The Hierarchical Information System for Management of the Targeted Advertising

Karina Melnyk[1], Natalia Borysova[1] and Viktoriia Melnyk[2]

[1] *National Technical University "Kharkiv Polytechnic Institute", Kirpichova street, 2, Kharkiv, 61002, Ukraine*
[2] *Kharkiv general education school of I-III degrees № 145, Amosova street, 24a, Kharkiv, 61171, Ukraine*

Abstract
The problems of target audience identification and user segmentation for managing a process of the targeted advertising to customers are considered. It has proposed to combine the solution of these two tasks within the framework of one information system. An overview of the existing methods for identifying and segmenting of the target audience are presented. In addition, the existing applications and tools for solving the given problem are considered. The formalization of these tasks is presented. The functional models of business processes corresponding to the identifying and segmenting of the target audience are developed. The architecture of the information system is proposed. It is presented in the form of a deployment scheme, a database model is developed. The results of numerical studies and evaluation of the effectiveness of the developed information system are presented.

1. Introduction

Profitability of any B2B or B2C company depends on many factors. One of the important one is advertising, which presents a product or service to a potential buyer for the selling purpose. Development of information technology has led to possibility to identify the target audience of customers and to influence on this audience using targeted advertising. Many various methods of Market Research are used to determine the target audience. Depending on the needs of the market, the strategic goals of an enterprise and characteristics of the advertised product or service, all potential buyers are divided into groups. There are several cases of dividing the customers into groups. Manager can create two groups: included in the target audience and not included in the target audience; three groups: primary target audience, secondary target audience, not included in the target audience; four groups: core, primary target audience, secondary target audience, not are included in the target audience. The advantages of determining the target audience are obvious:

- development of a more effective advertising campaign is a reason of saving the advertising budget;
- increasing of a brand awareness and loyalty of clients contribute to a class of regular customers;
- quick return on investment;
- a clearer understanding of the needs of their customers, it means more effective interaction with them, etc.

However, to improve the effectiveness of an advertising campaign, manager can divide the target audience into segments based on the personal characteristics of the clients. The customer segmentation models are used in this case. There are many benefits of this process. The segmentation

process can be used for promotion of a certain product to a group of buyers from the target audience. Other advantages of using the segmentation models are following: drawing up personalized price proposals, improving customer experience, searching for ideal customer, reducing the customer churn, prices optimization, searching for new market opportunities, etc.

Therefore, the purpose of this research is to develop an integrated hierarchical information system (HIS) that would combine the functions of systems for determining the target audience and a system of user segmentation for the distribution of targeted advertising.

2. Formal problem statement

The Hierarchical Information System is a system that solves tasks stage-by-stage. The first task is to identify the target audience. It is a classification task since it divides customers into several predefined groups. Input information is a set of characteristics of potential buyers. To solve the problem, it is necessary to do:
- analyze the domain area for creating a portrait of buyer persona;
- undertake review of approaches and methods for determining the targeted audience;
- formalize a method for the resolving the task.

The second level of the hierarchical management system is designed to solve the problem of segmentation of the target audience. This task is a clustering task and belongs to the class of unsupervised learning tasks. Potential customers should be grouped into clusters in such a way that objects from one cluster are closer to one another than objects from other clusters by any criterion. To solve the clustering problem, it is necessary to do:
- analyze methods and applications to solve the clustering task;
- formalize proposed method;
- perform numerical study;
- evaluate the proposed approach.

3. Overview of existing models, methods and applications to solve the given issue

This study proposes to solve the given issue in two stages. Therefore, it is necessary to conduct review the existing methods and applications separately for each stage: for identifying the target audience and for customer segmentation. There are many generally accepted models in direct marketing, which can be used for both target audience identifying and customer segmentation. The usage of these models depends on user's data (Table 1) [1-4].

Table 1
Models for target audience identifying and customer segmentation

Model name	Characteristics
Demographic based	Age, gender, occupation, income, education, marital status, ethnicity, race, religion, profession or role in the company
Geographic based	Continent, country, region or state, city, district, postal code, timezone
Psychographic based	Social class, lifestyle, personality, values, presence in digital and/or social media space, personal convictions, beliefs, attitude, interests
Technographic based	Usage of devises, applications and software
Behavioral	Habits, spending, consumption, usage and desired benefits, usage, loyalty, awareness, types of payment, demands, quality fanatics, price and/or brand sensitiveness

In addition, it is possible to use multiple models or mixed models.

3.1. Review of target audience identifying methods and applications

Any classification method can be used for resolving of the identifying task of the target audience. For example, Naive Bayes, logistic regression, Support Vector Machine, k-nearest neighbor, decision trees etc. allow dividing objects into different groups [5-9]. The set of buyers is divided into two classes: included in the target audience and not included in the target audience. The classification signs are different characteristics of the portrait of an ideal buyer. The portrait has built before the launch of an advertising campaign for a product or service and can be adjusted. The characteristics correspond to one or more used models or a mixed model (Table 1). The best case when a marketing specialist who understands the intricacies of promoting a company's product or service draws up the portrait. She/he can use various approaches, for example, Mark Sherrington's "5W" approach. 5W means answering such questions: What? (type of product or service we sale); Who? (our customer); Why? (motivation for buying); When? (conditions for buying); Where? (place of buying). There are alternative marketing approaches to the 5W method: the method "from the opposite", the method "from the product", the method "from the market", and the method "from the target" [1]. The result of using any of the marketing methods is a set of values for the characteristics of an ideal buyer. This is the most important and crucial stage for the further success of the advertising campaign. It influences on the size of the company's profit. An interesting tool from HubSpot named Make My Persona [10] can come in handy when drawing up a portrait of an ideal buyer. This tool allows building some virtual image of buyer personas in seven simple steps:
1. Creating Personas' avatar and choosing Personas' name.
2. Identifying Personas' demographic characteristics, such as age and level of education.
3. Identifying Personas' business, such as working industry and size of company.
4. Identifying Personas' careers, such as their job title, their job measuring, their boss.
5. Identifying Personas' job characteristics, such as their goals or objectives, their biggest challenges, their job responsibilities
6. Identifying Personas' lovely tools to do their job and to communicate with vendors and other businesses.
7. Identifying Personas' consumption habits, such as their lovely social networks and their training in their job.

