

Modeling of an interactive distance learning platform by means of modern information technologies.

Tatiana Grober^{*,1} and *Oleg Grober*¹

¹ Don state technical University, Gagarin square, 1, 344003 Rostov-on-Don, Russia

Abstract. According to the requirements of the Ministry of Science and Education, teachers of various qualifications are required to regularly improve their professional skills. Using the distance learning platform allows you to automate the learning process, starting with the paperwork, ending with the final testing and obtaining a diploma. The problem is that during the course of training, you have to wait a long time for a response from a consultant on the subject or on the course support. The study of this problem was carried out on the basis of the existing database of the company "InfoTeacher". To determine the most promising courses, we used cluster data analysis using the R software environment. At the beginning, we calculated the optimal number of clusters according to the criterion of rocky scree, and then continued the cluster analysis using the k-means method. Thus, the main input parameters for building an interactive simulation model were identified. This model is a queuing system and was implemented in the GPSS World programming language. The parameters of the model were selected so that it was effective: to minimize the queues for consultation with the curator, while the curators should be loaded by at least 50%.

1 Introduction

Teachers who want to improve their skills can use any educational platform located on the Internet. They differ in the design, the specifics of the offered product, the way of interaction with the client and payment methods. As a rule, all resources are licensed and a state-issued diploma is issued upon completion of the training. A common problem with such educational platforms is the lack of advisory support in the learning process, or the student is forced to wait a long time for a response from the site managers. When such situations occur, a negative attitude is formed in the consumer, and it is unlikely that he will ever return to this site, which entails a loss of the company's profit.

Creating an educational platform requires a lot of time and resources. Therefore, to create an effective distance learning system, it is advisable to first build a simulation model of the system.

Simulation methods allow you to create models of various systems and processes using programs written in specially developed programming languages. In this way, the created

* Corresponding author: grober71@mail.ru

models of real systems can be studied on a computer, changing the parameters, assessing their impact on the characteristics of the model, and predicting the behavior of the system, depending on the changes made.

Simulation modeling is very popular in the study of almost any processes and systems, starting from COVID-19 modeling [1], and ending with the source of groundwater pollution [2]. There are publications in which simulation models of processes are based on the data clustering algorithm [3, 4].

There are three main and most important simulation methods: discrete event modeling [5-9], agent-based modeling [10-12], and system dynamics [13-15]. To study a particular system, the researcher determines which method to use.

An interactive educational platform for teaching teachers can be focused not only on the buyer, but also on the company's employees. This means that it depends on the specific consumer which model will be built. The functioning of the platform under consideration is a queuing system from the perspective of the consumer, so further research will be conducted in terms of the theory of queuing.

The purpose of the work is to study the client database of the company "InfoTeacher", identify the most promising and popular courses using cluster data analysis, create and study a simulation model of the distance education system to optimize the organizational structure of the site.

Modeling the distance education system is especially relevant during the pandemic and isolation, as the learning process allows you to improve your skills while staying at home.

2 Materials and methods

The educational platform of the company "InfoTeacher", which implements programs of advanced training and professional retraining for teachers throughout Russia, was considered as the object of study. On the main page of the site there is a list of categories into which the available courses are distributed: advanced training, retraining, subject, psychological and pedagogical education, for educators, extracurricular and extracurricular activities, additional education, intersubject courses. There, just below the categories, are colorful icons of the most popular courses.

The educational platform in question provides services for viewing, convenient search and ordering of the course, and also features a fully automated process for completing the entire training. The system is fully adapted to the specifics of Russian education, in particular, it meets the requirements of the Federal State Educational Standard (FSES) for the "Electronic Information and Educational Environment".

This study focuses on how to organize the work of managing the flow of applications for studying courses in order to eliminate internal queues that arise when serving the course curator.

Research problem:

- apply cluster data analysis using the R software environment and determine the most promising courses;
- develop a simulation model of the educational platform in the GPSS World language;
- identify problem areas in the functioning of the educational platform (formation of queues);
- eliminate the detected problems to improve the effectiveness of distance learning (determine the optimal number of course curators to minimize the queues for consultation with the curator, while the curators must be loaded by at least 50%).

Research methods.

