form a solution with some accuracy. In addition, it will allow to determine the best values for hyperparameters of clustering methods, as well as significantly reduce the simulation time.

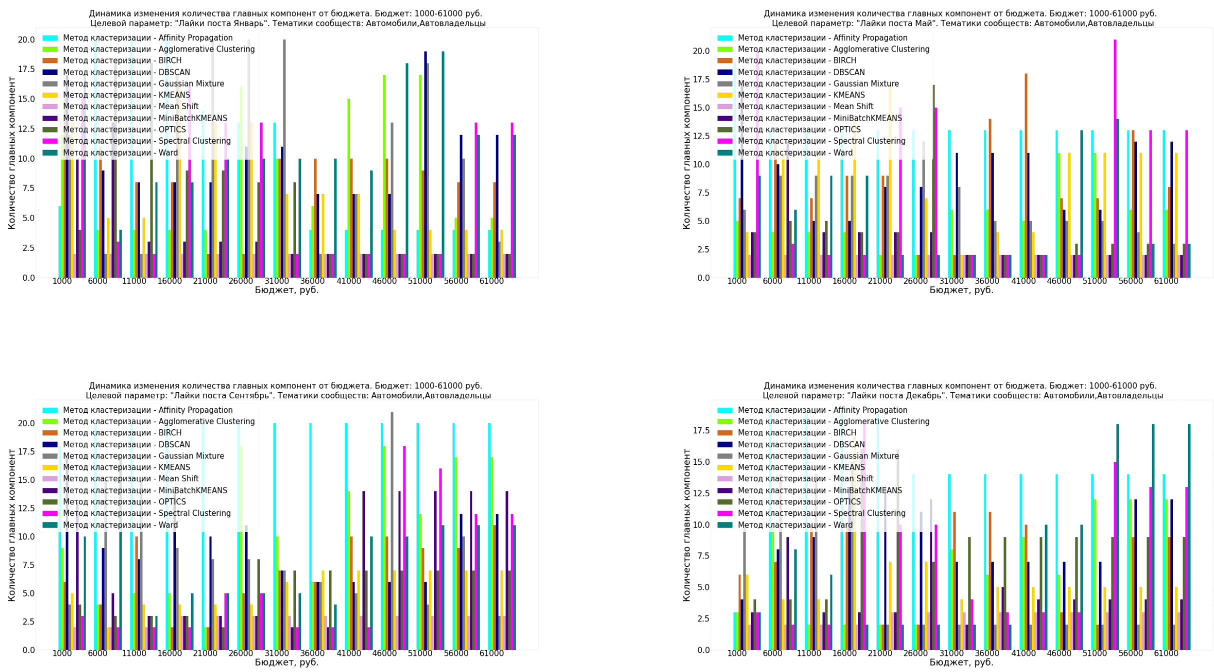


Figure 2.27: Evolution of the number of principal components from the budget. Block 2

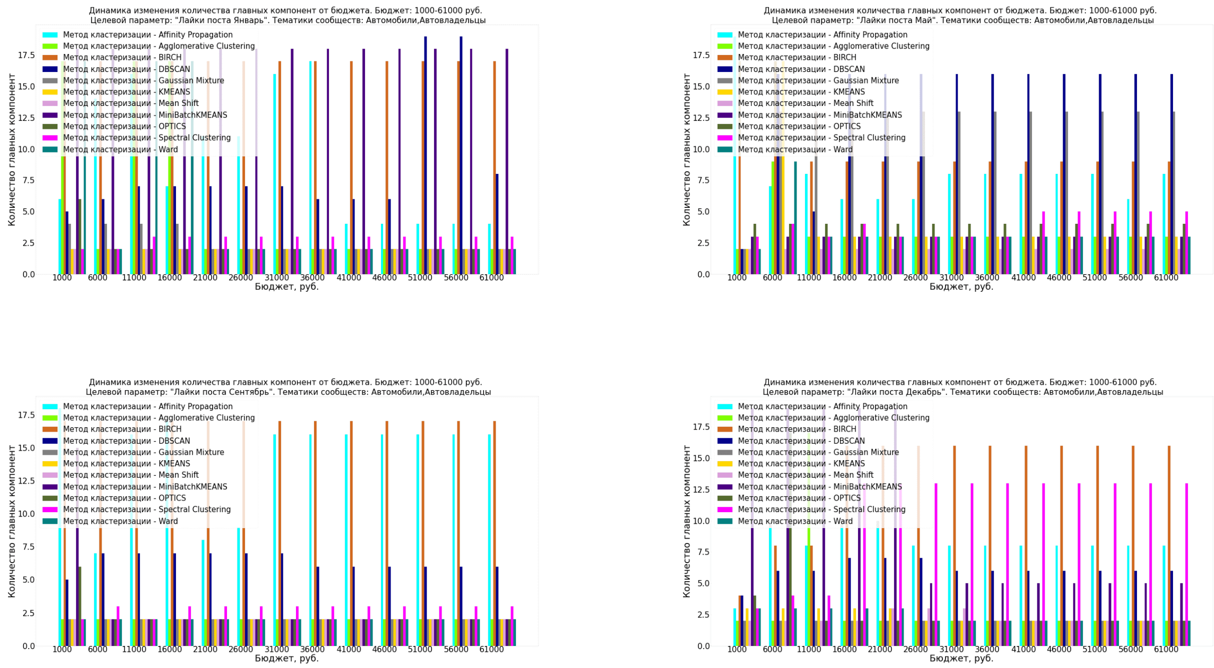


Figure 2.28: Evolution of the number of principal components from the budget. Block 4

2.4. Conclusions to the second chapter

The chapter formulates problem statements for modeling the process of information dissemination in MCM using machine learning and optimization methods with preliminary clustering to solve the problem of dimensionality reduction and time reduction of information dissemination scenarios formation. The architecture is proposed and a complex model with visualization is implemented programmatically, allowing to form a set of social network communities with recommendations on information placement in them. The methods of selection and feature extraction for the tasks of learning without a teacher are considered, the results of application of the mentioned methods are given. The methods of cluster analysis are considered and their hyperparameters are specified, and metrics for evaluating the quality of the obtained partitions are given. The software implementation of the complex model blocks is described. Training of clustering models and comparative analysis of modeling results are carried out, which showed that the application of cluster analysis methods allows solving the problem of dimensionality reduction in the optimization problem, which reduced the time of formation of information dissemination scenarios. The application of the developed software components allows to correct user preferences by adjusting hyperparameters of machine learning methods. The analysis of the importance of features allowed us to determine the basic set of significant characteristics of objects by selected methods of feature space compression.

Thus, the developed complex model with machine learning methods is a tool for scenario modeling, which allows to form various scenarios of information dissemination. The use of this complex model with machine learning methods will help management structures to be more flexible and adaptive when making decisions in the digital environment based on the modeled scenarios.

Chapter 3.
Intelligent scenario modeling system

3.1. Designing an intelligent system

Machine learning and artificial intelligence methods are being universally integrated and adapted for use in various application domains. Designing a management decision support system [2, 36, 67, 91] as a scenario modeling tool is a key step in creating an effective and reliable tool for business process management [16, 96, 104, 109]. Scenario modeling allows to evaluate possible variants of events development, the results of decisions made and analyze their impact on business processes.

When designing an intelligent management decision support system using scenario modeling, it is also important to take into account the specifics of business and industry [10, 88]. It is also necessary to determine the methods and tools of scenario modeling that will be used within the system. These can be mathematical models, analytical tools, statistical methods and other technologies that allow data analysis and numerical modeling. In developing this system, problem statements were formulated to define a set of information dissemination sites (Sections 1.1 and 2.1). Note that optimization methods as well as teacherless machine learning methods were used to analyze big data to form different scenarios of information promotion.

An important aspect of intelligent system design is also to ensure its reliability, security and scalability. The system should be flexible and adaptive to changing market conditions and business environment to ensure high efficiency of management decisions. As mentioned earlier, the architecture of the proposed solution is such that it satisfies the above aspects. Figure 3.1 below shows the integration scheme of the developed intelligent system for modeling the information dissemination process on the example of social networks.

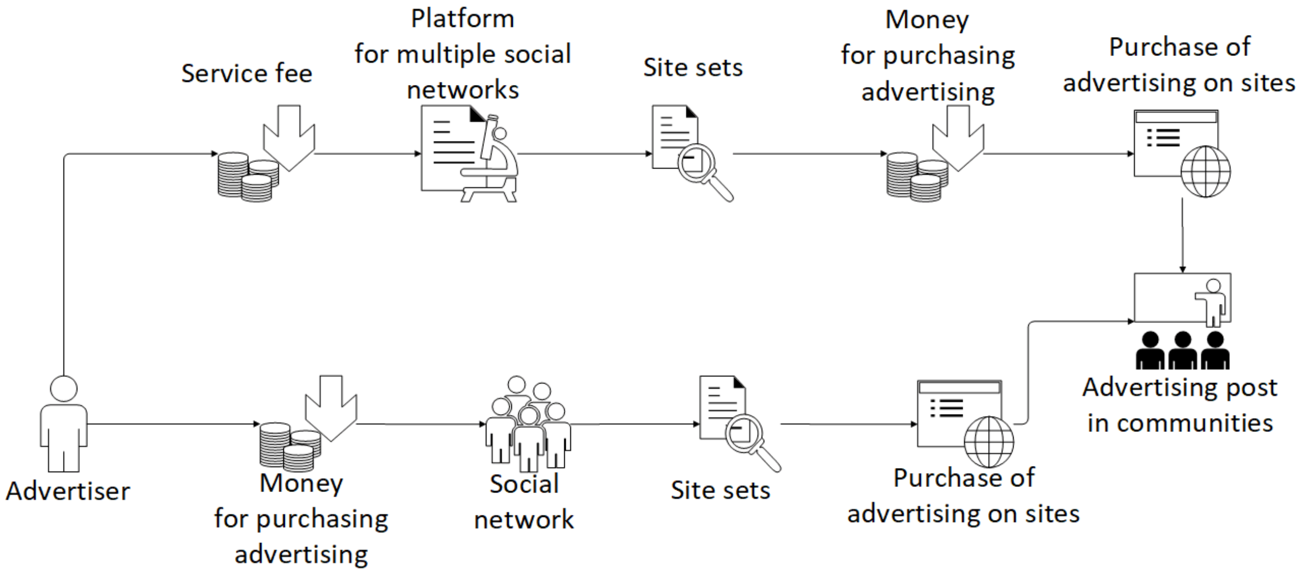


Figure 3.1: Scheme of integration of the intelligent system

In this case, in terms of feasibility of cooperation and possible benefits for each party, the proposed approach is mutually beneficial. As the quality of the provided service increases, the term of advertising campaigns is reduced, due to the implementation of big data analysis and the construction of recommendations for posting entries in the relevant communities. It should be noted that advertisers receive a service of proper quality with the involvement of fewer resources, and social networks receive an additional inflow of working capital. Implementation of such a tool is possible both on a cross-platform basis and for a specific social network, it will be beneficial in both the first and the second case. The difference between these options for the realization of intelligent system is in the functionality and where the advertiser will address directly to the tools of the selected social network or to the platform providing a similar service for several Internet sites.

In addition, at the current moment of time in the considered social network there is no choice of different scenarios of information promotion, for the invested money is offered one set of platforms for publishing an advertising record.

The problem of scenario selection for DM is the need to evaluate and analyze different scenarios in order to make an informed and effective decision. Also, DM are faced with uncertainty, complexity and a variety of factors that can affect the outcome of a decision. Choosing the best scenario requires a thorough understanding of business processes, analyzing data, predicting outcomes, and considering risks. The wrong choice of scenario can lead to undesirable consequences, loss of time, resources and loss of competitiveness of the company.

For successful management decision-making, it is necessary to use scenario modeling tools, conduct data analysis and take into account possible changes in the external environment. It is important to have such a decision support tool that will help the DM in choosing the optimal scenario and minimizing risks. It should also be noted that the DM will be able to evaluate different scenarios of information promotion and their impact on the key performance indicators of an advertising campaign on the Internet before the purchase of advertising and publication of the record in social network communities, which will allow for more accurate planning and budget allocation when conducting advertising campaigns.

The developed intellectual system as a tool of scenario modeling due to its architecture allows solving the set tasks, can be applied and adapted in different subject areas, is scalable and, due to its block structure, can be easily changed in accordance with the requirements for the development of its functionality. The architecture of the system can be familiarized by studying fig. 3.2.