After answering these questions, the user is taken to a page where she/he can see her/his Buyer Personas 'Overview, and can also expand it by adding own fields to this Overview. It can be saved, downloaded or shared on social networks after filling out a special pop-up form.

Existing applications, services, tools and software, which are intended for target audience identifying, for example, such as Google Analytics, Facebook Insights, Twitter Followers Dashboard and others can only be used to analyze the company's existing customer data, creating various analytical reports, visualizing analysis results, making forecasts, tracking the activity of regular customers, inflow and outflow of customers, etc. However, all of these feature-rich applications, services, tools and software have "a cold start" problem. They cannot be used for analysis in the absence of data. For example, when some startup is launched and the portrait of the ideal buyer has already been built, manager have to create advertising, but target audience is unknown. Obviously, the target audience identifying task has not been completely solved, and in some cases it has not been solved at all.

Thus, in order to identify potential buyers, it is necessary to form a portrait of the buyer persona and compare it with the considered objects. There are many ways of comparing objects. For instance, manager of a company can perform calculation of similarities between objects. Input data of buyer persona and potential customers are information of mixed type. Therefore, to calculate a similarity between them, are encouraged to use corresponding metrics. It can be Voronin similarity measure, Zhuravlev metric, Gower coefficient etc. [11].

3.2. Review of customer segmentation methods and applications

After identifying of the target audience, HIS can start customer segmentation. Manager can use rule-based methods or cluster analysis methods for customer segmentation. The use of rule-based

methods implies the creation of rules or the selection of a priori thresholds for strict customer segmentation. Such way of segmentation can lead to situation, when customers from one group have significant differences. It is also quite difficult to perform segmentation in more than two dimensions. In addition, the segmentation results are more consistent with the initial assumptions of the marketer and do not always reveal significant differences between customers. In this sense, cluster analysis methods show higher efficiency in comparison with rule-based methods. Since they are unsupervised machine learning methods, they do not need a training sample; they can work directly with the input data without prior training. These methods are more practical, they divide the training set of customers into more homogeneous groups, within the differences between customers are very small, in addition, cluster analysis methods allow to conduct the dynamic clustering, which is fully reflect the state of the available data at a given time [12].

There are special applications and software for customer segmentation on the market. For example, Segmentor the customer segmentation tool from Optimove [13], CleverTap [14], HubSpot [15], Experian [16], SproutSocial [17], Qualtrics [18], MailChimp [19] etc. In addition, there are applications and software for customer segmentation that only use the behavioral model. For example, Yieldify [20], Amplitude [21], Indicative [22], Mixpanel [23]. All of them certainly have their own advantages and perform the function of customer segmentation using one or several models. However, not all of them are free, but it is important for novice businesspersons or startups. In addition, there is no description of the used algorithms and methods, which is not will allow the user to double-check the obtained results and may lead to erroneous conclusions.

3.3. Review of targeted advertising management applications and services

After segmentation of the target audience, it is necessary to prepare special advertisements for each group of customers and send them out. For this, it is advisable to use special targeted advertising management applications and services. The article [24] provides a brief description of some of these services. Of course, there is a huge number of such applications and services, there are free and paid ones, they have different functionality, and some even solve the problem of finding the target audience in addition. Nevertheless, none of them solves three problems: target audience search, target audience segmentation and targeted advertising delivery.

4. Development of the hierarchical information system
4.1. Formalization of the management process of the targeted advertising

Let's consider designing and using the hierarchical information system to resolve the management task of the Targeted Advertising in a more detailed way. To solve the problem of determining the targeted audience and the segmentation of potential users, it is necessary to develop a functional model of the business process for managing targeted advertising delivery. For this, it is proposed to use the IDEF0 methodology (Figure 1).

The first step is marketing research. It allow to find, to collect and to analyze the received information for reducing the uncertainty in making managerial decisions. For example, it is important for a new startup to find a product or service that can be successfully implemented. For existing companies, it is possible to assess the prospects for demand for a specific product or service based on the study of consumer behavior. There are many methods to perform market research: observation, surveys, focus groups, personal interviews etc. Each method has its own advantages, disadvantages, limitations and is capable of providing information of varying completeness and accuracy.

The goal of the next stage is to highlight informative features that will fully reflect the portrait of a potential buyer. Next, a profile of ideal buyer persona is created. Defining a buyer or audience persona helps to create product or service to better target ideal customer. Depending on the final product is designed for, the appropriate templates are used: B2B or B2C Buyer Persona templates, as well as the results of marketing research.
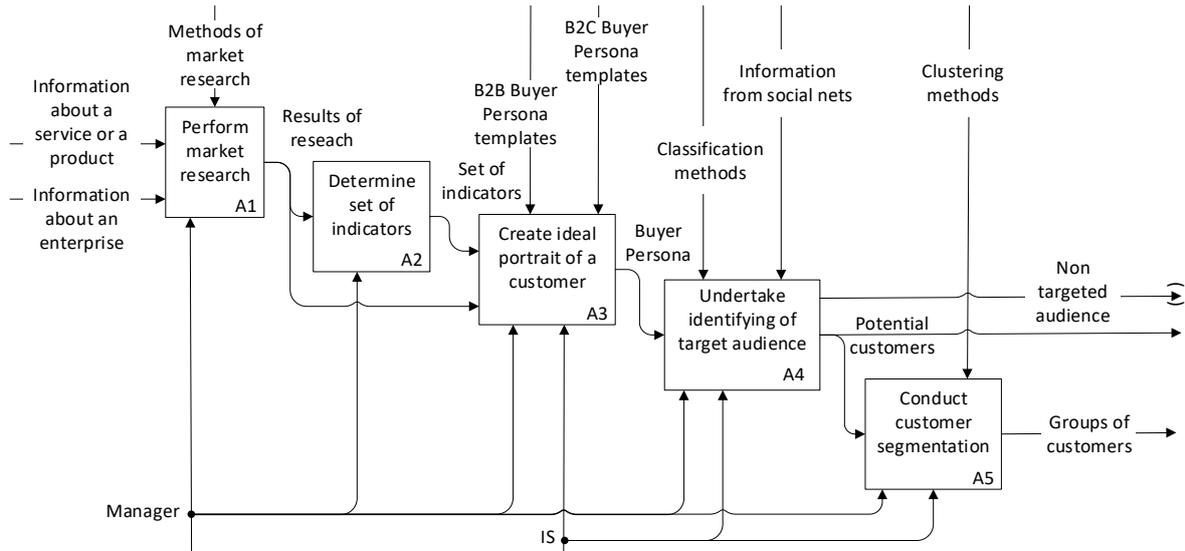
Figure 1: The model of Management of the Targeted Advertising