To determine the most promising courses of the company "InfoTeacher", a cluster analysis of the client database was carried out using the R software environment. At the beginning, using the criterion of rocky scree, we calculated the optimal number of clusters,

and then applied the k-means method for further research. The results of the cluster analysis were used to determine the main input parameters for building a simulation model of an interactive distance learning platform. This model is a queuing system and was implemented in the GPSS World programming language. As a result of multiple runs of the program, the model parameters were selected that meet the criterion of system efficiency.

3 Results

To study the target audience of the site, we analyzed data obtained during the 6 months of the company's operation (Fig. 1). It should be noted that the company is quite young, but rapidly developing, and the sample size was about 200 entries, including self – selected courses on the site and with the help of a manager-by phone. It should be noted that the educational platform on the Internet works in full only for 4 months, and the head of the company is interested in its development and, as a result, increasing profits.

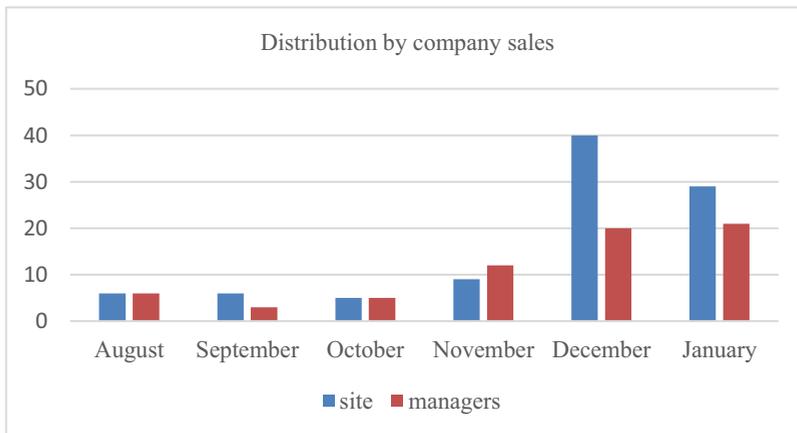


Fig. 1. Diagram of distribution of the company's sales

For a more detailed view of the consumers of this type of product (distance learning courses), we apply a cluster analysis of available statistical data.

Cluster analysis is a multidimensional statistical procedure that collects data containing information about a selection of objects, and then organizes the objects into relatively homogeneous groups.

The task of cluster analysis is to divide objects into groups (clusters) so that the objects in each group are somewhat similar. This reveals the internal structure of the data. The R software environment was used for cluster analysis.

First of all, we will find the optimal number of clusters according to the criterion of rocky scree. To do this, we will create a program code that allows you to visualize the results of the program in the form of a graph (Fig. 2).