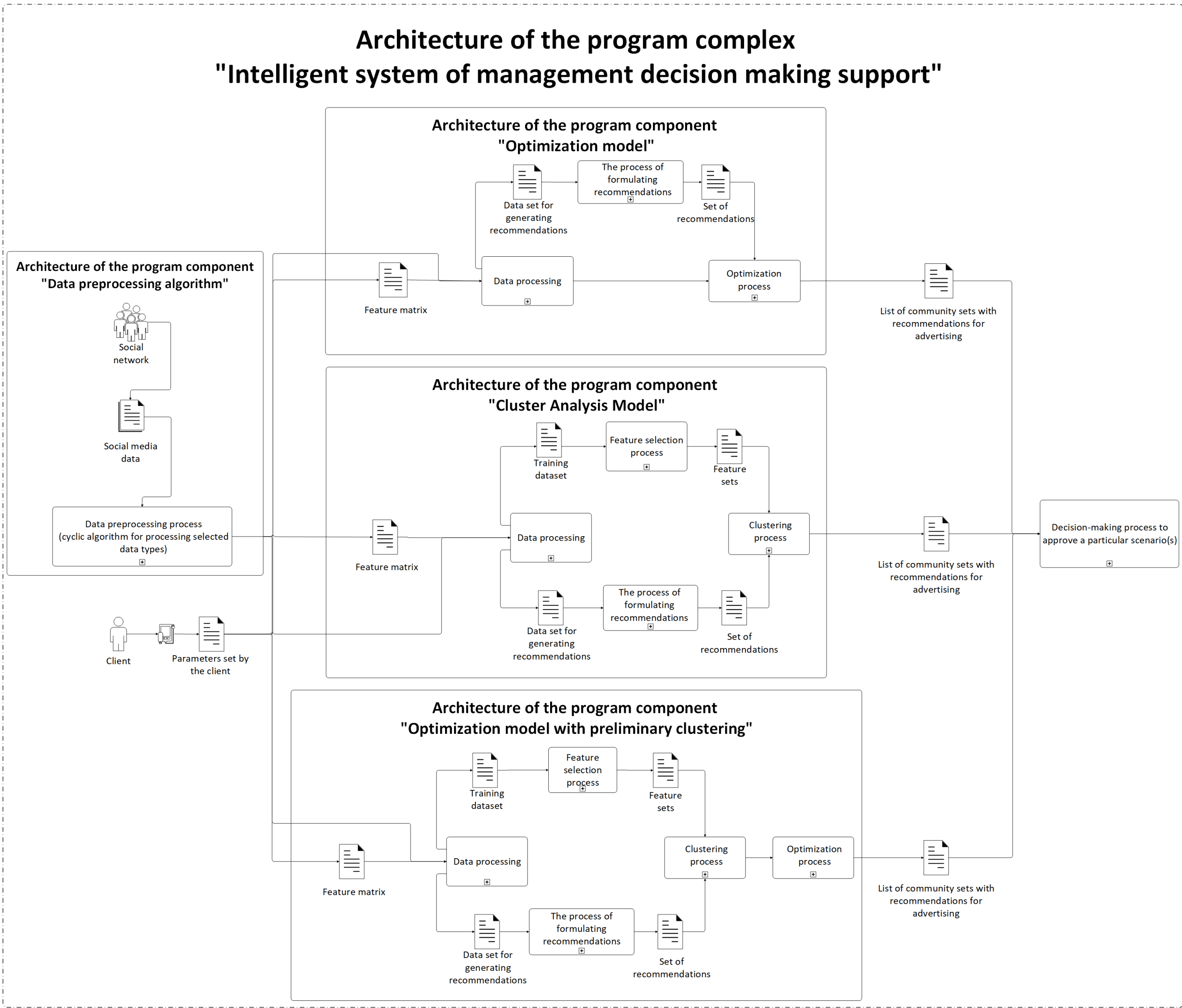


Figure 3.2: Architecture of software complex «Intelligent system of management decision support system» as a tool of scenario modeling

In addition, it should be noted that when developing a prototype of an intelligent system for the DM, machine learning, optimization and data analytics methods are used to process information and provide users-clients with relevant and reliable information, as well as form recommendations for making optimal strategic and operational decisions.

3.2. Realization and peculiarities of application of the intelligent system

The implementation of an intelligent management decision support system starts with the definition of business goals and tasks to be solved with its help. Then the data required for the system to function is collected and analyzed. After that, machine learning algorithms are configured and trained to work with specific data and tasks, as well as sensitivity analysis for determining advertiser preferences in a multi-criteria optimization problem. The components of the system are described in detail in the previous chapters and sections.

The developed Intelligent Management Decision Support System (IMDSS) is an innovative tool that combines the capabilities of artificial intelligence, data analytics and business processes to assist executives and managers in making informed and effective decisions. Features of the IMDSS application include:

- *Automation of the decision-making process:* IMDSS provides an opportunity to automate a part of the management decision-making process based on data analysis, scenario modeling (forecasting of results). This reduces the time required to make decisions and reduces the likelihood of errors.

- *Analyzing large amounts of data:* IMDSS is able to process and analyze large amounts of data from various sources, revealing hidden patterns and dependencies. This helps managers to make informed decisions based on facts and analytics.

- *Personalizing recommendations:* IMDSS is able to create personalized recommendations and suggestions based on each manager's individual needs and goals. This allows you to take into account the unique characteristics of your business and make decisions that are appropriate to your specific circumstances.

- *Monitoring and evaluation of results:* IMDSS provides the ability to monitor and evaluate the results of decisions made, as well as analyze their effectiveness. This allows to adjust action strategies in real time and optimize business processes.

- *Integration with other systems:* IMDSS can be easily integrated with other organizational information systems, such as CRM, ERP and BI systems, providing unified access to data and increasing the efficiency of the entire company.

- *Scalability:* the system can be easily scaled to handle different data volumes and tasks, making it a versatile tool for different organizations and industries.

Thus, the application of this IMDSS allows organizations to improve the quality of decision-making, optimize business processes and achieve a competitive advantage in the market when conducting information dissemination activities on the Internet, as well as allows advertising organizations to improve management efficiency, minimize risks and errors, and make informed strategic and operational decisions based on data and analytics.

The developed prototype of the intelligent system consists of 6 software components, each of which can work autonomously and provide scenarios to promote information [34, 45, 52]. Based on the client's needs, the necessary number of program components from the whole program complex will be used to obtain a solution in a short period of time. Let us consider in more detail the scheme of realization and functioning of the prototype system.

It should be noted that to ensure the stability of the system operation at external failures it was decided to implement a program complex consisting of 6 program blocks:

1. Program realization of cyclic algorithm of preprocessing of statistical data on user activity of information sites in the task of information dissemination in MCM. It is realized as a single file with the extension «.ipynb» for the convenience of making changes when writing the program code.

2. Software implementation of the optimization model, including the solution of the multicriteria optimization problem. The model was implemented in such a way that it consists of 6 files: 2 of them with extension «.py», which are used as libraries for data processing and forming recommendations; 1 with extension «.ipynb» and is used to conduct numerical simulation; the remaining 3 files with extension «.ipynb» are needed for visualization, analysis and interpretation of results.

3. Program implementation of cluster analysis methods using interpreted feature

selection methods. The block consists of 15 files: 11 files with «.py» extension used as libraries and implementing clustering methods, 2 files with «.py» extension used as libraries for data processing and forming recommendations, 1 file with «.ipynb» extension for numerical simulation, 1 file with «.ipynb» extension for visualization, analysis and interpretation of results.

4. Program implementation of cluster analysis methods with application of the method of principal components to compress the feature space. The structure is similar to item 3.

5. Software implementation of the optimization model with preliminary clustering and using interpreted feature selection methods. The structure is similar to item 3.

6. Software implementation of the optimization model with preliminary clustering and using the method of principal components to compress the feature space. The structure is similar to item 3.

Next, let's consider the developed structure of modeling results storage. The development of such structures is an important step in the process of working with data and analytics. This system allows efficient storage, management and processing of modeling results, which has a number of advantages:

- *Saving results:* the proposed storage structure allows saving all modeling results in a structured and easy-to-access form. This provides the possibility of reusing the results, analyzing and comparing different models.

- *Ease of access and sharing:* the proposed storage structure allows easy access to modeling results, sharing them with colleagues and other project participants. This facilitates collaborative work and knowledge sharing.

- *Improved decision-making process:* access to saved modeling results helps you make informed decisions based on data and analytics. Analyzing previous results can help identify trends and optimal action strategies.

- *Improving operational efficiency:* the proposed structure for storing modeling results simplifies data management processes, reduces the time required to find the necessary information, and increases the overall efficiency of the team.

Thus, designing a storage structure for modeling results plays a key role in the successful handling of data and analytics, providing security, ease of access to information, and improving the efficiency of management decision-making processes.

Based on the above, it was decided to save data by blocks in «.csv» and «.xlsx» formats. So, for example, in block 2 data are stored by relevant topics, which have subfolders by months of the year, where 2 more folders were created for researchers and for clients. The latter are distinguished by a set of columns. And if we consider blocks 3-6, in these cases there will be much more different information that will have to be stored and processed.

Let's take the block with the number 5, there is a folder named «Simulation Results», where the output summary tabular data for the corresponding topics are stored, each one has the following storage structure as shown in the figure 3.3.

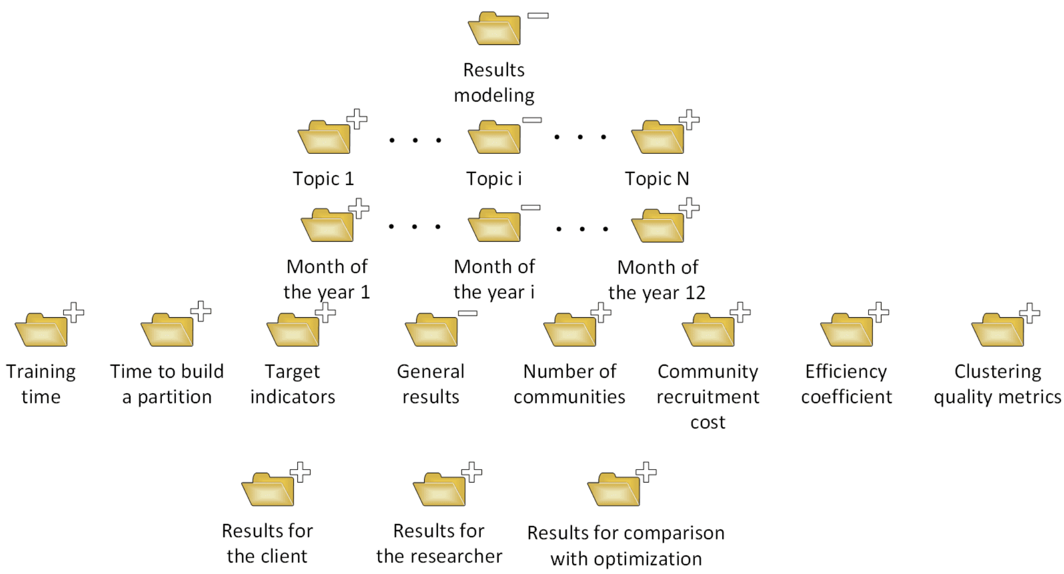


Figure 3.3: Scheme of data storage structure

Each folder located in «Month of the year i» has 11 files of format either «.csv» or «.xlsx», the number of which is equal to the number of applied methods of cluster analysis. Note that the «General Results» folder is divided into 3 more sections, which differ from each other by the set of columns in the files. Examples of the file structure of each folder are shown in the corresponding figures 3.4, 3.5, 3.6.

	A
1	,Dispersion,MAD,R-Dispersion,Laplacian Score,MCFS 1,MCFS 2,MCFS 3,Метод кластеризации,Бюджет клиента,Выбранные тематики,Целевой параметр,Месяц
2	0,0 days 00:00:01.260295,0 days 00:00:01.319103,0 days 00:00:01.241291,0 days 00:00:01.292892,0 days 00:00:01.274525,0 days 00:00:01.255446,0 days 00:00:01.266508,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Лайки поста Январь,Январь
3	1,0 days 00:00:01.260295,0 days 00:00:01.319103,0 days 00:00:01.241291,0 days 00:00:01.292892,0 days 00:00:01.274525,0 days 00:00:01.255446,0 days 00:00:01.266508,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Репосты поста Январь,Январь
4	2,0 days 00:00:01.260295,0 days 00:00:01.319103,0 days 00:00:01.241291,0 days 00:00:01.292892,0 days 00:00:01.274525,0 days 00:00:01.255446,0 days 00:00:01.266508,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Комментарии поста Январь,Январь
5	3,0 days 00:00:01.260295,0 days 00:00:01.319103,0 days 00:00:01.241291,0 days 00:00:01.292892,0 days 00:00:01.274525,0 days 00:00:01.255446,0 days 00:00:01.266508,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Просмотры поста Январь,Январь
6	4,0 days 00:00:01.272931,0 days 00:00:01.327778,0 days 00:00:01.245443,0 days 00:00:01.292303,0 days 00:00:01.275406,0 days 00:00:01.254725,0 days 00:00:01.270597,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Лайки поста Январь,Январь
7	5,0 days 00:00:01.272931,0 days 00:00:01.327778,0 days 00:00:01.245443,0 days 00:00:01.292303,0 days 00:00:01.275406,0 days 00:00:01.254725,0 days 00:00:01.270597,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Репосты поста Январь,Январь
8	6,0 days 00:00:01.272931,0 days 00:00:01.327778,0 days 00:00:01.245443,0 days 00:00:01.292303,0 days 00:00:01.275406,0 days 00:00:01.254725,0 days 00:00:01.270597,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Комментарии поста Январь,Январь
9	7,0 days 00:00:01.272931,0 days 00:00:01.327778,0 days 00:00:01.245443,0 days 00:00:01.292303,0 days 00:00:01.275406,0 days 00:00:01.254725,0 days 00:00:01.270597,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Просмотры поста Январь,Январь