The next step is to split buyers into two groups: targeted audience and non-targeted audience. Various classification methods can be used for this. According to the above analysis, the paper proposes to use an approach based on calculating a measure of similarity between an ideal customer and a potential customer. To improve the effectiveness of an advertising campaign, the target audience can be segmented depending on the needs of potential customers. It is necessary to solve the clustering task according to the selected indicators. For example, if you advertise certain sport clothes, then the younger generation will choose bright colors, and the middle age will prefer convenience to appearance. In general, the task of segmenting potential customers can be reduced to the task of determining the target audience. The expert for each group sets its own boundary values of the similarity measure. Thus, she/he can create several target groups: core audience, several groups with different values of similarity measures, non-targeted people. This approach can be applied in the case of limited finances. However, the expert can miss significant differences between clients, therefore, for a more efficient segmenting process, it is recommended to use automatic segmentation, namely clustering methods, since these are methods of unsupervised learning.

4.2. The model of resolving the identifying task of the target audience

Consider the task of identifying potential customers. The model for solving the task is described by the following activity diagram in Figure 2.

Let K be the set of clients selected by the manager to determine the value for the advertising company. The input data for the task of identifying the target group is an ideal customer profile. Let's specify F as a set of indicators of a potential client, and F_i $(F_i \in F)$ is a set of values of the i-th indicator. Then x_i^b and x_{ki} $(k \in K, \ i \in F_i)$ are the value of the i-th indicator of profile of buyer persona and of k-th client accordingly.

Let's designate s_k $(k \in K)$ as a measure of similarity between the profile of an ideal client and k-th client, then s_{ki} is a similarity in the по i-th indicator. Data about a potential customer can be qualitative data, in the form of categories, and in the form of numbers. For the simultaneous processing of such data, there are special proximity measures. Let us consider the calculation of the similarity on the example of using the Gower coefficient [11, 25]. To calculate the measure, it is necessary to turn qualitative data into categorical data, and then use formula (1) for them:

$$s_{ki} = \begin{cases} 1, x_i^b = x_{ki} \\ 0, x_i^b \neq x_{ki} \end{cases}. \qquad (1)$$
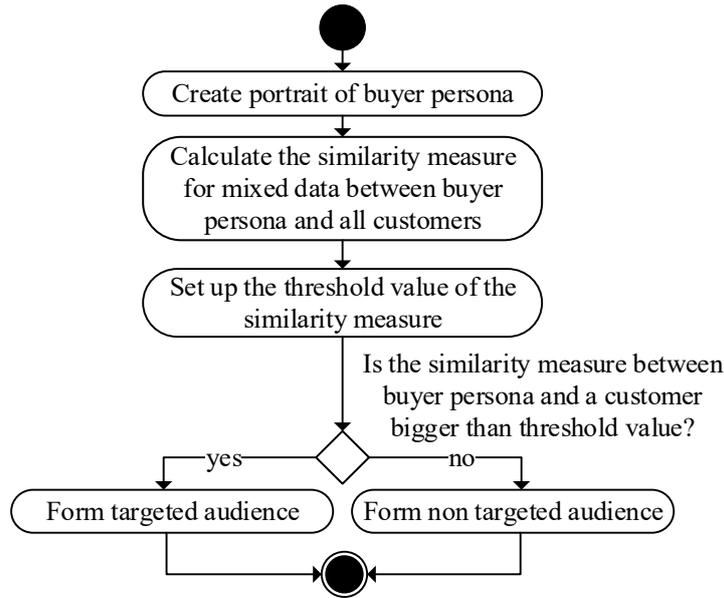
Figure 2: The model of the identifying process of the targeted audience

If the indicators of the portrait of audience persona are quantitative, then formula (2) is applied:

$$s_{ki} = 1 - \frac{\left|x_i^b - x_{ki}\right|}{\max\{x_i^b\} - \min\{x_i^b\}}. \tag{2}$$

To calculate the Gower coefficient s_k according to formula (3), it is necessary to determine w_i is the coefficient of the presence of the i-th indicator for the client: $w_i = 0$, if information about the indicator is absent in the profile of the ideal client or potential client, it equals to 1 if all the information is available .

$$s_k = \frac{\sum_i w_i s_{ki}}{\sum_i w_i}. \tag{3}$$

Next, the manager determines the similarity values that are acceptable for the target group of buyers and forms the corresponding sets. The obtained information allows formulating solutions for the implementation of an advertising campaign.

4.3.   The model of resolving the segmenting task

Review of related works according to resolving the segmenting task has showed the feasibility of using the clustering methods. Clustering process is a machine learning technique that groups of objects according to chosen indicators. Modern science knows a lot of clustering methods: Affinity Propagation, Balanced Iterative Reducing, K-Means Clustering, Clustering using Hierarchies, Mean shift clustering, Agglomerative Hierarchical Clustering, Expectation–Maximization Clustering using Gaussian Mixture Models etc. [26, 27]. One of the simplest and most commonly used method is the K-Means Clustering method. Let's consider a model for solving the task of segmenting a targeted audience using the proposed method (Figure 3).

Let the manager has a hypothesis about the number n of clusters or segments. It can be based on theoretical considerations or the results of marketing research. If the assumptions are not obvious, then a series of experiments with different numbers of clusters can be performed to find the optimal partition.