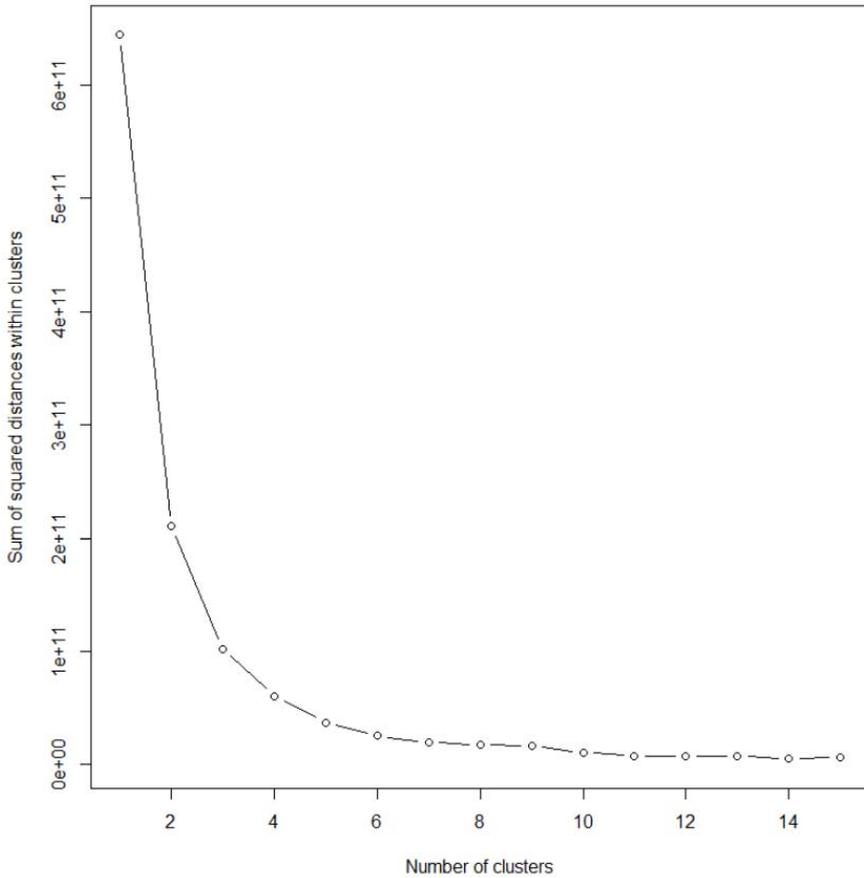


Fig. 2. Application of the scree test to determine the number of clusters

Figure 2 shows that the point at which the graph's decreasing rate changes corresponds to the value of argument 3, so it is better to use 3 clusters.

Next, we will continue the cluster analysis by the k-means method in the R software environment.

It is one of the most popular, non-hierarchical algorithms, due to its ease of implementation and speed of operation, but you need to know in advance the estimated number of clusters k .

Let's determine the average values of all the analyzed parameters in each of the clusters using the construction: `aggregate(d2, by=list(kc$cluster), FUN=mean)`. The result is shown in Figure 3.

Group.1	age	sum	managers	site	raznitsa
1	36.42308	1411.282	95083.33	67379.10	27704.231
2	43.56140	1855.789	135818.25	127115.26	8702.982
3	30.89796	2129.592	43846.73	20191.02	23655.714

Fig. 3. Average values of variables in each cluster

The resulting table shows the difference between entries in different clusters: the first cluster includes respondents whose average age is 36 years, they tend to spend less money

on classes, the difference between the transition from the site and from the manager is of the greatest importance.

The respondents of the second cluster are characterized by the highest average age-44 years, they are willing to pay for courses at an average level, the difference between the flows from the site and from managers is the smallest of all three clusters, both flows are quite large.

The respondents from the third cluster are characterized by a low average age – 31 years, they are more willing to pay for courses than other groups, the difference between flows from the site and from managers is quite large, both flows are small.