(a) Program component 3

	A
1	,2-D,3-D,4-D,5-D,6-D,7-D,8-D,9-D,10-D,11-D,12-D,13-D,14-D,15-D,16-D,17-D,18-D,19-D,20-D,21-D,Метод кластеризации,Бюджет клиента,Выбранные тематики,Целевой параметр,Месяц
2	0,0 days 00:00:03.698868,0 days 00:00:03.629852,0 days 00:00:03.939195,0 days 00:00:03.513331,0 days 00:00:03.690824,0 days 00:00:03.772574,0 days 00:00:03.769776,0 days 00:00:03.814897,0 days
3	1,0 days 00:00:03.698868,0 days 00:00:03.629852,0 days 00:00:03.939195,0 days 00:00:03.513331,0 days 00:00:03.690824,0 days 00:00:03.772574,0 days 00:00:03.769776,0 days 00:00:03.814897,0 days
4	2,0 days 00:00:03.698868,0 days 00:00:03.629852,0 days 00:00:03.939195,0 days 00:00:03.513331,0 days 00:00:03.690824,0 days 00:00:03.772574,0 days 00:00:03.769776,0 days 00:00:03.814897,0 days
5	3,0 days 00:00:03.698868,0 days 00:00:03.629852,0 days 00:00:03.939195,0 days 00:00:03.513331,0 days 00:00:03.690824,0 days 00:00:03.772574,0 days 00:00:03.769776,0 days 00:00:03.814897,0 days
6	4,0 days 00:00:03.871909,0 days 00:00:03.755882,0 days 00:00:03.967932,0 days 00:00:03.684865,0 days 00:00:03.886913,0 days 00:00:03.827364,0 days 00:00:03.759791,0 days 00:00:03.821898,0 days
7	5,0 days 00:00:03.871909,0 days 00:00:03.755882,0 days 00:00:03.967932,0 days 00:00:03.684865,0 days 00:00:03.886913,0 days 00:00:03.827364,0 days 00:00:03.759791,0 days 00:00:03.821898,0 days
8	6,0 days 00:00:03.871909,0 days 00:00:03.755882,0 days 00:00:03.967932,0 days 00:00:03.684865,0 days 00:00:03.886913,0 days 00:00:03.827364,0 days 00:00:03.759791,0 days 00:00:03.821898,0 days
9	7,0 days 00:00:03.871909,0 days 00:00:03.755882,0 days 00:00:03.967932,0 days 00:00:03.684865,0 days 00:00:03.886913,0 days 00:00:03.827364,0 days 00:00:03.759791,0 days 00:00:03.821898,0 days

(b) Program component 4

Figure 3.4: Training time

	A
1	,Dispersion,MAD,R-Dispersion,Laplacian Score,MCFS 1,MCFS 2,MCFS 3,Метод кластеризации,Бюджет клиента,Выбранные тематики,Целевой параметр,Месяц
2	0,0 days 00:00:00.056014,0 days 00:00:00.075042,0 days 00:00:00.046011,0 days 00:00:00.051090,0 days 00:00:00.041510,0 days 00:00:00.053013,0 days 00:00:00.054012,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Лайки поста Январь,Январь
3	1,0 days 00:00:00.063015,0 days 00:00:00.063015,0 days 00:00:00.055013,0 days 00:00:00.045011,0 days 00:00:00.064016,0 days 00:00:00.071018,0 days 00:00:00.054014,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Репосты поста Январь,Январь
4	2,0 days 00:00:00.065016,0 days 00:00:00.063015,0 days 00:00:00.055013,0 days 00:00:00.045011,0 days 00:00:00.064016,0 days 00:00:00.071018,0 days 00:00:00.054012,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Комментарии поста Январь,Январь
5	3,0 days 00:00:00.045011,0 days 00:00:00.064015,0 days 00:00:00.055013,0 days 00:00:00.045011,0 days 00:00:00.061015,0 days 00:00:00.039010,0 days 00:00:00.054014,MiniBatchKMEANS,1000,"Автомобили,Автовладельцы",Просмотры поста Январь,Январь
6	4,0 days 00:00:00.041011,0 days 00:00:00.052013,0 days 00:00:00.051727,0 days 00:00:00.048012,0 days 00:00:00.064015,0 days 00:00:00.054014,0 days 00:00:00.033009,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Лайки поста Январь,Январь
7	5,0 days 00:00:00.041011,0 days 00:00:00.052013,0 days 00:00:00.040010,0 days 00:00:00.052014,0 days 00:00:00.064015,0 days 00:00:00.054014,0 days 00:00:00.033009,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Репосты поста Январь,Январь
8	6,0 days 00:00:00.074019,0 days 00:00:00.068017,0 days 00:00:00.053013,0 days 00:00:00.052014,0 days 00:00:00.064015,0 days 00:00:00.054014,0 days 00:00:00.033009,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Комментарии поста Январь,Январь
9	7,0 days 00:00:00.074019,0 days 00:00:00.052013,0 days 00:00:00.040010,0 days 00:00:00.052014,0 days 00:00:00.064015,0 days 00:00:00.054014,0 days 00:00:00.033009,MiniBatchKMEANS,6000,"Автомобили,Автовладельцы",Просмотры поста Январь,Январь

(a) Program componentт 3

	A
1	,2-D,3-D,4-D,5-D,6-D,7-D,8-D,9-D,10-D,11-D,12-D,13-D,14-D,15-D,16-D,17-D,18-D,19-D,20-D,21-D,Метод кластеризации,Бюджет клиента,Выбранные тематики,Целевой параметр,Месяц
2	0,0 days 00:00:00.065016,0 days 00:00:00.047011,0 days 00:00:00.061014,0 days 00:00:00.048011,0 days 00:00:00.063015,0 days 00:00:00.049012,0 days 00:00:00.078019,0 days 00:00:00.054013,0 days 00:00:00.049013,0 days 00:00:00.064015,
3	1,0 days 00:00:00.058015,0 days 00:00:00.053012,0 days 00:00:00.049013,0 days 00:00:00.047012,0 days 00:00:00.055013,0 days 00:00:00.045011,0 days 00:00:00.051012,0 days 00:00:00.041010,0 days 00:00:00.035009,0 days 00:00:00.045011,
4	2,0 days 00:00:00.042010,0 days 00:00:00.047011,0 days 00:00:00.061014,0 days 00:00:00.046011,0 days 00:00:00.055013,0 days 00:00:00.045011,0 days 00:00:00.051012,0 days 00:00:00.041010,0 days 00:00:00.036008,0 days 00:00:00.076019,
5	3,0 days 00:00:00.042010,0 days 00:00:00.053012,0 days 00:00:00.049013,0 days 00:00:00.047012,0 days 00:00:00.055013,0 days 00:00:00.045011,0 days 00:00:00.051012,0 days 00:00:00.041010,0 days 00:00:00.035009,0 days 00:00:00.052012,
6	4,0 days 00:00:00.034009,0 days 00:00:00.039009,0 days 00:00:00.063015,0 days 00:00:00.058014,0 days 00:00:00.035008,0 days 00:00:00.046012,0 days 00:00:00.042010,0 days 00:00:00.040010,0 days 00:00:00.036009,0 days 00:00:00.040009,
7	5,0 days 00:00:00.093022,0 days 00:00:00.040009,0 days 00:00:00.052013,0 days 00:00:00.058014,0 days 00:00:00.035008,0 days 00:00:00.061014,0 days 00:00:00.077018,0 days 00:00:00.040009,0 days 00:00:00.061014,0 days 00:00:00.052012,
8	6,0 days 00:00:00.093022,0 days 00:00:00.064015,0 days 00:00:00.052013,0 days 00:00:00.048012,0 days 00:00:00.035008,0 days 00:00:00.061014,0 days 00:00:00.077018,0 days 00:00:00.040009,0 days 00:00:00.061014,0 days 00:00:00.052012,
9	7,0 days 00:00:00.093022,0 days 00:00:00.064015,0 days 00:00:00.048011,0 days 00:00:00.048012,0 days 00:00:00.062015,0 days 00:00:00.061014,0 days 00:00:00.077018,0 days 00:00:00.040009,0 days 00:00:00.061014,0 days 00:00:00.052012,

(b) Program component 4

Figure 3.5: Time to build a partition

	Dispersion	MAD	R-Dispersion	Laplacian Score	MCFS 1	MCFS 2	MCFS 3	Метод кластеризации	Бюджет клиента	Выбранные тематики	Целевой параметр	Месяц
0	2188	2377	2045	2519	2723	2569	2118	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
1	333	333	312	419	333	258	167	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
2	104	104	135	112	102	104	102	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
3	83573	79861	107877	105276	88003	79234	75502	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь
4	4338	3005	3239	3376	3260	4618	4708	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
5	605	769	678	620	746	1030	929	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
6	251	251	251	325	362	409	397	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
7	230149	257851	250716	359934	415120	378971	351877	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь

(a) Program component 3

	2-D	3-D	4-D	5-D	6-D	7-D	8-D	9-D	10-D	11-D	12-D	13-D	14-D	15-D	16-D	17-D	18-D	19-D	20-D	21-D	Метод кластеризации	Бюджет клиента	Выбранные тематики	Целевой параметр	Месяц
0	1753	1753	2403	3007	2403	2546	3007	3007	3067	3007	3007	3007	3007	3007	3007	3067	3007	2546	3007	3007	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
1	333	294	294	301	426	265	265	390	260	258	344	252	333	169	419	419	333	344	344	344	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
2	171	102	119	119	158	150	150	202	113	113	113	102	102	102	112	112	102	102	102	102	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
3	100876	79861	59875	79999	125400	85685	85685	131086	83277	71008	111005	97693	80698	80698	105276	105276	80698	111005	111005	111005	MiniBatchKMEANS	1000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь
4	4166	4885	4194	4017	3888	4069	4339	4708	4069	4339	4645	5106	5106	3518	4655	4516	4516	4516	4516	4516	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
5	575	596	765	818	489	448	767	890	767	448	767	767	890	928	890	890	584	890	718	718	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
6	280	292	330	296	259	251	301	330	301	251	301	301	330	330	330	330	251	330	329	329	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
7	259814	329856	329856	313967	230149	230149	358864	421727	358864	230149	358864	358864	421727	421727	421727	421727	230149	421727	385042	385042	MiniBatchKMEANS	6000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь

(b) Program component 4

Figure 3.6: Target indicators

The remaining file types are similar in structure to the files presented in the figures above, except for those stored in the «Clustering quality metrics» folder. The distinctive feature is that all three metrics are recorded for each target indicator (see Fig. 3.7).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Dispersion	MAD	R-Dispersion	Laplacian Score	MCFS 1	MCFS 2	MCFS 3	Метод кластеризации	Целевой параметр	Метрика качества кластеризации	Бюджет клиента	Выбранные тематики	Месяц
2	0	0,85561824	1,06851219	0,61374523	1,178847295	1,06210917	0,39392585	0,23759553	MiniBatchKMEANS	Лайки поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
3	1	14,9887812	17,3991378	18,6982348	16,24810323	40,5909114	864,839065	3180,44508	MiniBatchKMEANS	Лайки поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
4	2	0,15615398	0,21416506	0,199733142	0,174943928	0,27589233	0,32546819	0,39101134	MiniBatchKMEANS	Лайки поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
5	3	0,53575756	0,58670434	0,5297623	0,884737492	0,45494039	0,25991196	0,26537787	MiniBatchKMEANS	Репосты поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
6	4	13,0468405	11,8331911	14,66170944	15,06577038	25,7381193	792,193079	96,5748055	MiniBatchKMEANS	Репосты поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
7	5	0,07978887	0,09955982	0,120026873	0,20633215	0,11137279	0,18987478	0,30277678	MiniBatchKMEANS	Репосты поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
8	6	0,60003009	0,58670434	0,5297623	0,884737492	0,45494039	0,25991196	0,23759553	MiniBatchKMEANS	Комментарии поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
9	7	11,9256032	11,8331911	14,66170944	15,06577038	25,7381193	792,193079	3180,44508	MiniBatchKMEANS	Комментарии поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
10	8	0,10899229	0,09955982	0,120026873	0,20633215	0,11137279	0,18987478	0,39101134	MiniBatchKMEANS	Комментарии поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
11	9	0,90890946	0,4729517	0,5297623	0,884737492	0,53022566	0,35647167	0,26537787	MiniBatchKMEANS	Просмотры поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
12	10	16,6899798	11,4767291	14,66170944	15,06577038	23,9662817	1334,58847	96,5748055	MiniBatchKMEANS	Просмотры поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
13	11	0,18929714	0,10724718	0,120026873	0,20633215	0,16381003	0,46789677	0,30277678	MiniBatchKMEANS	Просмотры поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
14	12	1,20956988	1,26200702	1,003853153	1,281367522	0,40738542	0,35647004	0,34501851	MiniBatchKMEANS	Лайки поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
15	13	23,8531845	20,3539896	20,94793929	20,30415834	24,3850373	643,346611	3960,12465	MiniBatchKMEANS	Лайки поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
16	14	0,24009359	0,22591166	0,232336288	0,191415301	0,09362569	0,61964449	0,56193381	MiniBatchKMEANS	Лайки поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь
17	15	1,20956988	1,26200702	1,218260815	0,763627794	0,40738542	0,35647004	0,34501851	MiniBatchKMEANS	Репосты поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
18	16	23,8531845	20,3539896	26,79521241	13,73757032	24,3850373	643,346611	3960,12465	MiniBatchKMEANS	Репосты поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
19	17	0,24009359	0,22591166	0,253454463	0,155720235	0,09362569	0,61964449	0,56193381	MiniBatchKMEANS	Репосты поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь
20	18	0,49380578	0,50451547	0,519688578	0,763627794	0,40738542	0,35647004	0,34501851	MiniBatchKMEANS	Комментарии поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
21	19	11,3849994	11,4478394	17,02441998	13,73757032	24,3850373	643,346611	3960,12465	MiniBatchKMEANS	Комментарии поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
22	20	0,05501641	0,08239105	0,111514183	0,155720235	0,09362569	0,61964449	0,56193381	MiniBatchKMEANS	Комментарии поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь
23	21	0,49380578	1,26200702	1,218260815	0,763627794	0,40738542	0,35647004	0,34501851	MiniBatchKMEANS	Просмотры поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
24	22	11,3849994	20,3539896	26,79521241	13,73757032	24,3850373	643,346611	3960,12465	MiniBatchKMEANS	Просмотры поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
25	23	0,05501641	0,22591166	0,253454463	0,155720235	0,09362569	0,61964449	0,56193381	MiniBatchKMEANS	Просмотры поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь

(a) Program component 3

	A	B	C	D	E	F	G	H	I	J
1		2-D	3-D	4-D	5-D	6-D	7-D	8-D	9-D	10-D
2	0	0,32903267	0,57363858	0,46641725	1,07143367	0,50907582	0,8999555	0,81990874	0,53230614	0,70170741
3	1	94,2118842	67,0964931	24,3303225	24,3609733	14,7229058	13,0144451	10,3457647	8,98735477	8,36719672
4	2	0,11204582	0,2343865	0,08529562	0,22374404	0,08638841	0,14954679	0,11599464	0,06372881	0,06530952
5	3	0,32348213	0,47041216	0,57997804	0,82288386	0,62098699	1,20744316	0,73355148	1,42070674	1,05307759
6	4	196,659987	45,8746111	27,2749318	20,0064008	13,0392892	15,776097	9,8232297	16,3613395	10,255943
7	5	0,1952218	0,13418002	0,16354399	0,17149682	0,10261479	0,19190799	0,08477245	0,18839467	0,12462876
8	6	0,69834512	0,57363858	0,46641725	0,66077045	0,62098699	1,20744316	0,73355148	1,42070674	1,25581837
9	7	141,371065	67,0964931	24,3303225	18,3043459	13,0392892	15,776097	9,8232297	16,3613395	11,4383189
10	8	0,36306682	0,2343865	0,08529562	0,12907886	0,10261479	0,19190799	0,08477245	0,18839467	0,12820504
11	9	0,69834512	0,47041216	0,57997804	0,82288386	0,62098699	1,20744316	0,73355148	1,42070674	1,05307759
12	10	141,371065	45,8746111	27,2749318	20,0064008	13,0392892	15,776097	9,8232297	16,3613395	10,255943
13	11	0,36306682	0,13418002	0,16354399	0,17149682	0,10261479	0,19190799	0,08477245	0,18839467	0,12462876
14	12	0,7635208	0,93280578	0,94912382	1,18802659	1,29325556	1,20744316	1,45245067	1,42070674	1,25581837
15	13	151,129171	69,9179006	36,2863534	30,0173427	24,3200851	15,776097	17,9937529	16,3613395	11,4383189
16	14	0,39263659	0,32523826	0,26752886	0,24374616	0,22724048	0,19190799	0,17919254	0,18839467	0,12820504
17	15	0,33485336	0,69372682	1,00284835	1,18802659	1,29325556	0,49042527	0,81990874	0,83661825	0,80432571
18	16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
19	17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
21	19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
22	20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
23	21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
24	22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
25	23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

	A	K	L	M	N	O	P	Q	R	S	T	U
1		11-D	12-D	13-D	14-D	15-D	16-D	17-D	18-D	19-D	20-D	21-D
2	0	0,84288562	0,55569397	0,55286189	0,56045919	0,56189754	[illegible]	0,70378773	0,59346491	1,26177527	0,60311314	0,60311356
3	1	10,1585072	7,58502376	7,57471353	6,96179499	7,11412678	[illegible]	7,44464663	6,06644938	10,1182127	6,24143088	6,24140358
4	2	0,15221657	0,04695073	0,06416243	0,05192637	0,05312038	[illegible]	0,08747062	0,02623825	0,12522261	0,02828945	0,02828882
5	3	0,73999749	1,2397411	0,85261956	0,56045919	0,75156226	[illegible]	1,2553215	0,59346491	1,26177527	1,26237424	1,26237689
6	4	7,7688305	10,479196	9,60520341	6,96179499	6,60742372	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	5	0,06354168	0,12013788	0,13301049	0,05192637	0,04082846	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	6	0,79714486	0,79823494	0,55286189	0,56045919	0,56189754	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	7	8,21892381	8,41103188	7,57471353	6,96179499	7,11412678	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	8	0,09279826	0,11409596	0,06416243	0,05192637	0,05312038	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11	9	0,68169961	1,2397411	0,66677625	0,72175895	0,75156226	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	10	7,59149997	10,479196	6,89499972	6,82832765	6,60742372	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	11	0,08166203	0,12013788	0,05604636	0,05596739	0,04082846	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14	12	1,39457648	1,46108844	1,54381201	1,43662335	0,68371331	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15	13	15,156194	14,6811451	13,4075296	13,0315553	7,5663273	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
16	14	0,1688186	0,16886327	0,13794317	0,14448318	0,08565208	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
17	15	0,54147282	0,69608017	0,66677625	0,70378196	1,5498617	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
18	16	7,79960898	6,86815453	6,89499972	7,9826776	13,155932	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
19	17	0,0552324	0,05602022	0,05604636	0,06988136	0,15105098	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	18	0,54147282	0,69608017	0,66677625	0,70378196	0,75156226	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
21	19	7,79960898	6,86815453	6,89499972	7,9826776	6,60742372	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
22	20	0,0552324	0,05602022	0,05604636	0,06988136	0,04082846	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
23	21	0,54147282	0,69608017	0,66677625	0,70378196	0,75156226	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
24	22	7,79960898	6,86815453	6,89499972	7,9826776	6,60742372	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
25	23	0,0552324	0,05602022	0,05604636	0,06988136	0,04082846	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

	A	V	W	X	Y	Z	AA
1		Метод кластеризации	Целевой параметр	Метрика качества кластеризации	Бюджет клиента	Выбранные тематики	Месяц
2	0	MiniBatchKMEANS	Лайки поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
3	1	MiniBatchKMEANS	Лайки поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
4	2	MiniBatchKMEANS	Лайки поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
5	3	MiniBatchKMEANS	Репосты поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
6	4	MiniBatchKMEANS	Репосты поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
7	5	MiniBatchKMEANS	Репосты поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
8	6	MiniBatchKMEANS	Комментарии поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
9	7	MiniBatchKMEANS	Комментарии поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
10	8	MiniBatchKMEANS	Комментарии поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
11	9	MiniBatchKMEANS	Просмотры поста Январь	davies_bouldin_score	1000	Автомобили,Автовладельцы	Январь
12	10	MiniBatchKMEANS	Просмотры поста Январь	calinski_harabasz_score	1000	Автомобили,Автовладельцы	Январь
13	11	MiniBatchKMEANS	Просмотры поста Январь	silhouette_avg	1000	Автомобили,Автовладельцы	Январь
14	12	MiniBatchKMEANS	Лайки поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
15	13	MiniBatchKMEANS	Лайки поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
16	14	MiniBatchKMEANS	Лайки поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь
17	15	MiniBatchKMEANS	Репосты поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
18	16	MiniBatchKMEANS	Репосты поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
19	17	MiniBatchKMEANS	Репосты поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь
20	18	MiniBatchKMEANS	Комментарии поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
21	19	MiniBatchKMEANS	Комментарии поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
22	20	MiniBatchKMEANS	Комментарии поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь
23	21	MiniBatchKMEANS	Просмотры поста Январь	davies_bouldin_score	6000	Автомобили,Автовладельцы	Январь
24	22	MiniBatchKMEANS	Просмотры поста Январь	calinski_harabasz_score	6000	Автомобили,Автовладельцы	Январь
25	23	MiniBatchKMEANS	Просмотры поста Январь	silhouette_avg	6000	Автомобили,Автовладельцы	Январь

(b) Program component 4

Figure 3.7: Clustering quality metrics

We should also note that when writing the program code for visualization of simulation results and further analysis, appropriate directories were created for all program components, where the necessary graphs and bar charts are saved.

It is obvious that for further design, development, testing and implementation of this IMDSS in the industrial circuit [24, 33] it is necessary to create a data storage system using software solutions for creating relational databases and possibly non-relational ones. In this case, the current method of saving simulation results with specified file formats allows to create tables linked and not in appropriate database management systems to automate the process of information storage and processing without any problems.

3.3. Comparative analysis of modeling results

Within the framework of the dissertation research the results of modeling were considered, scenarios proposed by the system were analyzed and appropriate conclusions were made within the framework of the set tasks. However, it should be noted that the software component «Cluster analysis model» can be applied for other problem statements, where it is required not so much to maximize the values of the target indicator under given budget constraints for some list of information sites, as to obtain such a set, in which these sites will be combined according to certain principles, and the choice of the best set will be carried out according to the developed system of rules. In this sense, the considered program complex has the scalability property described in the previous section. In addition, cluster analysis methods can be used not only as a separate tool for solution formation, but also as a method for dimensionality reduction in an optimization problem.

In this paragraph, the modeling results presented in the form of summary tables will be discussed. Summary tables play an important role in data analysis and information presentation. They allow summarizing, grouping and aggregating data from different sources, which helps to quickly and efficiently identify patterns and compare different parameters. Thus, the use of summary tables is an essential tool for effective data analysis and informed decision making.

For example, let's take the following input parameters:

1. Topics: «Automobiles, Car owners»;

2. Month: January;

3. Target metrics: Likes, Reposts, Comments, Views.

Numerical modeling was carried out according to the given parameters. Let's proceed to the comparative analysis of different scenarios of information promotion. Below are 4 tables that allow us to evaluate the numerical results and the speed of their formation. Note that justification of the analysis of modeling results is a key element in the process of data-driven decision-making, it helps to make sure that the conclusions drawn from the modeling are correct and provides a basis for the development of strategies and further actions based on the results obtained.

Program unit number	Post likes	Post reposts	Post comments	Post Views	Client's budget	Total cost	Number of communities	Optimization run time	Clustering run time	Training time
2	3392	334	81	111452	1000	1000	10	0 days 00:00:00.004001		
3	3067	107	46	58642	1000	974	8		0 days 00:00:00.012003	0 days 00:00:03.708894
4	3130	94	44	55888	1000	974	8		0 days 00:00:00.003001	0 days 00:00:02.141678
5	3392	334	81	111452	1000	1000	10	0 days 00:00:00.003001	0 days 00:00:00.105024	0 days 00:00:14.424262
6	3201	117	69	81774	1000	986	11	0 days 00:00:00.002000	0 days 00:00:00.106024	0 days 00:00:14.410813
2	9972	1140	475	529371	6000	5998	26	0 days 00:00:00.012003		
3	8123	559	240	280543	6000	5954	31		0 days 00:00:00.101024	0 days 00:00:07.766416
4	7281	494	168	236201	6000	5451	25		0 days 00:00:00.001000	0 days 00:00:01.664153
5	9972	1140	475	529371	6000	5998	26	0 days 00:00:00.004095	0 days 00:00:00.001000	0 days 00:00:10.580808
6	9449	1118	467	522588	6000	5882	25	0 days 00:00:00.003000	0 days 00:00:00.015003	0 days 00:00:17.013147
2	12387	1804	688	783894	11000	10992	44	0 days 00:00:00.007002		
3	11481	1587	628	652937	11000	10217	37		0 days 00:00:00.079019	0 days 00:00:09.455923
4	9339	924	288	378373	11000	8543	37		0 days 00:00:00.038009	0 days 00:00:03.947927
5	12387	1804	688	783894	11000	10992	44	0 days 00:00:00.004000	0 days 00:00:00.000001	0 days 00:00:16.665632
6	11693	1800	718	789461	11000	10640	43	0 days 00:00:00.003001	0 days 00:00:00.015004	0 days 00:00:17.013322
2	13621	2150	799	965854	16000	15991	52	0 days 00:00:00.010002		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	10407	1641	564	817133	16000	15945	28		0 days 00:00:00.050012	0 days 00:00:09.520101
5	13621	2150	799	965854	16000	15991	52	0 days 00:00:00.004001	0 days 00:00:00.001000	0 days 00:00:14.874318
6	12831	2059	818	913892	16000	15365	58	0 days 00:00:00.010002	0 days 00:00:00.016003	0 days 00:00:20.773879

Figure 3.8: Summary table of results. Subjects: «Automobiles, Car owners». Month: January. Target parameter: Likes

Program unit number	Post likes	Post reposts	Post comments	Post Views	Client's budget	Total cost	Number of communities	Optimization run time	Clustering run time	Training time
2	1597	587	129	161250	1000	987	6	0 days 00:00:00.002000		
3	1043	419	112	105276	1000	699	4		0 days 00:00:00.045011	0 days 00:00:01.292892
4	1058	442	121	121012	1000	915	6		0 days 00:00:00.005002	0 days 00:00:03.508108
5	1597	587	129	161250	1000	987	6	0 days 00:00:00.001000	0 days 00:00:00.000001	0 days 00:00:05.016178
6	1597	587	129	161250	1000	987	6	0 days 00:00:00.002000	0 days 00:00:00.106024	0 days 00:00:13.750781
2	6698	1736	604	660727	6000	5991	25	0 days 00:00:00.002000		
3	5126	1271	482	467840	6000	5847	10		0 days 00:00:00.006002	0 days 00:00:00.584957
4	3739	1052	356	422726	6000	5424	5		0 days 00:00:00.107026	0 days 00:00:08.179697
5	6343	1727	595	652826	6000	6000	23	0 days 00:00:00.002000	0 days 00:00:00.109025	0 days 00:00:24.606837
6	5957	1728	594	653051	6000	5954	23	0 days 00:00:00.001000	0 days 00:00:00.102023	0 days 00:00:22.498651
2	8374	2277	706	909412	11000	10992	33	0 days 00:00:00.004001		
3	5337	1811	577	662371	11000	10074	16		0 days 00:00:00.100709	0 days 00:00:00.247215
4	4756	1651	615	675489	11000	10415	12		0 days 00:00:00.089021	0 days 00:00:08.092901
5	8441	2267	707	899505	11000	10980	32	0 days 00:00:00.003000	0 days 00:00:00.108025	0 days 00:00:33.718610
6	7214	2254	702	879700	11000	10995	31	0 days 00:00:00.004001	0 days 00:00:00.106024	0 days 00:00:28.992862
2	10359	2674	850	1087611	16000	15989	47	0 days 00:00:00.003001		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	5869	1953	614	816331	16000	14597	12		0 days 00:00:00.008002	0 days 00:00:00.590138
5	10771	2657	843	1090513	16000	15991	43	0 days 00:00:00.003001	0 days 00:00:00.106024	0 days 00:00:36.735366
6	9049	2638	814	1081554	16000	15991	35	0 days 00:00:00.003001	0 days 00:00:00.104024	0 days 00:00:32.382823