The K-Means algorithm starts with randomly selected clusters and then reassigns objects to them to minimize intra-cluster variability and maximize inter-cluster variability. The main drawback of the K-Means algorithm is that it only works with numeric values. The input customer information is mixed information. Therefore, in this paper, it is proposed to use the Gower coefficient to calculate the distances between the centers of the clusters and the objects, since it works with mixed data.
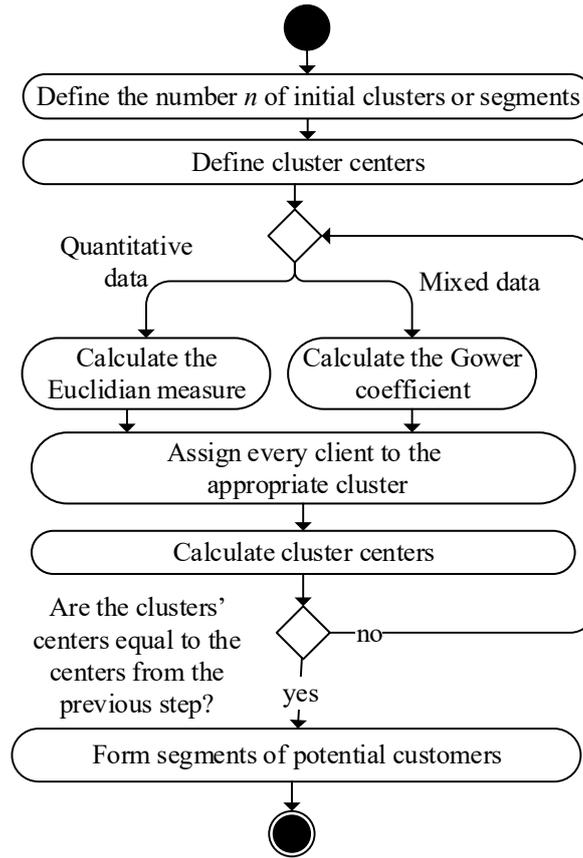
Figure 3: The model of the segmenting process of the targeted audience

The algorithm of K-Means is following. The center of each cluster is randomly determined. Denote $c_q(\{x_i\})$ $(q = \overline{1,n})$ as the center of q-th cluster with the set of values of all indicators $\{x_i\}$. Then $\|C_q\|$ is the cardinality of the set of objects of the q-th cluster. Then it is necessary to calculate the distance between the centers of the clusters and each object s_{qt} $(t \in T)$ according to formulas (1)-(3), where T is the set of potential buyers or target audience obtained in the previous step, and t is a specific buyer from this set. The object or client is assigned to the closest cluster

$$t_q = \mathrm{argmax}_{q=\overline{1,n}}\{s_{qt}\}, \tag{4}$$

where t_q is designation of belonging of t-th customer to q-th cluster.

After calculating the distances and assigning objects to the new cluster, it is necessary to find the coordinates of the new center for each cluster. The new values of each indicator of the center of the cluster are found based on the use of the formula for the arithmetic mean of all values of the i-th indicator of objects that belong to the current cluster:

$$c_q(\{x_i\}) = \frac{1}{\|C_q\|}\sum_{t \in C_q} x_{ti}, \; i \in F_i. \tag{5}$$

Next, the algorithm again calculates the distance from each object to the cluster centers using formulas (1)-(3) and assigns the objects to the nearest cluster using formula (4). The centers of gravity of the clusters are calculated again according to (5). This process is repeated until the centers of gravity stop "migrating" in space.

Thus, the model for target audience segmentation has been proposed.

4.4.   Architectural solution of the Information System

The HIS design process is revised from the development of the software requirement specification (SRS). All requirements from SRS are divided into functional and non-functional ones. Functional

requirements are subdivided into business requirements, user requirements, and system requirements. Non-functional requirements describe how the HIS should work and which properties and quality attributes it should have. The main requirements described in the SRS are presented in the form of Requirement Diagram (Figure 4).
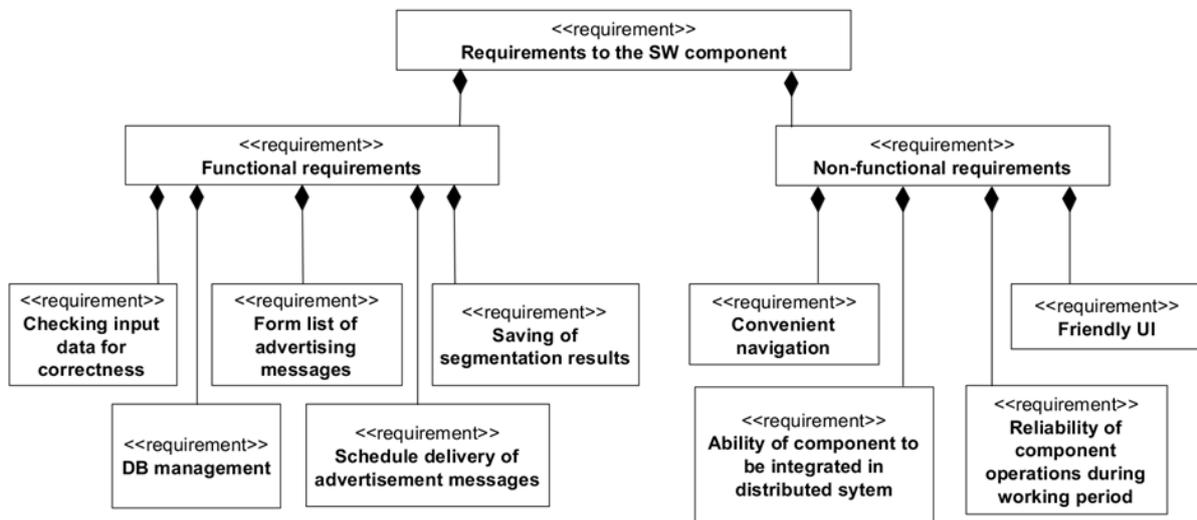
Figure 4: The Requirements Diagram for the HIS

The list of functional requirements for the system is following:
- The HIS has to verify the correctness of the entered data.
- The HIS has to generate a list of advertising messages and advertising objects or generate an error in the process of creating advertising messages.
- The HIS has to create the advertising messages and customer groups after the customer segmentation process is completed.
- The HIS has to provide access to all necessary data and documents (list of product categories, information on discounts, rules of message formation, advertising budget, segmentation rules, list of customer wishes, description of the segmentation method, list of customer preferences) to the advertiser at the stage of formation of advertising messages.
- The HIS has to provide access to the database (Products_DB, Customers_DB etc.) at any time.
- The HIS has to provide opportunities to work with files that have been created in other systems.
- The HIS has to allow the advertiser to suspend his/her work and save to a file, then to resume his/her work from the saved file.
- The HIS has to send advertising messages at the scheduled time or allow the system user to do so manually.