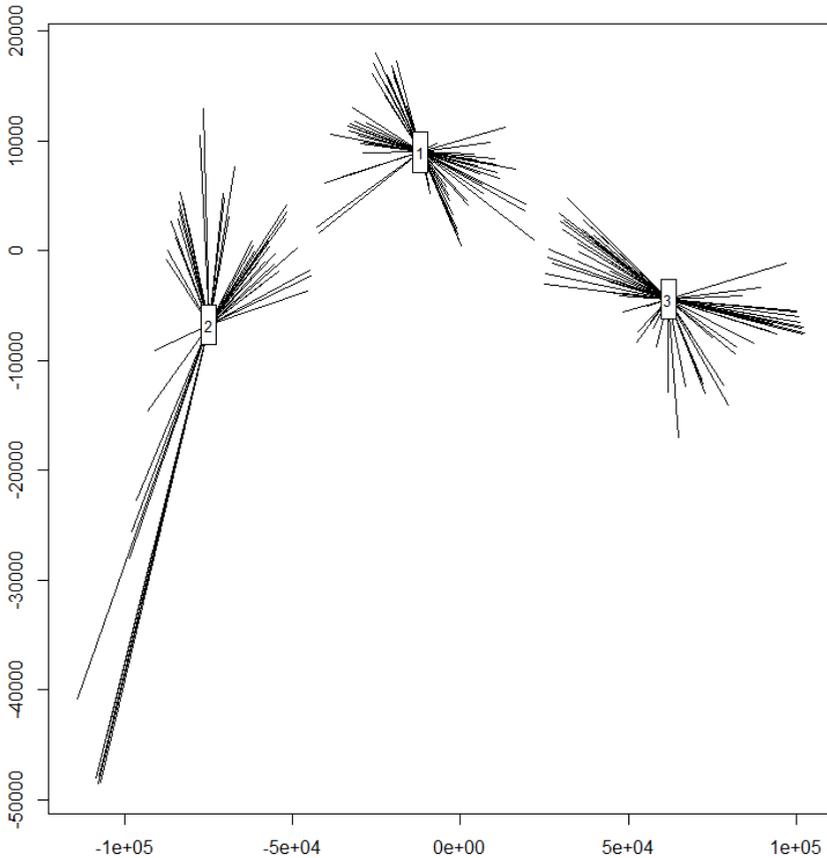


Fig. 4. The spread of the values of each cluster relative to the average value

Figure 4 shows how far the elements are located from the middle of each cluster.

Clusters 1 and 3 are characterized by greater proximity to the center, which indicates a greater uniformity of these groups.

You can see the clusters in more detail in the figure . To do this, use the following function in R: `clusplot(d2, kc$cluster, color=TRUE, shade=TRUE, labels=2, lines=0)`.

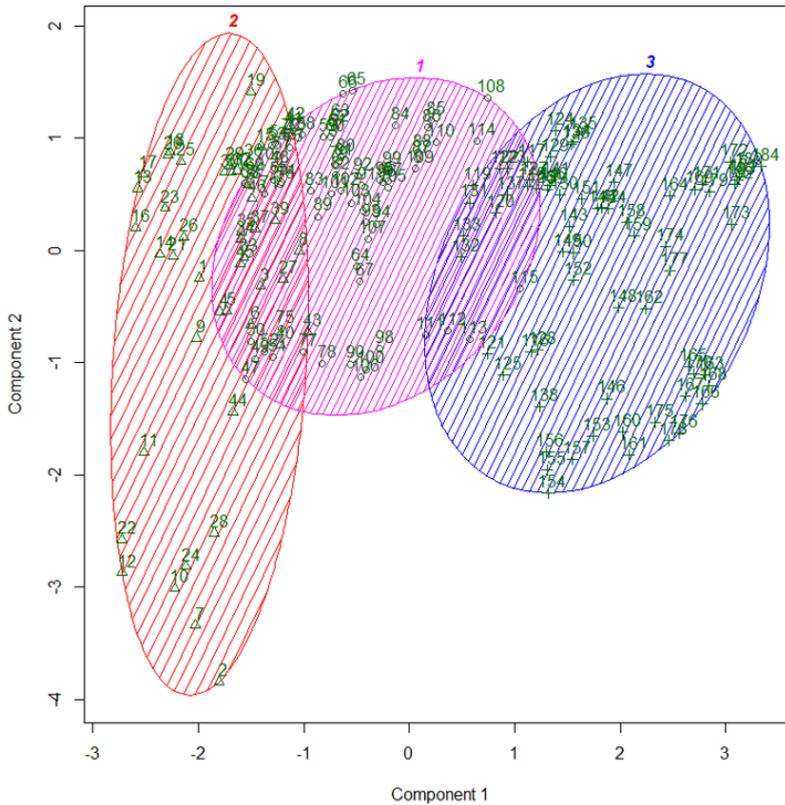


Fig. 5. Intersection of cluster elements

The shaded areas are blue and red and make up the fields of the parameters of objects belonging to different clusters (Fig. 5). Between the clusters there are areas of intersection, namely: between the first and second, first and third. This indicates a certain similarity of observations in these clusters. Note that the clusters are approximately evenly filled, but the first one is somewhat more numerous.

The first cluster is represented by middle-aged respondents, the cluster is the most numerous. It includes people with different interests and needs, according to age.

The second cluster is represented by respondents over 40 years old who want to update their knowledge and master innovative teaching methods, and they also learn English, which may be required both in connection with updating their knowledge in the specialty, and for mastering English-language literature for work.

The third cluster is represented by young people who are willing to pay more than others for courses, they are interested in the basics of first aid, pedagogy of preschool education and innovative teaching methods, which are basic courses.

It should be noted that the percentage of the explained variance of the original data set is 74%, which indicates a good quality of classification.