Figure 3.9: Summary table of results. Subjects: «Automobiles, Car owners». Month: January. Target parameter: Reposts

Program unit number	Post likes	Post reposts	Post comments	Post Views	Client's budget	Total cost	Number of communities	Optimization run time	Clustering run time	Training time
2	875	365	219	128951	1000	1000	7	0 days 00:00:00.003001		
3	385	312	135	107877	1000	944	6		0 days 00:00:00.039010	0 days 00:00:08.175789
4	634	390	202	131086	1000	944	5		0 days 00:00:00.041010	0 days 00:00:03.814897
5	871	357	218	123571	1000	974	6	0 days 00:00:00.001999	0 days 00:00:00.108025	0 days 00:00:16.061855
6	871	357	218	123571	1000	974	6	0 days 00:00:00.002000	0 days 00:00:00.105024	0 days 00:00:13.909219
2	6554	1494	757	655108	6000	5991	27	0 days 00:00:00.002001		
3	2838	924	529	416453	6000	5855	11		0 days 00:00:00.177775	0 days 00:00:09.330109
4	3739	1052	356	422726	6000	5424	5		0 days 00:00:00.107026	0 days 00:00:08.179697
5	6910	1485	756	646598	6000	5994	26	0 days 00:00:00.002001	0 days 00:00:00.109025	0 days 00:00:24.606837
6	5766	1396	751	644514	6000	5964	25	0 days 00:00:00.001000	0 days 00:00:00.001000	0 days 00:00:10.641500
2	9105	1799	952	900895	11000	10971	47	0 days 00:00:00.002001		
3	11409	1589	649	659514	11000	10649	38		0 days 00:00:00.002000	0 days 00:00:02.626978
4	4756	1651	615	675489	11000	10415	12		0 days 00:00:00.089021	0 days 00:00:08.092901
5	8639	1770	944	883800	11000	10837	44	0 days 00:00:00.002001	0 days 00:00:00.001001	0 days 00:00:16.349813
6	8027	1671	929	846029	11000	10564	42	0 days 00:00:00.002001	0 days 00:00:00.001001	0 days 00:00:13.154460
2	10964	2118	1073	1120841	16000	15976	61	0 days 00:00:00.002000		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	5122	1798	664	767519	16000	12431	13		0 days 00:00:00.092022	0 days 00:00:08.104014
5	9607	2111	1051	1081789	16000	15481	51	0 days 00:00:00.000999	0 days 00:00:00.001001	0 days 00:00:17.544121
6	9302	2048	1044	1071474	16000	15575	48	0 days 00:00:00.001001	0 days 00:00:00.001001	0 days 00:00:14.685395

Figure 3.10: Summary table of results. Subjects: «Automobiles, Car owners». Month: January. Target parameter: Comments

Program unit number	Post likes	Post reposts	Post comments	Post Views	Client's budget	Total cost	Number of communities	Optimization run time	Clustering run time	Training time
2	1382	537	148	194726	1000	992	10	0 days 00:00:00.002000		
3	385	312	135	107877	1000	944	6		0 days 00:00:00.039010	0 days 00:00:08.175789
4	634	390	202	131086	1000	944	5		0 days 00:00:00.041010	0 days 00:00:03.814897
5	1382	537	148	194726	1000	992	10	0 days 00:00:00.000999	0 days 00:00:00.107025	0 days 00:00:16.061855
6	1382	537	148	194726	1000	992	10	0 days 00:00:00.001000	0 days 00:00:00.106024	0 days 00:00:13.909219
2	5917	1539	662	746262	6000	5990	27	0 days 00:00:00.002001		
3	5126	1271	482	467840	6000	5847	10		0 days 00:00:00.006002	0 days 00:00:00.584957
4	3739	1052	356	422726	6000	5424	5		0 days 00:00:00.107026	0 days 00:00:08.179697
5	5814	1534	661	743160	6000	5983	28	0 days 00:00:00.001000	0 days 00:00:00.108026	0 days 00:00:22.318820
6	5917	1539	662	746262	6000	5990	27	0 days 00:00:00.001000	0 days 00:00:00.104024	0 days 00:00:24.078659
2	9650	2019	822	1010102	11000	11000	44	0 days 00:00:00.004001		
3	5337	1811	577	662371	11000	10074	16		0 days 00:00:00.100709	0 days 00:00:00.247215
4	6509	1394	499	678536	11000	10537	21		0 days 00:00:00.106025	0 days 00:00:09.774296
5	7007	1888	801	994974	11000	10978	40	0 days 00:00:00.003013	0 days 00:00:00.158036	0 days 00:00:27.603781
6	8509	1986	816	1001131	11000	11000	42	0 days 00:00:00.002003	0 days 00:00:00.001001	0 days 00:00:14.717530
2	10064	2480	922	1235878	16000	15990	46	0 days 00:00:00.002000		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	5437	1613	607	849227	16000	14259	13		0 days 00:00:00.112027	0 days 00:00:08.628373
5	8438	2488	904	1215932	16000	15961	48	0 days 00:00:00.001001	0 days 00:00:00.105185	0 days 00:00:33.014212
6	9056	2401	883	1186803	16000	15338	42	0 days 00:00:00.001001	0 days 00:00:00.104024	0 days 00:00:32.382823

Figure 3.11: Summary table of results. Subjects: «Automobiles, Car owners». Month: January. Target parameter: Views

The structure of the tables is such that it allows comparing scenarios of different program blocks. The tables do not contain all the information available to decision makers, but it is sufficient to analyze the application of the models and methods under consideration, both in terms of time spent on building a solution and comparing the numerical values of various parameters. Using the data from the presented example, the relevant observations and conclusions are formulated.

Conclusion 3.1. *In the obtained scenarios of information promotion on given budgets the best result is demonstrated by program blocks where optimization methods are applied.*

Indeed, the optimization approach allows us to solve the problem exactly. However, if we look at the table in Figure 3.8, where the budget equals 6000 rubles, and compare the results of blocks 2 and 3, i.e. optimization and clustering with the use of interpretive methods of feature selection, then the values of the targets of these two scenarios will differ by 18.54% in favor of block 2. But if we compare the number of communities in which the record will be placed, it turns out that in block 3 there are 5 more such communities, which may be crucial for the decision-maker.

Similarly, consider the scenarios presented in Figure 3.11, namely the budget of 11,000 rubles, blocks 2 and 3. Here, the target values of these two scenarios will differ by 34.42% in favor of block 2, but the number of communities is 2.75 times greater. This suggests that this set of sites, which contains 36.36% of communities from the optimal scenario, gives 65.58% of the result, provided that there is a balance from the used budget of 926 rubles, in contrast to Block 2, where the budget is fully utilized. This, among other things, may indicate a significant difference in user activity indicators in different scenarios.

Thus, scenario modeling helps to make more informed decisions based on modeling and data analysis, assessing the possible risks and benefits of different strategies. In addition, researchers can conduct simulations to assess the likely consequences of different actions. This approach helps to generate and consider different scenarios and choose the most effective solutions.

Therefore, the design of this intelligent system took into account the possibility of analyzing the resulting decisions on various aspects, including the ratio of the sum of the values of the relevant targets per unit of money invested (see Fig. 3.12). Analyzing this kind of data helps to evaluate different sets of communities and make decisions about publication placement.

	Dispersion	MAD	R-Dispersion	Laplacian Score	MCFS 1	MCFS 2	MCFS 3	Метод кластеризации	Бюджет клиента	Выбранные тематики	Целевой параметр	Месяц
0	3,79527559	3,34257206	2,91707921	3,400513479	2,92481203	2,47794118	1,89166667	Spectral Clustering	1000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
1	0,35233161	0,42045455	0,42045455	0,420454545	0,42190889	0,51190476	2,31944444	Spectral Clustering	1000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
2	0,14300847	0,14300847	0,14300847	0,16375	0,10725552	0,20634921	0,08286674	Spectral Clustering	1000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
3	114,276483	114,276483	114,276483	124,5825	101,855324	84,6517094	76,2534722	Spectral Clustering	1000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь
4	1,48184818	1,03250658	1,70389232	0,981002742	1,22746579	1,36429291	1,36327053	Spectral Clustering	6000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
5	0,18307463	0,24784483	0,22247254	0,30260599	0,18802041	0,21450713	0,23802203	Spectral Clustering	6000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
6	0,06166277	0,07093569	0,17430556	0,067037954	0,08350896	0,08472825	0,09035013	Spectral Clustering	6000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
7	70,4142037	87,7269089	96,4498794	74,24381188	96,3227513	79,0028932	71,1277541	Spectral Clustering	6000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь
8	1,48184818	1,03250658	1,70389232	0,981002742	1,22746579	1,04909718	1,12371538	Spectral Clustering	11000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
9	0,18307463	0,10907424	0,22247254	0,110435758	0,12908096	0,18537524	0,15062916	Spectral Clustering	11000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
10	0,03554549	0,03858845	0,03554549	0,039274359	0,04425363	0,06137147	0,06146618	Spectral Clustering	11000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
11	41,3368532	56,4137489	42,5061242	57,36525154	59,600585	70,517105	63,9069198	Spectral Clustering	11000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь
12	1,48184818	1,03250658	1,70389232	0,442118432	0,53541638	0,84613592	1,12371538	Spectral Clustering	16000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
13	0,18307463	0,11204867	0,0728773	0,113210075	0,1187941	0,15215751	0,15062916	Spectral Clustering	16000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
14	0,03554549	0,0390625	0,03554549	0,041791252	0,03601073	0,05219417	0,06146618	Spectral Clustering	16000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
15	33,9881523	56,2868917	36,4579869	56,04642283	54,4480909	60,266898	63,9069198	Spectral Clustering	16000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь
16	1,48184818	0,29225077	1,70389232	0,442118432	0,60711438	0,84613592	0,65365132	Spectral Clustering	21000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
17	0,0522259	0,11204867	0,06731946	0,113210075	0,1187941	0,1381416	0,12578974	Spectral Clustering	21000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
18	0,03554549	0,03350878	0,02096083	0,041791252	0,03765716	0,04389654	0,0413797	Spectral Clustering	21000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
19	31,8731688	47,4596904	36,9070787	56,04642283	55,1034652	58,2426407	52,1630144	Spectral Clustering	21000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь
20	0,44777373	0,29225077	1,70389232	0,442118432	0,60711438	0,84613592	0,65365132	Spectral Clustering	26000	Автомобили,Автовладельцы	Лайки поста Январь	Январь
21	0,06300756	0,11204867	0,06731946	0,113210075	0,1187941	0,11275791	0,12578974	Spectral Clustering	26000	Автомобили,Автовладельцы	Репосты поста Январь	Январь
22	0,01884226	0,03350878	0,02096083	0,041791252	0,03765716	0,03653482	0,0413797	Spectral Clustering	26000	Автомобили,Автовладельцы	Комментарии поста Январь	Январь
23	36,0494459	47,4596904	36,9070787	56,04642283	55,1034652	51,3689522	52,1630144	Spectral Clustering	26000	Автомобили,Автовладельцы	Просмотры поста Январь	Январь

Figure 3.12: Table of efficiency ratios by target indicators

Conclusion 3.2. *In the considered examples, the procedure of preliminary clustering for dimensionality reduction in the optimization problem allows to form scenarios of information advancement no later than using the optimization model and with minimal deviation in the values of target parameters.*

The use of such an approach will allow, in case of large dimensions of the optimization problem, to significantly reduce the running time of the algorithm, provided that a solution equivalent to the optimal one is obtained in quantitative terms, or the same at all, due to the use of an appropriate machine learning method for the construction of the partitioning. The presence of a large number of objects in the selected topic(s) is a relevant problem in large advertising campaigns.

Remark 3.1. *The above examples demonstrate the necessity of applying data prepartitioning in an optimization problem when the number of objects increases.*

It should be noted that the difference in the values of target indicators between blocks 2, 5, 6, if there is any, is within 5-10% of the optimal values, and the speed of solution search procedure in the optimization problem decreases significantly. Let's take for example blocks 2 and 5 from the tables presented in figures 3.9, 3.10, and compare, in the first case scenarios in the budget of 11000 rubles (see Fig. 3.13), in the second case 16000 (see Fig. 3.14).