The result of turning out the requirements into a structured solution that meets both technical and business requirements is an architectural solution for HIS. Deployment Diagram is used to visualize the topology of the physical components of the system. The proposed architectural solution for The Hierarchical Information System for the Management of Targeted Advertising in the form of a Deployment Diagram is shown in Figure 5.

This research proposes to use the "client-server" architectural pattern for the development of the HIS architecture. Such architecture allows sharing the data processing function to several separate servers. It separates the functions of storing, processing and presenting data for more efficient use. The presentation component or a "client" is responsible for the user interface. Application logic is executed at the middle level of the architecture, namely the application server layer. It provides data exchange between users and databases. The middle layer is split into two separate components to improve the performance of the HIS:
- IIS Application Server is responsible for the application's logic;
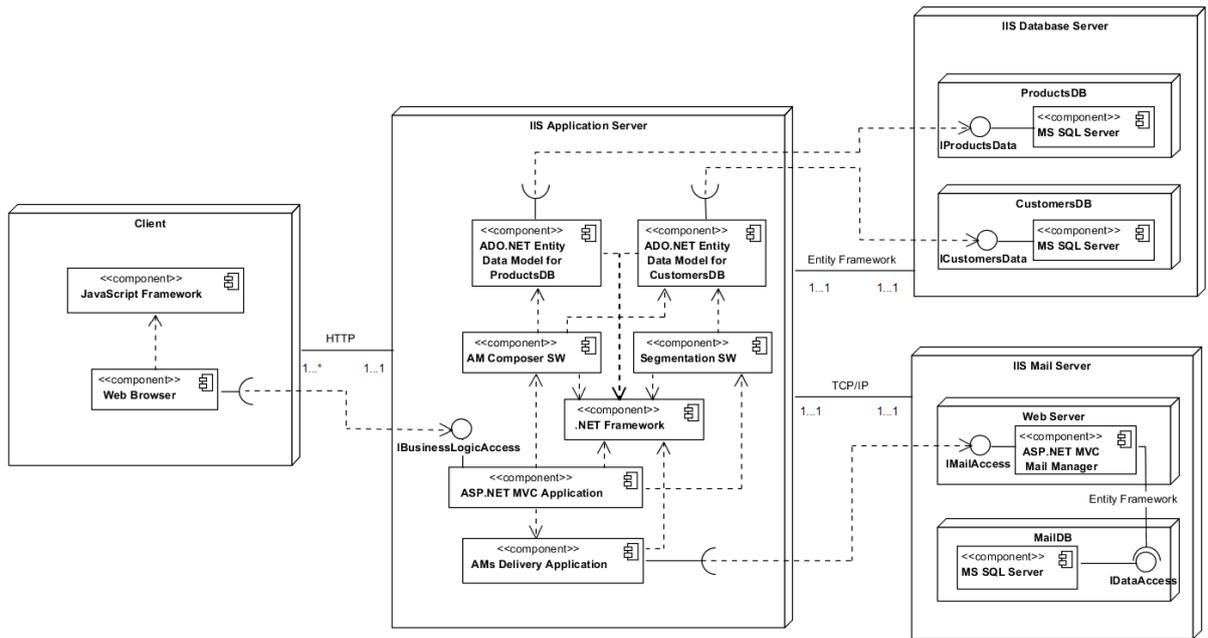- IIS Mail Server is devoted to process advertising messages.

Figure 5: The Architecture of the HIS

The data layer is designed to store and manage the information processed by the HIS. Here is a database that allows to implement the interpretation of information about the domain area in the form of formalized data in accordance with certain requirements (Figure 6).
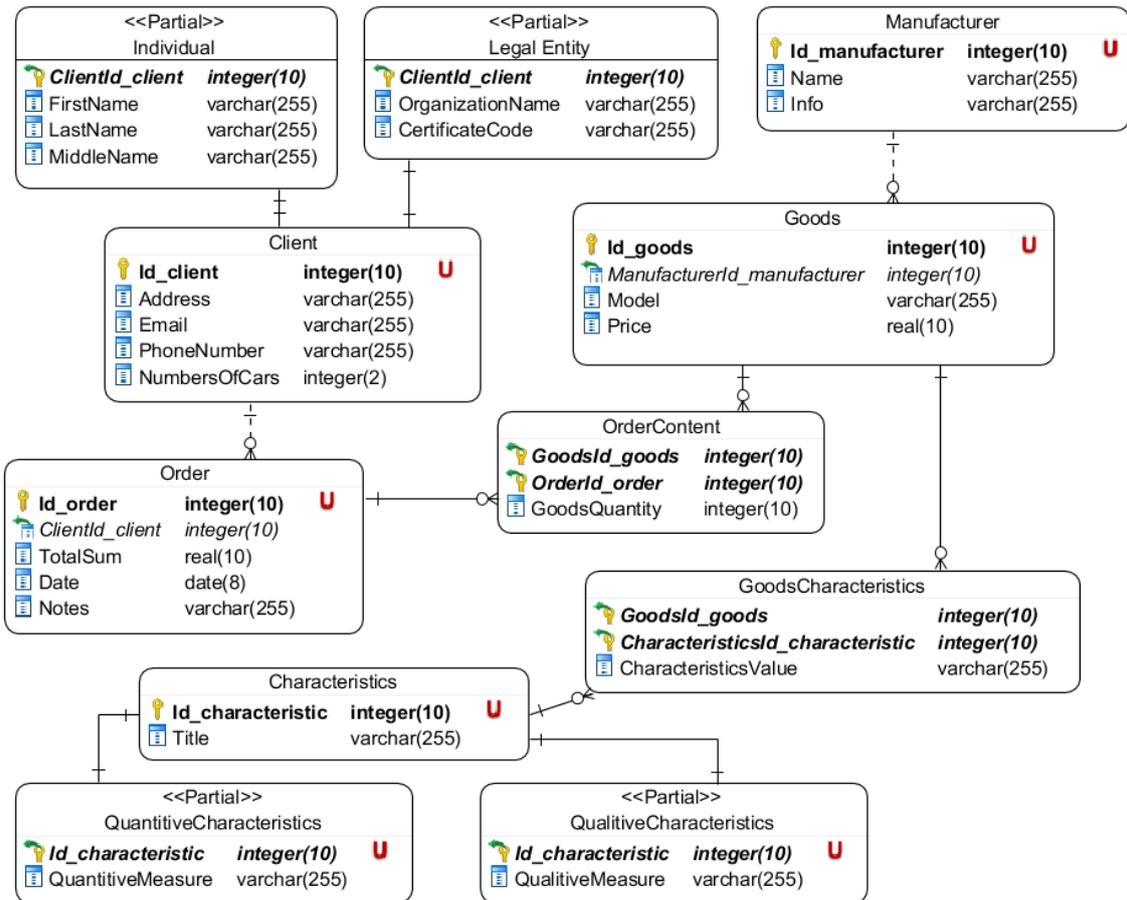