Cluster analysis of the data showed that the most popular are variations of the following courses: "Fundamentals of first aid", "Implementation of lean technologies in the activities of educational institutions in accordance with FSES", "Innovative methods and technologies of teaching the discipline in the implementation of the FSES", as well as refresher courses: "Pedagogy and methodology of preschool education" and "teaching English".

Let's move on to building a simulation model of the educational platform, using the information obtained.

The block diagram of the distance learning process is shown in Figure 6.

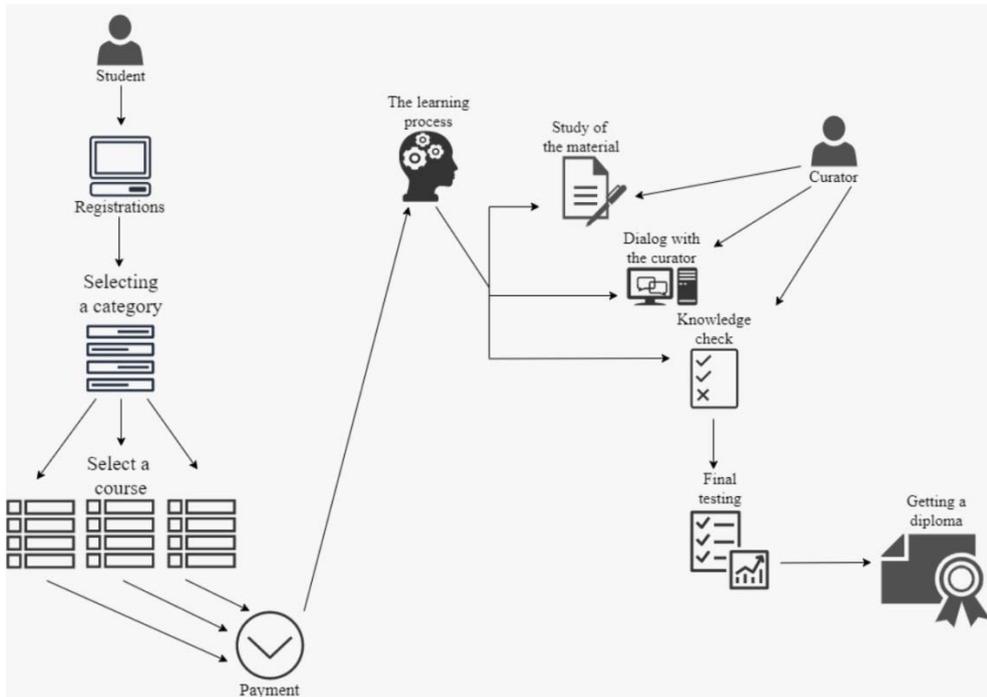


Fig. 6. The process of distance learning on the site

In this paper, as a criterion for the effectiveness of the site's functioning, we use the following: to determine the optimal number of course curators in order to minimize the queues for consultation with the curator, while the curators must be loaded by at least 50%.

The leader among all courses is "Fundamentals of first Aid" (more than 25% of applications). Despite the numeric data for the course "Implementation of lean technologies in the activities of educational institutions in accordance with FSES" (14% of applications), greater attention should be given to the courses: "Pedagogy and methodology of primary education", "Innovative methods and technologies of teaching disciplines in the implementation of the GEF" and "teacher education: the English language". These courses are the most promising, as they fall into the intersection of different clusters.

After we have decided on the courses on which the main profit of the company depends, we will move on to creating a simulation model of the distance learning system from the consumer's point of view.

We will assume that we have a multi-channel queuing system (QMS) with a uniform flow of requests and an unlimited storage capacity.

This model is stochastic, since the processes of receiving and servicing applications are random, that is, the time intervals between incoming applications and the duration of their service in the devices are random variables (students have different levels of training, so in the training process, the time for consultation is a random variable), described by the corresponding distribution laws.

As input parameters of the initial model, we consider the following: the number of customer requests, the number of curators and managers of the company (site), the intervals of receipt and service of customer requests were similar to the average statistical indicators taken from the company's database. The duration of advanced training courses is generally either 36 hours or 72 hours, and the duration of refresher courses is 504 hours. We believe

that the time intervals between the receipt of applications for training have an exponential distribution, the unit of model time is an hour. The service time of the majority of clients of the educational platform is subject to an even distribution. We considered a simulation model of the site functioning for a month (that is, 720 hours).