	Block 2		Block 5
	Community_name		Community_name
0	AUDI	0	AUDI
1	AUTO	1	AUTO
2	Deutsche Autos	2	Best Cars - Auto Club
3	JDM	3	JDM
4	JDM РФ	4	JDM РФ
5	LOW BASS TEAM | Autosound	5	LOW BASS TEAM | Autosound
6	Men`s Academy	6	Men`s Academy
7	Nissan Club	7	Nissan Club
8	Vazzz i Basss™ (18+) | Bot	8	TURBO GARAGE
9	[IN] Road | AUTO	9	Vazzz i Basss™ (18+) | Bot
10	AUTOBUGURT	10	AUTOBUGURT
11	CAR MARKET PERM TO BUY A USED CAR	11	Auto jokes
12	Auto jokes	12	Auto market | Orel | Orel region
13	Auto market | Orel | Orel region	13	Car market of DNR-LNR (Donetsk, Makeyevka, Khartsyzsk)
14	Auto market | Cheboksary | Chuvashia	14	Korchey car market
15	Car market of DNR-LNR (Donetsk, Makeyevka, Khartsyzsk)	15	Auto market Crimea (Simferopol, Sevastopol)
16	Korchey car market	16	Combat Korch
17	Combat Korch	17	Video sale AUTO Sverdlovsk region
18	Video sale AUTO Sverdlovsk region	18	Main Road
19	Main Road	19	Traffic accident Kostroma
20	Traffic accident Kostroma	20	Road wars!
21	Road wars!	21	TRUCK AND OFF-ROAD WORLD
22	TRUCK AND OFF-ROAD WORLD	22	Car music
23	Car music	23	German predators
24	German predators	24	Auto Sale | Auto Selection
25	Auto Sale | Auto Selection	25	A hundred-dollar car
26	A hundred-dollar car	26	A car for a hundred dollars | Ural
27	A car for a hundred dollars | Ural	27	Cars for pennies
28	Cars for pennies	28	Cars for pennies'77 | Moscow
29	Cars for pennies'77 | Moscow	29	IT'S DRIFTING, BABY
30	IT'S DRIFTING, BABY	30	Japanese Guns | JDM
31	Japanese Guns | JDM	31	ministry of operatives
32	ministry of operatives		

Figure 3.13: Communities from solutions for a budget of 11000 rubles. Target parameter: Reposts

	Block 2				Block 5		
	Community_name		Community_name		Community_name		Community_name
0	+300 km/h | Autojournal	35	Video sale AUTO Sverdlovsk region	0	Best Cars - Auto Club	35	Behind the wheel
1	AUDI	36	Off-road vehicles	1	JDM	36	Automobile Club
2	AVTO ZONA 163 Samara Togliatti Auto Barakholka	37	GARAGE | AUTO | GIF	2	JDM | DRIFT	37	Lada Vesta & Lada Vestav
3	Best Cars - Auto Club	38	Garage	3	JDM RF	38	Best cars
4	JDM	39	Main Road	4	Life 4x4: four-wheel drive!	39	Auto enthusiasts
5	JDM | DRIFT	40	Main Road: Auto / Accident	5	MQ: Men's Quality	40	TRUCK AND OFF-ROAD WORLD
6	JDM RF	41	Traffic accident Kostroma	6	Nissan Club	41	Car music
7	Life 4x4: four-wheel drive!	42	Truckers	7	Off-Road | Off-Road World	42	German predators
8	MQ: Men's Quality	43	Road wars!	8	[Uaz 4×4 Off-Road Club]™	43	A hundred-dollar car
9	Nissan Club	44	Behind the wheel	9	[IN] Road | AUTO	44	A car for a hundred dollars | Ural
10	Off-Road | Off-Road World	45	Lada Vesta & Lada Vestav	10	AUTOBUGURT	45	Cars for pennies
11	[Uaz 4×4 Off-Road Club]™	46	Best cars	11	AUTO MECHANIC BY GOD	46	Cars for pennies'77 | Moscow
12	[IN] Road | AUTO	47	Auto enthusiasts	12	CAR THEME™ | CAR | CARS | CARS | CARS	47	IT'S DRIFTING, BABY
13	CAR UP TO 1000$ BELARUS [1K.AUTO]	48	TRUCK AND OFF-ROAD WORLD	13	Auto + Painter	48	JAPANESE BRIGS
14	AUTO CLUB	49	Car music	14	Auto - Overheard	49	Japanese Guns | JDM
15	AUTOBUGURT	50	German predators	15	Auto Almaty	50	ministry of operatives
16	AUTO MECHANIC BY GOD	51	Underneath the tuning	16	Auto Perm		
17	CAR MARKET PERM TO BUY A USED CAR	52	Car for 200	17	Auto after inspection | Belarus		
18	CAR THEME™ | CAR | CARS | CARS | CARS	53	A hundred-dollar car	18	Auto jokes		
19	Auto + Painter	54	A car for a hundred dollars | Ural	19	Auto-selection		
20	Auto - Overheard	55	Cars for pennies	20	Auto market | Orel | Orel region		
21	Auto Almaty	56	Cars for pennies'77 | Moscow	21	Auto market | Cheboksary | Chuvashia		
22	Auto Perm	57	IT'S DRIFTING, BABY	22	Car market of DNR-LNR (Donetsk, Makeyevka, Khartsyzsk)		
23	Auto after inspection | Belarus	58	JAPANESE BRIGS	23	Korchey car market		
24	Auto jokes	59	Japanese Guns | JDM	24	Auto market Krasnodar		
25	Auto-selection	60	ministry of operatives	25	Auto market Crimea (Simferopol, Sevastopol)		
26	Auto market | Orel | Orel region			26	Auto market Peter SPB		
27	Auto market | Cheboksary | Chuvashia			27	Video sale AUTO Sverdlovsk region		
28	Auto market Belarus | Up to 5000$			28	Off-road vehicles		
29	Car market of DNR-LNR (Donetsk, Makeyevka, Khartsyzsk)			29	GARAGE | AUTO | GIF		
30	Korchey car market			30	Main Road		
31	Auto market Krasnodar			31	Main Road: Auto / Accident		
32	Auto market Crimea (Simferopol, Sevastopol)			32	Traffic accident Kostroma		
33	Auto market Peter SPB			33	Truckers		
34	Battle Korch			34	Road wars!		

Figure 3.14: Communities from solutions for a budget of 16000 rubles. Target parameter: Comments

In the case of the «Reposts» target parameter, the differences are minimal both in the number of communities and in the total values of the indicator for the modeled set of sites. But if we turn to the target parameter «Comments», we will notice that the number of communities differs by as much as 10 pieces, and the total values of the selected indicator by 22 comments, or by only 2.05% of the optimal one. Analyzing and comparing such scenarios will help DM to objectively evaluate the necessity of publishing an advertising record in the proposed sets of communities. For this purpose, this paper introduces relative metrics, which are defined in the second paragraph of the first chapter. For example, the distribution coefficients, sociability coefficients, average age of the target audience, and the intensity of publishing activity in a given time period for the respective social network groups can play an important role. All these factors can lead to the fact that the option with a lower value of the target parameter will be chosen, but it will meet the advertiser's requirements by other equally important criteria.

Remark 3.2. *Based on the above results, it is established that interpretable feature selection methods should be applied when solving the clustering problem to compress the feature space.*

The conducted comparative analysis of modeling results is an important stage of the dissertation research, as it allows to optimize the modeling process, to conclude a number of conclusions, remarks and observations regarding the considered mathematical models and methods to achieve scientific and practical results in solving the set tasks taking into account the peculiarities of the specified subject area.

3.4. Conclusions to the third chapter

This chapter is devoted to the construction of the scheme, development and program implementation of an intelligent system of support for managerial decision-making in the task of information dissemination in the MCM. The application of the intelligent system of scenario modeling will allow to optimize the distribution of company resources, to form different scenarios of information dissemination taking into account the peculiarities of sites, behavioral activity of their participants and audience feedback modeling.

The features of implementation and application of the developed intelligent system were considered, schemes of its integration into the industrial circuit on the example of social networks were given, as well as the architecture of the software complex and the scheme of data storage and modeling results with the possibility of transformation and visualization of information were presented. Numerical modeling and comparative analysis of the results have been carried out, which demonstrated the importance of using the scenario approach as a tool for strategic analysis, as well as intelligent tools for processing large data sets in the task of information dissemination in the MCM to support managerial decision-making. Note that the system is programmatically implemented in a cross-platform integrated development environment for the Python programming language — PyCharm.

Conclusion

The dissertation research is devoted to the analysis of information dissemination process in MCM and development of methodology for scenario modeling using optimization and machine learning methods without a teacher. In the course of the conducted research tools for intellectual analysis of statistical data of information sites and formation of scenarios of information dissemination in MCM were developed. This made it possible to realize an intellectual system of management decision support in the field of information and communication technologies for automation of management processes on the example of marketing task of information dissemination in social networks, where the communities of this network act as information platforms.

The following main results were obtained within the framework of the completed dissertation research:

1. Problem statements for modeling the process of information dissemination in MCM using optimization methods are formulated.

2. Problem statements for modeling the process of information dissemination in MCM using machine learning methods are formulated.

3. A software component has been developed that implements a cyclic algorithm for preprocessing statistical data on the user activity of information sites in the task of information dissemination in the MCM in the Python programming language in the cross-platform integrated development environment PyCharm.

4. A software component with a recommendation block for forming scenarios of information dissemination in MCM and solving optimization problems with the possibility of transformation and visualization of information in the Python programming language in the cross-platform integrated development environment PyCharm has been developed and implemented.

5. We developed software components using machine learning and teacherless feature selection methods to solve the problem of clustering of information sites

and the problem of dimensionality reduction in an optimization problem with the possibility of transforming and visualizing information in the Python programming language in the cross-platform integrated development environment PyCharm.

6. The architecture has been developed and the intellectual system of management decision support in the task of information dissemination in the MCM has been programmatically implemented, as well as the scheme of data storage and modeling results with the possibility of transformation and visualization of information in the Python programming language in the cross-platform integrated development environment PyCharm has been proposed. The comparison of information dissemination scenarios is carried out and the feasibility of forming several sets of information sites through numerical modeling is demonstrated.

7. A tool for numerical simulation of the system under study has been developed, which allows to perform sensitivity analysis of criteria, as well as to analyze the process of forming unique scenarios of information dissemination as a result of changing preferences in the problem of multi-criteria optimization on the example of the market of goods and services in the digital environment, taking into account the nomenclature of goods, budget and time interval. Application of the developed software components allows to correct user preferences by adjusting hyperparameters of machine learning methods, as well as to reduce the time of formation of scenarios of information dissemination. Analyzing the importance of attributes allowed us to determine the basic set of significant characteristics of objects using the selected methods of feature space compression.

In conclusion, we note that all the tasks formulated in the framework of the study have been accomplished, and the set goal has been achieved in full.

Further work on this subject can be aimed at developing the functional capabilities of the intelligent system in the framework of solving the problem of modeling scenarios of information dissemination in the MCM. Let us present promising directions and tasks of future research:

- Integration with other IT solutions to perform datamining of various additional characteristics of information sites.
- Application of time series and text analysis methods for predicting values of key characteristics and forming recommendations on information placement in different MCM.
- Detection by hidden patterns of bots and fake accounts on information sites.
- Application of deep machine learning algorithms and neural network technologies to analyze and generate new media content.

In addition, the actual tasks will be to improve the presented schemes of applied mathematical models and program complex, by editing or adding new program and functional modules to achieve greater efficiency, in terms of the time of formation of scenarios of information dissemination and their number at given input parameters of the system, when solving problems in various fields of knowledge (economics and management, sociology, political science).

Bibliography

1. Avtoposting v sotsial’nykh setiakh — «Pur Ninja». — URL: https://pur.ninja/ (accessed 11.06.2022).
2. Adigamov A. E. Sistema podderzhki priniatiia reshenii pri proektirovanii slozhnykh tekhnicheskikh sistem / A. E. Adigamov, P. V. Makarov, I. N. Simacheva, N. D. Simachev // Istoriia i pedagogika estestvoznaniia. — 2013. — №2. — P. 22—24.
3. Azhmukhamedov, I.M. Modelirovanie protsessa rasprostraneniia informatsii v sotsial’nykh setiakh / I. M. Azhmukhamedov, D. A. Machueva, D. A. Zholobov // Fundamental’nye issledovaniia. — 2017. — № 5. — P. 9—14.
4. Azhmukhamedov, I.M. Modelirovanie rasprostraneniia informatsii v sotsial’nykh setiakh / I.M. Azhmukhamedov, N. A. Anshakov, V. M. Gudonis // Problemy Nauki. — 2017. — № 20(102) — P. 16—19.
5. Analiz otcheta «Global Digital 2023». — URL: https://www.web-canape.ru/business/statistika-interneta-i-socsetej-na-2023-god-cifry-i-trendy-v-mire-i-v-rossii/ (accessed 15.03.2024).
6. Astaf’eva, E. V. Model’ reklamnoi kompanii s effektom «nedoedaniia» reklamy / E.V. Astaf’eva, A.F. Terpugov // Vestnik Tomskogo gosudarstvennogo universiteta. — 2004. — № 284. — P. 34—36.
7. Akhmedova, D. D. Matematicheskaia model’ funktsionirovaniia strakhovoi kompanii s uchetom raskhodov na reklamu / D. D. Akhmedova, A. F. Terpugov // Izvestiia vuzov. Fizika. — 2001. — № 1. — P. 25—29.
8. Akhmedova, D. D. Optimizatsiia deiatel’nosti strakhovoi kompanii s uchetom raskhodov na reklamu / D. D. Akhmedova, O. A. Zmeev, A. F. Terpugov // Vestnik Tomskogo gosudarstvennogo universiteta. — 2002. — № 275. — P. 181—184.
9. Akhmedova, D. D. Optimizatsiia raskhodov na reklamu pri deiatel’nosti strakhovoi kompanii / D. D. Akhmedova, O. A. Zmeev // Izvestiia vuzov. Fizika. — 2001. — № 6. — P. 3—7.