Figure 6: The Database structure

The structure of the developed database consists of 11 entities:
- The entity of "Clients" is the company's customers.
- The categorical entity "Individual" describes common clients or people.
- The categorical entity of "LegalEntity" describes a legal entity.
- The entity of "Order" is a table with orders.
- The entity of "Manufacturer" is a list of producers of goods.
- The entity of "Goods" is a list of goods.
- The entity of "OrderContent" is the content of the order.
- The entity of "Characteristics" is a list of product characteristics.
- The categorical entity of "QuantitiveCharacteristics" is a list of quantitative characteristics of the product.
- The categorical entity of "QualitiveCharacteristics" is a list of quality characteristics of the product.
- The entity of "GoodsCharacteristics" shows the value of a specific characteristic of a particular product.

Thus, the architectural solution of the hierarchical information system was proposed. It allows solving the task of determining and segmentation of the target audience. In addition, the component, which is responsible for sending messages to clients from the target audience, was developed.

5. Experiments

To check the performance of the developed HIS, three tasks have been solved to promote a new sports club service: target audience identifying task, customer segmentation task and management task of targeted advertising. The target audience has been determined based on the database of the club's clients and among users of the social networks Instagram and Facebook.

The marketing research of the fitness services market, the communication with clients and analysis of their requests made it possible to see the increased interest of the club's clients in losing weight, quick recovery after heavy physical exertion, developing muscle endurance and strength, increasing their elasticity, strengthening the cardiovascular system, etc. In this regard, management of the fitness club has been decided to create a new type of service in this fitness center as aqua aerobics. This type of fitness contributes to the normalization of weight, hardening of the body and strengthening its immunity, smoothing out the manifestations of cellulite, increasing skin tone, relieving muscle and emotional stress, neutralizing the negative effects of stress, strengthening the nervous system, normalizing sleep. The aforementioned marketing research would determine the set of indicators for identifying the target audience:
- x_1 – income of clients: x_{11} – low, x_{12} – average, x_{13} – high;
- x_2 – field of activity: x_{21} – office worker, x_{22} – student, x_{23} – business manager, x_{24} – housewife, x_{25} – creative activity;
- x_3 – social media interests in Instagram and/or Facebook: x_{31} – presence of likes, followers and subscriptions of pages with information about sport clubs, diet, weight loss, child goods, mam's publics, x_{32} – absence of such records, x_{33} – private account or absence of account in social networks;
- x_4 – geographic location of a potential client: x_{41} – residents of the area where the fitness club is located, x_{42} – residents of other areas;
- x_5 – age: x_{51} – less, than 21 years, x_{52} – 21-35 years, x_{53} – 35-45 years, x_{54} – more, than 45 years.

The management of the fitness club has been created the profile of the target audience: x_{12} or x_{13}; x_{21} or x_{23} or x_{24}; x_{31} or x_{33}; x_{41} or x_{42}; x_{52} or x_{53} or x_{54}. Let's consider several potential clients of aqua aerobics service. Information from club's databases, social networks or questionnaires is presented as a set of indicators in Table 2. Analysis of input data allow seeing that some indicators for the portrait of a potential consumer have several values. It, in turn, adds uncertainty during the deciding whether to include a particular potential client in the target group. The usage of the proposed

technology could clearly define the future client based on the varying process of the threshold value, which helps to increase or filter the customer base.

Table 2
Result of the resolving the target audience identifying task

Potential client	Values of the input indicators					Gower coefficient S_k	Threshold value	Belonging to the target audience
	x_1	x_2	x_3	x_4	x_5			
1	x_{12}	x_{23}	x_{33}	x_{41}	x_{53}	1		+
2	x_{11}	x_{21}	x_{31}	x_{42}	x_{52}	1		+
3	x_{13}	x_{24}	x_{31}	x_{42}	x_{53}	1	0,55	+
4	x_{11}	x_{25}	x_{32}	x_{41}	x_{54}	0,4		-
5	x_{12}	x_{22}	x_{31}	x_{41}	x_{51}	0,6		+

The customers 1-3 and 5 fell into the target group with the current threshold as can be seen from Table 2. If the threshold value increases to 0.61 to refine the set of potential customers, then we will consider only customers 1-3. Despite the fact that two of them live in a different area, the scope of their employment and interests allows to offer them the considered service.

Second stage of proposed technology is resolving the customer segmentation task for the potential clients of aqua aerobics service. Analysis of customer preferences and the impact of aqua aerobics on the human body shows that potential customers could be divided into 3 segments. The first segment is responsible for improving the muscles and visible changes of the body of clients, that is, this service can be used as an additional training. The aim of the second group of clients is psychological and physical rehabilitation, that is, to relieve muscle and emotional tension and problems. The purpose of the third segment of clients is the organization of leisure, that is, more entertaining than strengthening the body. The result in this case is the acquisition of new acquaintances, an interesting pastime and an increase in self-esteem. To solve the segmentation task, the following indicators have been proposed:

- x_1 – health status: x_{11} – poor health, x_{12} – average state, x_{13} – good health;
- x_2 – social identification: x_{21} – business people, x_{22} – student, x_{23} – athlete, x_{24} – former athlete, x_{25} – housewife;
- x_3 – social media interests in Instagram and/or Facebook: x_{31} – sport clubs, x_{32} – diet, x_{33} – weight loss, x_{34} – child goods, x_{35} – pools, x_{36} – private account or absence of account in social networks;
- x_4 – free time: x_{41} – all day, x_{42} – weekend, x_{43} – weekday evenings;
- x_5 – age: x_{51} – less, than 21 years, x_{52} – 21-35 years, x_{53} – 35-45 years, x_{54} – more, than 45 years;
- x_6 – marital status: x_{61} – single, x_{62} – married, x_{63} – have kids, x_{64} – childless;
- x_7 – regularity of training: x_{71} – 0-1 per week, x_{72} – 2-3 per week, x_{73} – more than 3 times per week;
- x_8 – possible problems of the target audience: x_{81} – a weight problem, x_{82} – problems with communication, x_{83} – problems with appearance, x_{84} – bad mood, x_{85} – narrow social circle, x_{86} – low immunity, x_{87} – absence of complaints.