The purpose of the study: to change the original model so that it meets the criterion of effectiveness. The problem with the functioning of this model is that during the modeling process, internal queues may arise (queues for service from course curators).

In this paper, a program was developed that simulates the operation of the site in the GPSS World programming language.

Examining the created model, the program varied the values of the input parameters. First of all, the number of working devices changed (2, 3 or 4 curators). In addition, the behavior of the model depending on the intensity of the flow of input applications was studied.

The input flow of applications is divided into 9 main classes of applications. The most popular courses, based on the conducted cluster analysis, are more likely to appear in the program compared to the rest. In this case, these are classes 1, 3, 4 and 9. Classes are grouped by categories, which are assigned curators. They advise students on issues that arise during the training process. This type of activity for them is additional to the main work. Therefore, the curator can promptly advise a limited number of students. Now there are 2 curators working on the site, and they are already overloaded (Fig. 7). The DELAY column contains the number of refusals from serving requests by category. In the AVE TIME column – the waiting time in the service request queue.

QUEUE	MAX	CONT.	ENTRY	ENTRY (0)	AVE. CONT.	AVE. TIME	AVE. (-0)	RETRY
1	2	1	6	4	0.471	8.433	25.298	0
2	0	0	0	0	0.000	0.000	0.000	0
3	1	1	2	1	0.010	0.553	1.106	0
4	1	1	5	4	0.010	0.225	1.124	0
5	0	0	0	0	0.000	0.000	0.000	0
6	0	0	0	0	0.000	0.000	0.000	0
7	1	0	2	2	0.000	0.000	0.000	0
8	0	0	0	0	0.000	0.000	0.000	0
9	1	0	1	0	0.023	2.424	2.424	0

STORAGE	CAP.	REM.	MIN.	MAX.	ENTRIES	AVL.	AVE. C.	UTIL.	RETRY	DELAY
KATEGORY1	5	0	0	5	7	1	2.890	0.578	0	2
KATEGORY2	5	0	0	5	6	1	3.101	0.620	0	1
REGISTR	70	69	0	2	17	1	0.157	0.002	0	0
TESTING	70	68	0	3	3	1	0.786	0.011	0	0

Fig. 7. Report on the operation of the model in December 2020

If, as expected, the flow of applications increases, then, as can be seen in Figure 8, denial of service is inevitable.

QUEUE	MAX	CONT.	ENTRY	ENTRY (0)	AVE.CONT.	AVE.TIME	AVE. (-0)	RETRY
1	2	2	6	4	0.489	7.847	23.540	0
2	0	0	0	0	0.000	0.000	0.000	0
3	2	1	4	1	0.536	12.883	17.177	0
4	1	1	4	3	0.260	6.254	25.017	0
5	0	0	0	0	0.000	0.000	0.000	0
6	0	0	0	0	0.000	0.000	0.000	0
7	1	0	1	1	0.000	0.000	0.000	0
8	0	0	0	0	0.000	0.000	0.000	0
9	2	0	3	1	0.447	14.323	21.485	0

STORAGE	CAP.	REM.	MIN.	MAX.	ENTRIES	AVL.	AVE.C.	UTIL.	RETRY	DELAY
KATEGORY1	5	0	0	5	7	1	3.412	0.682	0	3
KATEGORY2	5	0	0	5	7	1	3.928	0.786	0	1
REGISTR	100	100	0	3	18	1	0.187	0.002	0	0
TESTING	120	117	0	4	4	1	0.881	0.007	0	0

Fig. 8. Report on the operation of the model when the input stream increases

For the most popular courses, queues and refusals of service are formed (1, 3, 4, 9 classes of applications), which is unacceptable. You need to add more course curators.