10. Balashova I. V. SISTEMY PODDERZhKI PRINIaTIIa REShENII / I. V. Balashova, T. A. Tereshchenko // The Scientific Heritage. — 2021. — № 79(4). — P. 3—7.
11. Badryzlov, V. A. Otsenka effektivnosti rasprostraneniia informatsii v sotsial'nykh setiakh s ispol'zovaniem imitatsionnogo modelirovaniia / V. A. Badryzlov, V. V. Sidel'tsev // Kreativnaia ekonomika. — 2018. — № 9(12). — P. 1359—1372.
12. Biblioteka «SciPy», algoritmy optimizatsii «scipy.optimize». — URL: https://sociate.ru/ (accessed 21.09.2023).
13. Birzha reklamy — «Sociate». — URL: https://sociate.ru/ (accessed 11.06.2022).
14. Borodina, I. P. Model' optimal'nogo povedeniia firmy s uchetom vliianiia reklamy / I. P. Borodina, C. B. Zhak // Ekonomicheskaia nauka sovremennoi Rossii. — 2004. — №3. — P. 80—87.
15. Borodina, I. P. Sistemy upravleniia reklamnymi kommunikatsiiami firmy / I. P. Borodina. — Rostov n/D : izd-vo RBPKhL SP RF, 2004. — 123 p.
16. Visaitova, L. R. Sistemy podderzhki priniatiia reshenii: effektivnost' priniatiia reshenii / L. R. Visaitova // Nauchno-prakticheskii elektronnyi zhurnal Alleia Nauki. — 2020. — № 11(50). — P. 219—225.
17. Gomaniuk, S. O. Matematicheskoe modelirovanie reklamnoi kampanii obrazovatel'nogo tsentra / S.O. Gomaniuk // VKR. NIU VShE. Moskva. — 2015. — 53 p. URL: https://www.hse.ru/edu/vkr/153015895?ysclid=lwewjxodyp495766419.
18. Gorkovenko, D. K. Obzor modelei rasprostraneniia informatsii v sotsial'nykh setiakh / D.K. Gorkovenko // Molodoi uchenyi. — 2017. — № 8 (142). — P. 23—28.
19. Gorkovenko, D. K. Sravnitel'nyi analiz modelei epidemii i kletochnogo avtomata pri modelirovanii rasprostraneniia informatsii v sotsial'nykh setiakh / D. K. Gorkovenko // Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie. — 2017. — № 3(10). — P. 103—113.
20. Gorchakova, R. R. Osobennosti formirovaniia korporativnogo imidzha / R. R. Gorchakova // Izvestiia vysshikh uchebnykh zavedenii. Povolzhskii region.

Obshchestvennye nauki. — 2012. — № 2(22). — P. 185—192.

21. Gribanova E. B. Algoritmy modelirovaniia rasprostraneniia informatsii pri marketingovykh meropriiatiiakh v gruppakh onlainovoi sotsial'noi seti / E. B. Gribanova // Problemy upravleniia. — 2018. — № 1. — P. 66—73.
22. Gubanov D. A. Modeli vliianiia v sotsial'nykh setiakh / D. A. Gubanov, D. A. Novikov, A. G. Chkhartishvili // Upravlenie bol'shimi setiami. — 2009. — № 27. — P. 205—281.
23. Gubanov, D. A. Sotsial'nye seti: modeli informatsionnogo vliianiia, upravleniia i protivoborstva [Tekst] / D. A. Gubanov, D. A. Novikov, A. G. Chkhartishvili; Rossiiskaia akademiia nauk, Institut problem upravleniia. — Izd. 3-e, pererab. i dop. — Moskva : Izd-vo MTsNMO, 2018. — 223 p.
24. Guss, S. V. Razrabotka semeistva programmnykh sistem v spetsificheskoi predmetnoi oblasti / S. V. Guss // Matematicheskie struktury i modelirovanie. — 2011. — № 22. — P. 55–68.
25. Delitsyn, L. L. Kolichestvennye modeli rasprostraneniia novovvedenii v sfere informatsionnykh i telekommunikatsionnykh tekhnologii [Tekst] : monografiia / L. L. Delitsyn ; Moskovskii gos. un-t kul'tury i iskusstva. — Moskva : Moskovskii gos. un-t kul'tury i iskusstva, 2009. — 106 p.
26. Dzialoshinskii, I. M. Kommunikativnoe vozdeistvie: misheni, strategii, tekhnologii: monografiia / I. M. Dzialoshinskii. — M.: NIU VShE, — 2012. — 572 p.
27. Ershov, V. E. Issledovanie otnosheniia pol'zovatelei sotsial'nykh setei k aktivnosti po prodvizheniiu na ikh osnove / V. E. Ershov // Internet-zhurnal «Naukovedenie». — 2017. — № 5(9). — P. 1—13. URL: https://cyberleninka.ru/article/n/issledovanie-otnosheniya-polzovateley-sotsialnyh-setey-k-aktivnosti-po-prodvizheniyu-na-ih-osnove/viewer.
28. Ekhlakov, Iu. P. Mnogokriterial'naia zadacha podderzhki priniatiia reshenii pri vybore variantov postavki na tselevye rynki lineiki programmnykh produktov / Iu. P. Ekhlakov, D. N. Baraksanov // Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniia i radioelektroniki. — 2015. — № 3(37). — P. 114—119.

29. Ekhlakov, Iu. P. Osnovnye polozheniia po razrabotke programmy prodvizheniia programmnykh produktov v seti Internet / Iu. P. Ekhlakov, D. N. Baraksanov // Biznes-informatika. — 2012. — № 4(22). — P. 26—32.
30. Ekhlakov, Iu. P., Struktura i soderzhanie kommunikatsionnogo soobshcheniia pri organizatsii Internet-reklamy programmnykh produktov / Iu. P. Ekhlakov, D. N. Baraksanov // Internet-marketing. — 2016. — № 3(93). — P. 146—156.
31. Izmodenova, K. V. Ob optimal'nom upravlenii protsessom rasprostraneniia informatsii / A. P. Mikhailov, K. V. Izmodenova // Matematicheskoe modelirovanie. — 2005. — № 5(17). — P. 67—76.
32. Karamova, V. V. Pozitsionirovanie brenda/produktsii organizatsii kak instrument povysheniia effektivnosti ee deiatel'nosti / V. V. Karamova // Vestnik UGNTU. Nauka, obrazovanie, ekonomika. Seriia ekonomika. — 2018. — № 3(25). — P. 105—112.
33. Koznov, D. V. Instrumenty dlia upravleniia variantivnost'iu: gotovnost' k promyshlennomu primeneniiu / D. V. Koznov, I. A. Novitskii, M. N. Smirnov // Trudy SPIIRAN. — 2013. — № 3(26). — P. 297—331.
34. Koel'o, L. P. Postroenie sistem mashinnogo obucheniia na iazyke Python / L. P. Koel'o, V. Richart. — 2-e izdanie; per. s angl. A.A. Slinkina. — M.: DMK Press, 2016. — 302 p.
35. Kriazheva, Iu. S. Ispol'zovanie Social Media Marketing kak instrumenta prodvizheniia kompanii / Iu. S. Kriazheva // Formula menedzhmenta. — 2017. — № 1(1). — P. 5—11.
36. Larichev, O. I. Sistemy podderzhki priniatiia reshenii. Sovremennoe sostoianie i perspektivy ikh razvitiia [Elektronnyi resurs] / O. I. Larichev, A. B. Petrovskii // Itogi nauki i tekhniki. Ser. Tekhnicheskaia kibernetika. — Moskva : VINITI, 1987. — Vol. 21. — P. 131—164. — URL: http://www.raai.org/library/papers/Larichev/Larichev_Petrovsky_1987.pdf (accessed 23.04.2019).
37. Livshin, D. A. Matematicheskoe modelirovanie v marketinge pri postroenii reklamnykh kampanii / D. A. Livshin, L. I. Voronova // Sovremennye naukoemkie tekhnologii. — 2014. — № 5-2. — P. 207—209.

38. Marevtseva, N. A. Nekotorye matematicheskie modeli informatsionnogo napadeniia i informatsionnogo protivoborstva / N. A. Marevtseva // Sotsiologiia. — 2011. — № 3. — P. 26—35.
39. Marevtseva, N. A. Prosteishie matematicheskie modeli informatsionnogo protivoborstva / N. A. Marevtseva // Matematicheskoe modelirovanie sotsial'nykh protsessov. — 2010. — № 11 — P. 59—72.
40. Marketingovaia platforma iskusstvennogo intellekta «Albert». — URL: https://albert.ai/ (accessed 15.03.2024).
41. Mirskaia, S. Iu. Otsenka vliianiia reklamy na formirovanie potrebitel'skogo sprosa / S. Iu. Mirskaia, V. I. Sidel'nikov, I. P. Borodina // Izvestiia vuzov. Severo-Kavkazskii region. Seriia: Estestvennye nauki. — 2004. — № 6. — P. 13—18.
42. Mikhailov, A. P. Modeli informatsionnoi bor'by / A. P. Mikhailov, N. A. Marevtseva // Matematicheskoe modelirovanie. — 2011. — № 10(23). — P. 19—32.
43. Mikhailov, A. P. Model' informatsionnogo protivoborstva v sotsiume pri periodicheskom destabiliziruiushchem vozdeistvii / A. P. Mikhailov, A. P. Petrov, O. G. Proncheva, N. A. Marevtseva // Matematicheskoe modelirovanie. — 2017. — № 2(29). — P. 23—32.
44. Mikhailov, A. P. Razvitie modeli rasprostraneniia informatsii v sotsiume / A. P. Mikhailov, A. P. Petrov, N. A. Marevtseva, I. V. Tret'iakova // Matematicheskoe modelirovanie. — 2014. — № 3(26). — P. 65—74.
45. Miuller A., Gvido S. Vvedenie v mashinnoe obuchenie s pomoshch'iu Python: rukovodstvo dlia spetsialistov po rabote s dannymi / A. Miuller, S. Gvido. — per. s angl. — SPb.: OOO «Al'fa-kniga», 2017. — 480 p.
46. Netologiia. Soobshchestvo sotsial'noi seti. — URL: https://netology.ru/glossariy/gruppa-v-socialnoy-seti (accessed 09.02.2024).
47. Nosova, M. V. Modelirovanie rasprostraneniia informatsii v detsentralizovannykh setevykh sistemakh s nereguliarnoi strukturoi / M. V. Nosova, L. I. Sennikova // Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh. — 2014. — № 17. — P. 329—335.

48. Opisanie metodov klasternogo analiza. Biblioteka Scikit-learn. — URL: https://scikit-learn.org/stable/modules/clustering.html (accessed 22.03.2023).
49. Opisanie metodov otsenki kachestva klasterizatsii. Biblioteka Scikit-learn. — URL: https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation (accessed 22.03.2023).
50. Opisanie metoda poroga otkloneniia. Biblioteka Scikit-learn. — URL: https://scikit-learn.org/stable/modules/feature_selection.html (accessed 22.03.2023).
51. Optimizatsiia mediaplana s pomoshch'iu matematicheskogo modelirovaniia. — URL: https://www.marketing.spb.ru/lib-special/case/math_model.htm (accessed 17.03.2024).
52. Orel'en, Zh. Prikladnoe mashinnoe obuchenie s pomoshch'iu Scikit-Learn, Keras i TensorFlow: kontseptsii, instrumenty i tekhniki dlia sozdaniia intellektual'nykh sistem / Zh. Orel'en. — 2-e izd., per. s angl. — SPB.: OOO «Dialektika», 2020. — 1040 p.
53. Popov E. V. Teoriia vychisleniia okhvata SMI / E. V. Popov, G. A. Shmatov // Problemy upravleniia. — 2009. — № 5. — P. 22—27.
54. Popov E. V. Vychislenie okhvata SMI / E. V. Popov, G. A. Shmatov // Problemy upravleniia. — 2010. — № 2. — P. 34—38.
55. Realizatsiia i opisanie metoda glavnykh komponent. Biblioteka Scikit-learn. — URL: https://scikit-learn.org/stable/modules/decomposition.html#pca (accessed 22.03.2023)
56. Reklamnaia platforma na baze «Big Data MTS» — «MTS Marketolog». — URL: https://marketolog.mts.ru/ (accessed 15.03.2024).
57. Samarskii, A. A. Matematicheskoe modelirovanie: Idei. Metody. Primery. / A. A. Samarskii, A. P. Mikhailov. — 2-e izd., ispr. — M.: Fizmatlit, 2001. — 320 p.
58. Svirkin, M. V. Imitatsionnoe modelirovanie sotsial'noi seti s dvumia liderami / M. V. Svirkin, V. V. Karelin, V. M. Bure, A. V. Ekimov // Vestnik SPbGU. Seriia 10. Prikladnaia matematika. Informatika. Protsessy upravleniia. — 2016. — № 1. — P. 92—98.