The snippet of input data for ten potential clients and result of resolving the segmentation task is presented in Table 3. The clients 1-3 were chosen as initial centers for the three clusters. Result of using the K-Means Clustering method at the first stage is the set of values of similarity measure between clients 1-3 and 4-10. It is a base for defining the belonging of each client to one of the clusters. Next step is to calculate new cluster centers. After the sixth iteration, the cluster centers have stabilized.

An analysis of the content of each segment shows that the first cluster includes middle-aged housewives with overweight problems and a former athlete. Based on the values of other indicators, it can be understood that this segment of clients is aimed at rehabilitation, support and restoration of health. The second cluster of potential customers includes those people who are seriously into sports. For this group, it is necessary to create special training programs. The third segment is a group of

young people, mostly unmarried, who are focused not only on sports, but also on a pleasant and interesting pastime.

Table 3
The result of the segmenting process of the targeted audience

Clien t	Values of the input indicators								Number of iteration and content of each cluster					
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	1	2	3	4	5	6
1	x_{13}	x_{21}	x_{31}	x_{43}	x_{52}	x_{61}	x_{72}	x_{87}	I 1,	I 4,	I 3,	I 5,	I 2,	I 2,
2	x_{12}	x_{25}	x_{33}	x_{41}	x_{53}	x_{62}	x_{73}	x_{81}	4, 8,	5, 8	5, 10	8, 10	5, 10	5, 10
3	x_{11}	x_{22}	x_{31}	x_{43}	x_{51}	x_{61}	x_{71}	x_{82}	9					
4	x_{12}	x_{21}	x_{31}	x_{43}	x_{53}	x_{63}	x_{71}	x_{86}		II 2,	II 2,	II 2,	II 4,	II 4,
5	x_{13}	x_{25}	x_{32}	x_{41}	x_{52}	x_{64}	x_{73}	x_{83}	II 2,	6, 7,	4, 6,	4, 6,	6, 7	6, 7
6	x_{12}	x_{23}	x_{31}	x_{42}	x_{53}	x_{63}	x_{73}	x_{87}	5, 6,	10	7, 8	7		
7	x_{11}	x_{24}	x_{34}	x_{42}	x_{53}	x_{63}	x_{72}	x_{81}	7, 10				III 1,	III 1,
8	x_{11}	x_{21}	x_{35}	x_{43}	x_{52}	x_{62}	x_{71}	x_{84}		III 1,	III 1,	III 1,	3, 8,	3, 8,
9	x_{13}	x_{23}	x_{35}	x_{43}	x_{52}	x_{64}	x_{73}	x_{82}	III 3	3, 9	9	3, 9	9	9
10	x_{12}	x_{24}	x_{36}	x_{42}	x_{54}	x_{62}	x_{73}	x_{85}						

Thus, the practical obtained result shows the feasibility of using the proposed technology in the article in real conditions to create a target audience.

6. Discussion

Efficiency estimation of the HIS was carried out separately for each task: target audience identifying task, customer segmentation task and management task of targeted advertising. The classification method based on the calculation of the similarity measure has been used for target audience identifying task. Therefore the standard metrics for assessing the effectiveness of the Precision and Recall classification were chosen to assess the effectiveness of HIS. The Confusion matrix with the corresponding indicators for calculating these metrics is presented in Table 4.

Table 4
Confusion matrix for Precision and Recall metrics evaluation

		Predicted	
		Positive	Negative
Actual	Positive	TP = 1123	FN = 77
	Negative	FP = 41	TN = 429

Data for the matrix is the result of comparison of the results of the work of HIS and the opinion of an expert. The classifier assigned 1164 clients from 1670 to the target audience. It incorrectly attributed 41 clients to the target audience and did not attribute 77 clients from the target audience to it. The formulas for calculating and Precision and Recall metrics values are presented below:

$$Precision = \frac{TP}{TP + FP} = \frac{1123}{1123 + 41} = 0.96,$$

$$Recall = \frac{TP}{TP + FN} = \frac{1123}{1123 + 77} = 0.94$$

The values of the Precision and Recall metrics are quite high and very close to one, which indicates a high classification efficiency.

K-means clustering algorithm was used for resolving the customer segmentation task. The Rand index (RI) has been chosen for evaluating the quality of clustering algorithm. The RI calculates a proximity measure between two clusters based on the comparing results of pairs that are assigned in the same or different clusters in the predicted and true clustering process. The input data for clustering were 1200 clients of the target audience. The contingency table with the indicators values for calculating the RI is presented in Table 5.

Table 5
Contingency table for Rand index evaluation

	Same cluster	Different clusters
Same class	SS = 337	DS = 61
Different classes	SD = 53	DD = 749

An expert compared the clustering results of the HIS with the own results of the distribution of clients into groups. The formula for calculating and the value of the Rand index are presented below:

$$Rand = \frac{SS + DD}{SS + DS + SD + SS} = \frac{337 + 749}{337 + 61 + 53 + 749} = 0.91$$

The obtained value of the RI is rather close to one. It indicates an almost complete coincidence of clusters and classes, which showed a high efficiency of clustering.

Conversion rate (CR) was used to evaluate the effectiveness of targeted advertising mailing. The target audience of 1200 clients was split into two clusters. The first cluster included 418 clients, the second one 782. Different advertisements about the new service of the sports club were prepared for the clients of each cluster. The set of data was prepared according to the customers' actions within a month from the date of sending advertisements. The formula for CR calculating and CR values for the two clusters are presented below:

$$CR = \frac{purchase}{total\ number\ of\ ads}\ 100\%,$$

$$CR_{1claster} = \frac{57}{418}\ 100\% = 13.64\%,$$

$$CR_{2claster} = \frac{108}{782}\ 100\% = 13.81\%.$$

In SMM, an advertising company is considered as successful if the CR values have reached 3-5%. In our case, the rather high CR values can be explained by the fact that the target audience included both real customers from the sports club's database and new customers found in social networks.