QUEUE	MAX	CONT.	ENTRY	ENTRY (0)	AVE.CONT.	AVE.TIME	AVE. (-0)	RETRY
1	1	1	7	5	0.287	4.006	14.019	0
2	1	0	1	0	0.151	14.734	14.734	0
3	1	1	2	1	0.002	0.110	0.220	0
4	1	0	5	4	0.107	2.090	10.452	0
5	0	0	0	0	0.000	0.000	0.000	0
6	0	0	0	0	0.000	0.000	0.000	0
7	1	0	1	1	0.000	0.000	0.000	0
8	0	0	0	0	0.000	0.000	0.000	0
9	1	0	3	2	0.113	3.672	11.016	0

STORAGE	CAP.	REM.	MIN.	MAX.	ENTRIES	AVL.	AVE.C.	UTIL.	RETRY	DELAY
KATEGORY1	6	0	0	6	8	1	3.904	0.651	0	2
KATEGORY2	6	0	0	6	8	1	4.074	0.679	0	0
KATEGORY3	6	6	0	1	1	1	0.369	0.062	0	0
REGISTR	100	100	0	3	19	1	0.195	0.002	0	0
TESTING	120	116	0	5	5	1	1.072	0.009	0	0

Fig. 9. Report on the operation of the model with one added curator

Adding one more curator and increasing the load of each one by 20% leads to a greater number of requests being fulfilled, but there are still refusals (Fig. 9). If you add another curator, we get the following situation, as in Figure 9.

QUEUE	MAX	CONT.	ENTRY	ENTRY (0)	AVE.CONT.	AVE.TIME	AVE. (-0)	RETRY
1	1	0	6	5	0.024	0.396	2.378	0
2	1	0	1	0	0.151	14.734	14.734	0
3	1	1	2	1	0.002	0.110	0.220	0
4	1	0	5	5	0.000	0.000	0.000	0
5	0	0	0	0	0.000	0.000	0.000	0
6	0	0	0	0	0.000	0.000	0.000	0
7	1	0	2	2	0.000	0.000	0.000	0
8	0	0	0	0	0.000	0.000	0.000	0
9	1	0	3	3	0.000	0.000	0.000	0

STORAGE	CAP.	REM.	MIN.	MAX.	ENTRIES	AVL.	AVE.C.	UTIL.	RETRY	DELAY
KATEGORY1	6	0	0	6	8	1	3.904	0.651	0	1
KATEGORY2	6	1	0	5	5	1	3.297	0.549	0	0
KATEGORY3	6	5	0	1	2	1	0.632	0.105	0	0
KATEGORY4	6	5	0	3	3	1	0.996	0.166	0	0
REGISTR	100	100	0	3	19	1	0.195	0.002	0	0
TESTING	120	116	0	5	5	1	1.072	0.009	0	0

Fig. 10. Report on the operation of the model with two added curators

It can be seen that categories 3 and 4 are underloaded by 10% and 16%, respectively, but one of the conditions for the effectiveness of the system is the workload of performers by more than 50%. Therefore, this solution is not suitable for us.

The company's management decided to redistribute the courses between the categories in order to eliminate the queues that arise for consultations with curators.

After running the program for modeling the operation of the distance learning platform with the changed input parameters, we received the following report (Fig. 11).

QUEUE	MAX	CONT.	ENTRY	ENTRY(0)	AVE. CONT.	AVE. TIME	AVE. (-0)	RETRY
1	1	0	7	7	0.000	0.000	0.000	0
2	1	0	1	1	0.000	0.000	0.000	0
3	1	0	1	1	0.000	0.000	0.000	0
4	1	0	5	5	0.000	0.000	0.000	0
5	0	0	0	0	0.000	0.000	0.000	0
6	0	0	0	0	0.000	0.000	0.000	0
7	1	0	3	3	0.000	0.000	0.000	0
8	0	0	0	0	0.000	0.000	0.000	0
9	1	0	2	2	0.000	0.000	0.000	0

STORAGE	CAP.	REM.	MIN.	MAX.	ENTRIES	AVL.	AVE.C.	UTIL.	RETRY	DELAY
KATEGORY1	6	0	0	6	7	1	3.345	0.558	0	0
KATEGORY2	6	3	0	6	6	1	3.128	0.521	0	0
KATEGORY3	6	4	0	4	6	1	2.081	0.347	0	0
REGISTR	100	100	0	3	19	1	0.195	0.002	0	0
TESTING	120	113	0	8	8	1	1.525	0.013	0	0

Fig. 11. Report on the operation of the model with redistributed courses between categories

The simulation results show that with an increasing flow of requests, all requests will receive the necessary service. And the category curators will be quite loaded.