59. Semenova, O. V. Vazhnost' pozitsionirovaniia brenda i aktual'nye metody prodvizheniia brenda organizatsii / O. V. Semenova // Vestnik Samarskogo universiteta. Ekonomika i upravlenie. — 2023. — № 2(14). — P. 155—163.
60. Semiglazov, A. M. Matematicheskoe modelirovanie reklamnoi kampanii / A. M. Semiglazov, V. A. Semiglazov, K. I. Ivanov // Upravlenie, vychislitel'naia tekhnika i informatika, doklady TUSURa. — 2010. — № 2(22) — P. 342—349.
61. Semiglazov, V. A. Optimizatsiia raskhodov na reklamnuiu kampaniiu / V. A. Semiglazov // Marketing. — 2007. — № 1. — P. 63—70.
62. Servis analitiki kontenta sotsial'nykh setei — «Popsters». — URL: https://popsters.ru (accessed 11.06.2022).
63. Servis upravleniia sotsial'nymi setiami — «LiveDune». — URL: https://livedune.com/ru (accessed 11.06.2022).
64. Sotsial'naia set' «VKontakte». — URL: https://vk.com/ (accessed 15.03.2024).
65. Statistika akkauntov v sotsial'nykh setiakh — «JagaJam». — URL: https://jagajam.com (accessed 11.06.2022).
66. Statistika soobshchestv sotsial'noi seti «VKontakte» — «AllSocial». — URL: http://allsocial.ru (accessed 11.06.2022).
67. Terelianskii, P. V. Sistemy podderzhki priniatiia reshenii. Opyt proektirovaniia : monografiia / P. V. Terelianskii. — Volgograd : VolgGTU. 2009. — 127 p.
68. Timofeev, S. V., Matematicheskoe modelirovanie informatsionnogo protivoborstva / S. V. Timofeev, A. V. Baenkhaeva // Nauchno-tekhnicheskie vedomosti SPbGPU. Fiziko-matematicheskie nauki. — 2021. — № 1(14). — P. 164—176.
69. Trofimov, V. M. Model' informatsionnogo vozdeistviia v sotsial'nykh setiakh / V. M. Trofimov, L. A. Vidovskii, R. A.D'iachenko // Nauchnyi zhurnal KubGAU. — 2015. — № 110. — P. 1788—1801.
70. Flakh, P. Mashinnoe obuchenie. Nauka i iskusstvo postroeniia algoritmov, kotorye izvlekaiut znaniia iz dannykh [Tekst] : per. s angl. / P. Flakh. — Moskva : DMK Press, 2015. — 399 p.
71. Fursov, D. V. Analiz primeneniia metodov mashinnogo obucheniia i optimizatsii v zadache formirovaniia nabora soobshchestv dlia prodvizheniia informatsii

v sotsial'nykh media / D. V. Fursov (nauch. ruk. Krylatov A.Iu.) // Sbornik tezisov dokladov kongressa molodykh uchenykh. Elektronnoe izdanie. — SPb: Universitet ITMO. — 2024. URL: https://kmu.itmo.ru/digests/article/13613.

72. Fursov, D. V. Imitatsionnoe modelirovanie informatsionnogo vozdeistviia s ispol'zovaniem sredstv massovoi kommunikatsii / D. V. Fursov // Protsessy upravleniia i ustoichivost'. — 2018. — № 1(5). — P. 517—522.
73. Fursov, D. V. Imitatsionnoe modelirovanie prodvizheniia informatsii v slozhnykh organizatsionnykh sistemakh / D. V. Fursov // Protsessy upravleniia i ustoichivost'. — 2019. — № 1(6). — P. 486—490.
74. Fursov, D. V. Intellektual'naja sistema podderzhki prinjatija reshenij v zadache rasprostranenija informacii v sredstvah massovoj kommunikacii / D. V. Fursov, A. Iu. Krylatov // Informatsionnye sistemy i tekhnologii. — 2024. — № 2(142). — P. 80—89.
75. Fursov, D. V. Intellektual'naia sistema podderzhki priniatiia upravlencheskikh reshenii v zadache formirovaniia nabora ploshchadok rasprostraneniia informatsii v sredstvakh massovoi kommunikatsii / D. V. Fursov (nauch. ruk. Krylatov A.Iu.) // Sbornik tezisov dokladov kongressa molodykh uchenykh. Elektronnoe izdanie. — SPb: Universitet ITMO. — 2024. URL: https://kmu.itmo.ru/digests/article/13603.
76. Fursov, D. V. Primenenie metodov optimizacii v zadache rasprostranenija informacii v sredstvah massovoj kommunikacii / D. V. Fursov, A. Iu. Krylatov, M. V. Svirkin // Prikladnaia matematika i voprosy upravleniia. — 2024. — № 1. — P. 139—159.
77. Fursov, D. V. Primenenie metodov mashinnogo obuchenija v zadache rasprostranenija informacii v sredstvah massovoj kommunikacii / D. V. Fursov // Avtomatizatsiia protsessov upravleniia. — 2024. — № 2(76). — P. 75—86.
78. Fursov, D. V. Programmnyi kompleks «Intellektual'naia sistema podderzhki priniatiia upravlencheskikh reshenii v zadache rasprostraneniia informatsii v sredstvakh massovoi kommunikatsii» / D. V. Fursov // Programma dlia EVM. Svidetel'stvo o registratsii programmy dlia EVM RU2024614394, data registratsii 22.02.2024, nomer i data postupleniia zaiavki – 2024613251, 19.02.2024,

data publikatsii 22.02.2024, Biulleten' FIPS №3.

79. Fursov, D. V. Programmnyi komponent «Obrabotka statisticheskikh dannykh aktorov sredstv massovoi kommunikatsii» / D. V. Fursov // Programma dlia EVM. Svidetel'stvo o registratsii programmy dlia EVM RU2024615049, data registratsii 01.03.2024, nomer i data postupleniia zaiavki – 2024613270, 19.02.2024, data publikatsii 01.03.2024, Biulleten' FIPS №3.
80. Fursov, D. V. Programmnyi komponent «Programmnaia realizatsiia metodov klasternogo analiza dlia formirovaniia reshaiushchego nabora aktorov v zadache rasprostraneniia informatsii» / D. V. Fursov // Programma dlia EVM. Svidetel'stvo o registratsii programmy dlia EVM RU2024614766, data registratsii 28.02.2024, nomer i data postupleniia zaiavki – 2024613298, 19.02.2024, data publikatsii 28.02.2024, Biulleten' FIPS №3.
81. Fursov, D. V. Programmnyi komponent «Programmnaia realizatsiia metodov optimizatsii v zadache rasprostraneniia informatsii» / D. V. Fursov // Programma dlia EVM. Svidetel'stvo o registratsii programmy dlia EVM RU2024614893, data registratsii 29.02.2024, nomer i data postupleniia zaiavki – 2024613262, 19.02.2024, data publikatsii 01.03.2024, Biulleten' FIPS №3.
82. Fursov, D. V. Programmnyi komponent «Programmnaia realizatsiia metodov optimizatsii s predvaritel'noi klasterizatsiei v zadache rasprostraneniia informatsii» / D. V. Fursov // Programma dlia EVM. Svidetel'stvo o registratsii programmy dlia EVM RU2024614590, data registratsii 27.02.2024, nomer i data postupleniia zaiavki – 2024613238, 19.02.2024, data publikatsii 27.02.2024, Biulleten' FIPS №3.
83. Fursov, D. V. Razrabotka prototipa intellektual'noj sistemy podderzhki prinyatiya upravlencheskih reshenij v zadache rasprostraneniya informacii / D. V. Fursov // IT. Nauka. kreativ : Materialy VI Vserossijskoj konferencii, v ramkah I Mezhdunarodnogo foruma, Omsk, 14–16 maya 2024 goda. — Moskva: Obshchestvo s ogranichennoj otvetstvennost'yu "Izdatel'sko-knigotorgovyj centr "Kolos-s", 2024. — P. 432—436.

84. Fursov, D. V. Statisticheskoe imitatsionnoe modelirovanie prodvizheniia informatsii v sotsial'nykh setiakh / D. V. Fursov, M. V. Svirkin // Perspektivy nauki. — 2020. — № 6(129). — P. 39—44.
85. Khabr. Stat'ia: «Obzor metodov otbora priznakov». — URL: https://habr.com/ru/companies/jetinfosystems/articles/470622/ (accessed 22.03.2023).
86. Khrapov P. V., Matematicheskoe modelirovanie protsessa rasprostraneniia novostei v sotsial'nykh setiakh / P. V. Khrapov, V. A. Stolbova // Sovremennye informatsionnye tekhnologii i IT-obrazovanie. — 2019. — № 1(15). — P. 225—231.
87. Chasto zadavaemye voprosy. «Popsters». — URL: https://popsters.ru/app/faq/ (accessed 11.06.2022).
88. Chernorutskii, I. G. Metody priniatiia reshenii. — SPb.: BKhV–Peterburg, 2005. — 416 p.: il.
89. Chesnokov, V. O. Primenenie algoritma vydeleniia soobshchestv v informatsionnom protivoborstve v sotsial'nykh setiakh / V. O. Chesnokov // Voprosy kiberbezopasnosti. — 2017. — № 1(19). — P. 37—44.
90. Chibisova, A. V., Matematicheskoe modelirovanie reklamnoi kampanii / A. V. Chibisova, D. S. Shinakov // Matematicheskoe modelirovanie i chislennye metody. — 2022. — № 3. — P. 84—97.
91. Shabanov, R. M. Intellektual'naia informatsionnaia sistema podderzhki priniatiia reshenii / R. M. Shabanov, N. A. Mikushin // Molodoi issledovatel' Dona. — 2019. — № 4(19). — P. 91—97.
92. Shigina, Ia. I. Marketing v sotsial'nykh media: sovremennye instrumenty prodvizheniia dlia malogo biznesa / Ia. I. Shigina, K. A. Zorina // Vestnik tekhnologicheskogo universiteta. — 2015. — № 23(18). — P. 96—102.
93. Shmatov, G. A. K probleme poiska zakonomernostei v sfere mediavozdeistvii / G. A. Shmatov // Vestnik UGTU–UPI. Seriia ekonomika i upravlenie. — 2010. — № 6. — P. 102—111.
94. Shmatov, G. A. Matematicheskaia teoriia mediaplanirovaniia : monografiia / G. A. Shmatov ; Rossiiskaia akad. nauk, Ural'skoe otd-nie, In-t ekonomiki. —

Ekaterinburg : In-t ekonomiki UrO RAN, 2009. — 329 p.

95. Shmatov, G. A. Optimizatsiia perioda razmeshcheniia reklamy / G. A. Shmatov // Reklama: teoriia i praktika. — 2008. — № 1. — P. 8—15.
96. Alyoubi, B. A. Decision support system and knowledge-based strategic management / B. A. Alyoubi // ELSEVIER. — 2015. — № 65. — P. 278—284.
97. Bass, F. M. A new product growth for model consumer durables / F. M. Bass // Management Science. — 1969. — № 5(15). — P. 215—227.
98. Cai, D. Unsupervised feature selection for Multi-Cluster data / D. Cai, C. Zhang, X. He // Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2010. — P. 333—342.
99. Calinski, T. A dendrite method for cluster analysis / T. Calinski, J. Harabasz // Communications in Statistics. — 1974. — Vol. 3, iss. 1, — P. 1—27.
100. Davies, D. L. A Cluster Separation Measure / D. L. Davies, D. W. Bouldin // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1979. — Vol. PAMI-1, iss. 2. — P. 224—227.
101. Fursov, D. Problems of Data Processing in the Problem of Modeling Advertising Campaigns in Social Networks Using Python Libraries / D. Fursov, A. Krylatov, M. Svirkin, F. Prokhorenko // R. Silhavy, P. Silhavy, Z. Prokopova (eds) Data Science and Algorithms in Systems. CoMeSySo 2022. Lecture Notes in Networks and Systems. Springer, Cham. — 2023. — Vol 597, iss. 2. — P. 990—1001.
102. He, X. Laplacian Score for Feature Selection / X. He, D. Cai, P. Niyogi // Proceeding of Advances in Neural Information Processing Systems. — 2005. — Vol. 18. — URL: https://api.semanticscholar.org/CorpusID:14270830.
103. Jin, J. HPC Simulations of Information Propagation over Social Networks / J. Jin, S. J. Turner, B. S. Lee, J. Zhong, B. He // Procedia Computer Science. — 2012. — Vol. 9. — P. 292—301.
104. Khodashahri, N. G. Decision support system (DSS) / N. G. Khodashahri, M. H. Sarabi // Singaporean journal of business economics and management studies. — 2013. — №6. — P. 94—102.

105. Kleinberg, J. M. An Impossibility Theorem for Clustering / J. M. Kleinberg // Adv Neural Inform Process Syst (NIPS). — 2002 — Vol. 15. — URL: https://api.semanticscholar.org/CorpusID:1487131
106. Klepek, M. Marketing communication model for social networks / M. Klepek, H. Starzyczna // Journal of business economics and management. — 2018. — Vol. 19, iss. 3. — P. 500—520.
107. Pistol, L. Model of simulation for optimizing marketing mix through conjoint analysis case study: launching a product on a new market / L. Pistol, R. Bucea-Manea-Tonis // Economics World. — 2017. — Vol. 5, iss 4. — P. 311—315.
108. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis / P. J. Rousseeuw // Journal of Computational and Applied Mathematics. — 1987. — Vol. 20. — P. 53—65.
109. Tripathi, K. P. Decision support system is a tool for making better decisions in the organization / K. P. Tripathi // Indian Journal of Computer Science and Engineering. — 2017. — № 21. — P. 112—117.
110. Yekhlakov, Yu. P. Mathematical model and algorithm for selection of Internet sites and places for display of communication messages in planning advertising campaigns / Yu. P. Yekhlakov, D. N. Baraksanov // Business Informatics. — 2017. — № 1(39). — P. 55—60.
111. Yekhlakov, Yu. P. Mathematical model and algorithm of selecting software promotion options differentiated by functionality and business models / Yu. P. Yekhlakov, D. N. Baraksanov // Business Informatics. — 2015. — № 4(34). — P. 55—62.