7. Conclusion

In this study, the approach was proposed for solving the task of dividing potential customers into non-targeted audience and targeted audience with additional segmentation for a more effective advertising campaign. The methods and existing applications for solving the given task were considered. The model of resolving the identifying task of the targeted audience and the model of the segmenting process of the targeted audience were developed. The architectural solution for the HIS has been developed on the base of the chosen architectural pattern "client-server". The conducted experiment and the assessment of the performance of the HIS have showed the feasibility of usage of the developed HIS in real conditions for managers that conduct an advertising campaign in order to attract new customers and improve the financial condition of the enterprise.

8. References

[1] Matan Naveh, How To Identify a Target Audience for Your Business, 2022. URL: https://elementor.com/blog/how-to-identify-target-audience/.

[2] Customer Segmentation Models, 2017. URL: https://medium.com/think-with-startupflux/customer-segmentation-models-52ef7738823a.

[3] What is Customer Segmentation? – Types, Techniques, Models. URL: https://survicate.com/customer-segmentation/what-is-customer-segmentation/.

[4] K. Baker, The Ultimate Guide to Customer Segmentation: How to Organize Your Customers to Grow Better, 2020. URL: https://blog.hubspot.com/service/customer-segmentation.

[5] E. F. Ayetiran, A. B. Adeyemo, A Data Mining-Based Response Model for Target Selection in Direct Marketing, International Journal of Information Technology and Computer Science 4(1) (2012). doi:10.5815/ijitcs.2012.01.02.

[6] E. W. Maibach, A. Leiserowitz, C. Roser-Renouf, C. K. Mertz, Identifying like-minded audiences for global warming public engagement campaigns: an audience segmentation analysis and tool development, PLoS ONE 6(3) (2011). URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0017571 doi:10.1371/journal.pone.0017571

[7] G. Tirenni, C. Kaiser, A. Herrmann, Applying decision trees for value-based customer relations management: Predicting airline customers' future values, J Database Mark Cust Strategy Manag 14 (2007) 130–142. doi:10.1057/palgrave.dbm.3250044.

[8] K. Melnyk, S. Kirkin, Intelligent Data Processing in Creating Targeted Advertising, in: Proceedings of the 1st International Conference Computational Linguistics And Intelligent Systems, volume 1 of COLINS 2017, NTU «KhPI», Kharkiv Ukraine, 2017, pp. 131–132.

[9] M. Karim, R. M. Rahman, Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing, Journal of Software Engineering and Applications 6 (4) (2013). URL: https://www.scirp.org/html/6-9301587_30463.htm. doi:10.4236/jsea.2013.64025.

[10] HubSpot tools. Make My Persona. A Buyer Persona Generator from HubSpot. URL: https://www.hubspot.com/make-my-persona

[11] K. Melnyk, N. Borysova, Integrated Technology for Personnel Assessment Based on the Competencies Model, in: T. Hovorushchenko, A. Pakštas, V. Vychuzhanin, H. Yin and N. Rudnichenko (Eds.), Proceedings of the 9th International Conference "Information Control Systems & Technologies", ICST-2020, Odessa, 2020, pp. 343-357. URL: http://ceur-ws.org/Vol-2711/paper27.pdf. doi:10.13140/RG.2.2.26024.60169.

[12] Customer Segmentation via Cluster Analysis, 2020. URL: https://www.optimove.com/resources/learning-center/customer-segmentation-via-cluster-analysis.

[13] Segmentor. Customer segmentation tool, 2022. URL: https://segmentor.optimove.com/#/.

[14] CleverTap. Audience segmentation. Build actionable user segments with ease, 2022. URL: https://clevertap.com/segmentation/.

[15] HubSpot's Product and Services Catalog, 2022. URL: https://legal.hubspot.com/hubspot-product-and-services-catalog?_ga=2.129780293.818963249.1623571090-2102544527.1617356624.

[16] Experian. Marketing solutions, 2022. URL: https://www.experian.com/business/solutions/marketing-solutions.

[17] SproutSocial. Listening Tools. Inform your business strategy with social listening, 2022. URL: https://sproutsocial.com/features/social-media-listening/.

[18] Qualtrics. Platforma dlya segmentacii rynka. Izuchajte svoyu celevuyu auditoriyu s pomoshch'yu resheniya dlya segmentacii rynka [Qualtrics. Market segmentation platform. Research your target audience with a market segmentation solution], 2022. URL: https://www.qualtrics.com/ru/product-experience/po-dlya-segmentirovaniya-rynka/?rid=langMatch&prevsite=en&newsite=ru&geo=UA&geomatch=.

[19] MailChimp. Put your audience at the heart of your marketing, 2022. URL: https://mailchimp.com/audience/.

[20] Yieldify. Behavioral segmentation, 2022. URL: https://www.yieldify.com/platform/behavioral-segmentation/.

[21] Amplitude Analytics, 2022. URL: https://help.amplitude.com/hc/en-us/categories/360006505092-Amplitude-Analytics.

[22] Indicative. Segmentation, 2022. URL: https://www.indicative.com/feature/segmentation/.

[23] Mixpanel. Limitless segmentation. Analyze why metrics change, 2022. URL: https://mixpanel.com/segmentation/.

[24] D.O. Maidebura, K.V. Melnyk, N.V. Borysova, Analiz isnuiuchykh servisiv nalashtuvannia tarhetovanoi reklamy [Analysis of existing services of setting up targeted advertising], in: E. I. Sokol (Eds.), Proceedings of XXIX International scientific-practical conference in Information technologies: science, engineering, technology, education, health, part 1 of MicroCAD-2021, NTU "KhPI", Kharkiv Ukraine, 2021. p. 32.

[25] B. S. Everitt, S. Landau, D. Stahl, Cluster Analysis, Wiley, New York, NY, 2011.

[26] G. Seif. The 5 Clustering Algorithms Data Scientists Need to Know, 2018. URL: https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68.

[27] M. McGregor. 8 Clustering Algorithms in Machine Learning that All Data Scientists Should Know, 2020. URL: https://www.freecodecamp.org/news/8-clustering-algorithms-in-machine-learning-that-all-data-scientists-should-know/.