As for the effectiveness of the simulation model of the educational platform, it would be interesting to find out how the behavior of the system will change in the case of a non-uniform incoming flow.

4 Conclusions

Cluster analysis is widely used in marketing to segment consumers to determine the approach to each group. With the help of it, the most promising and popular courses of the educational platform were identified. In this study of the client database, the intersection between the first and second, first and third clusters was obtained due to such a discipline as "Innovative methods and technologies of training in the context of the implementation of the Federal State Educational Standard". Since innovative teaching methods have become particularly important at the moment, all categories of respondents were interested in this course. Accordingly, we can assume that the page of this course should expect the greatest flow of consumers.

The model of the interactive educational platform was implemented as a stochastic simulation model of queuing using the universal modeling environment GPSS World. This work allows us to solve a very urgent problem – reducing internal queues for consultation with the course curator. In addition, an effective system meets the needs of students, supervisors, and company management. Since the result of the work is a software product, it can be used to predict the behavior of the model when changing the input parameters, to form an optimal management strategy for other similar systems.

The results of the study can be used by most companies engaged in business related to distance learning.

References

1. S. Annas, M. I. Pratama, M. Rifandi, W. Sanusi, S. Side, *Chaos, Solitons & Fractals* **139**, 110072 (2020) doi:10.1016/j.chaos.2020.110072

2. H. Wang, W. Lu, J. Li, *Journal of Contaminant Hydrology* **234**, 103681 (2020) doi:10.1016/j.jconhyd.2020.103681
3. Z. Jin, Z. Liang, Y. Wang, W. Men, *Computer Communications* **166**, 262-270 (2021) doi:10.1016/j.comcom.2020.11.003
4. Y. Liu, X. Guo, J. We, H. Zhang, *Journal of Petroleum Science and Engineering*, 108422 (2021) doi:10.1016/j.petrol.2021.108422
5. M. Saghi-Jadid, H. Ketabchi, *Journal of Hydrology*, 125709 (2020) doi:10.1016/j.jhydrol.2020.125709
6. T. Grober, O. Grober, O. Savchenko, *E3S Web of Conferences* **210**, 11002 (2020) doi:10.1051/e3sconf/202021011002
7. D. Marek, A. Domański, J. Domańska, T. Czachórski, J. Klamka, J. Szygła, *Computer Communications* **166**, 40-48 (2021) doi:10.1016/j.comcom.2020.11.014
8. M. Camargo, M. Dumas, Oscar González-Rojas, *Decision Support Systems* **134**, 113284 (2020) doi:10.1016/j.dss.2020.113284
9. A. S. Elshall, M. Ye, M. Finkel, *Journal of Hydrology* **590**, 125427 (2020) doi:10.1016/j.jhydrol.2020.125427
10. Y. Sun, H. Liu, *Physica A: Statistical Mechanics and its Applications* **566**, 125652 (2021) doi:10.1016/j.physa.2020.125652
11. Li Kou, W. Fan, S. Song, *Computers & Electrical Engineering* **86**, 106744 (2020) doi:10.1016/j.compeleceng.2020.106744
12. J. Hyuck, C. Jinwoo Oh, H. Lee, *International Journal of Heat and Mass Transfer* **160**, 120222 (2020) doi:10.1016/j.ijheatmasstransfer.2020.120222
13. X. Hao, J. Yan, J. Sha, W. Gengyu He, Ci Song, Y. Ma, G. Zhang, *Journal of Cleaner Production* **287**, 125062 (2021) doi:10.1016/j.jclepro.2020.125062
14. D. Li, L. Deng, X. Zeng, Z. Cai, *Microprocessors and Microsystems* **83**, 104009 (2021) doi:10.1016/j.micpro.2021.104009
15. G. Guizzi, D. Falcone, F. De Felice, *Computers & Industrial Engineering* **137**, 106052 (2019) doi:10.1016/j.cie.2019.106052