STATISTICS IN TRANSITION new series, December 2023
Vol. 24, No. 5, pp. 147–166, https://doi.org/10.59170/stattrans-2023-068
Received – 04.09.2022; accepted – 22.09.2023

Marketing segmentation of banks' corporate clients based on data mining technique

Rostyslav Klochko[1], **Olena Piskunova**[2]

Abstract

In recent years, the banking services market has been developing dynamically, experiencing a sharp increase in competition. Banks that provide maximum profitability for each client gain the most significant competitive advantage. The sales model in corporate banking is associated with personal interaction between bank employees and business owners, and the subsequent establishment of individual service conditions. However, this approach is often ineffective when a bank faces the issue of maximising the efficiency of business activities. This study aims to segment a bank's corporate client base and develop a pricing strategy for each of the groups that have been singled out in the process. The study sample consisted of 4,500 corporate clients of a Ukrainian bank who were active users of euro accounts. The k-means data mining algorithm was used to develop marketing segments. The optimal number of clusters was determined by weighing the results of calculating 26 indices from the NbClust package and the bank's business requirements. Six similarity groups were found during the calculation of the algorithm. The study found that clusters 1 and 2 were a concentration of unprofitable customers for whom an introduction of a service fee was urgently needed. Marketing segments 3 and 4 were customers who did not record net losses but with whom it was deemed necessary to work to improve their profitability. The remaining segments were 'healthy' users of euro accounts. With regard to these customers, it was recommended no additional service fees should be imposed. The proposed methodology makes it possible for a bank to remain attractive in a competitive environment while not incurring unnecessary costs.

Key words: clusterisation, k-means, pricing strategy, customer value, cost optimisation

1. Introduction

In recent years, the banking services market has been developing exceptionally dynamically, experiencing a sharp increase in competition due to the deregulation of

[1] Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after Vadym Hetman, Ukraine. E-mail: rostislav.klochko@gmail.com. ORCID: https://orcid.org/0000-0003-2690-2785.
[2] Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after Vadym Hetman, Ukraine. E-mail: EPiskunova@kneu.edu.ua. ORCID: https://orcid.org/0000-0002-0373-5358.

banking activities, the introduction of new technologies, and increased consumer demand. Banks are increasingly turning to the practice of cost optimization in order to achieve maximum business efficiency. Banks that provide maximum profitability for each client gain the most significant competitive advantage. Accordingly, the concept of customer value and its maximization acquires vital importance.

For most universal banks that provide a wide range of services for the retail and corporate segments, the share of profit received from the corporate is system-forming. Corporate banking refers to the part of banking that deals with corporate clients. It typically serves a wide range of clients, from small and medium-sized businesses with a few million in revenue to large corporations with billions in sales. Banks need more timely and adequate information to serve their corporate clients better. In contrast to the retail business model, the traditional sales model in corporate banking is associated with personal interaction between bank employees and business owners, with the subsequent establishment of individual service conditions. This approach is often ineffective when the bank faces the issue of maximizing the efficiency of business activities, especially when they have thousands of corporate clients in their portfolios. It is not easy to provide an individual approach for every client, not lose the effectiveness of clients' value management, and not spend many human resources simultaneously. On the other hand, the competitive environment does not allow to provide the same pricing policy for all corporate clients.

One of the main options for solving this problem is classifying corporate clients into appropriate similarity groups - marketing segments and developing a price strategy separately for each. The term "marketing segmentation of corporate clients" was first stated in this study. However, the proposed approaches to customer differentiation and further development of individual marketing proposals for each of the groups allow us to consider the usage of this concept appropriate.

2. Literature Review

The application of machine learning methods and data mining for customer segmentation has become one of the main topics in the literature that examines the banking market (He et al., 2022; Aghaei, 2021; Chawla & Joshi, 2021). This is primarily because banks, by their very nature, accumulate vast volumes of customer data, which is crucial in building an effective customer clustering algorithm.

Studies concerning the bank's corporate clients mainly aim to improve the operational efficiency of the business. Usually, it is about determining the riskiest customers unable to repay debts (Oleynik & Formánek 2020) or making banking operations that violate domestic legislation (Hamal & Senvar 2021). In these studies, the authors have a list of confirmed risky customers, based on the example of which they try to classify all other unassessed clients. The basis of such studies is machine learning classification methods, such as support vector machine, naive Bayes, artificial neural network, k-nearest neighbour, random forest, and logistic regression.

However, if we analyse the research on all retail and corporate banking customer segments, customer segmentation is usually focused on achieving marketing goals (Rajaobelina et al., 2019; Firdaus & Utama, 2021; Fathima & Muthumani, 2019), where retail research has both a quantitative and a qualitative advantage. Two characteristics of the retail segment can explain this: this perimeter includes the largest number of customers and generates enormous volumes of data; the peculiarity of retail customers allows applying the same marketing policy to many customers, which is not usual if we are talking about corporate clients.

The range of problems covered in the literature on this topic is wide. Yanık & Elmorsy (2019) developed a clustering model based on the data of 40,000 credit card users, which made it possible to implement eight different marketing strategies according to the specifics of the use of banking products and socio-demographic characteristics of customers. The study is based on a self-organizing map and k-means clustering models. A similar study was conducted by Hung et al. (2019). The only difference is that they used the hierarchical agglomerative clustering algorithm to perform the analysis. In general, neural-network-based clustering approaches are gaining popularity among banking researchers. These algorithms are usually used in combination with hierarchical clustering (Kovacs et al., 2021) or minimum spanning tree (Barman & Chowdhury, 2019) to improve segmentation efficiency.

It is necessary to highlight studies on customer segmentation in the lending process. Vijayalakshmi et al. (2020) created a segmentation model based on random forest, logistic regression, and support vector machine learning algorithms to estimate the default of a future loan more accurately. At the same time, Umuhoza et al. (2020) used the k-means algorithm to better manage the portfolio of already issued loans.

Calvo-Porral & Lévy-Mangin (2020) proposed the algorithm for developing different communication policies based on the client's psycho type. They divide clients into groups using confirmatory factor analysis (CFA) and Anova test algorithms.

The presence of a large number of customer activity characteristics allowed Djurisic et al. (2020) to develop a segmentation model based on the combination of the customer significance assessment approach (Recency, Frequency, and Monetary) with clustering (k-means) and classification (Support Vector Machine) models. This algorithm helped to focus the bank's efforts on retaining only those customers who bring real value to the company. A similar study was proposed by Dang Tran et al. (2023) to address the problem of predicting bank customer churn.

In turn, due to mainly personal service and a small amount of structured data, there is no variety of research on the specifics of working with corporate clients, especially with the use of intelligent modelling technologies.

Formisano et al. (2020) made the first attempts at marketing research of the bank's corporate clients. With the help of the Kano model, the researchers tried to estimate the non-linear relationship between the level of customer satisfaction and the quality of banking services. The analysis results can be applied to identify groups of dissatisfied customers, the so-called irritants, and improve their customer experience.

Tungjitnob et al. (2021) focused on identifying corporate users (business owners) among mobile banking customers. The developed algorithms will allow companies to implement special offers specifically for the corporate segment of online banking users. The primary means of modelling were the Extreme Gradient Boosting algorithm and Convolutional Neural Network machine learning methods.

Osowski & Sierenski (2020) highlighted a vital research topic on corporate clients' activity. Based on the Neural Networks technology, they developed an algorithm that predicts the activity status of a corporate client. The proposed model will allow banks to identify groups of customers at risk of leaving the bank and influence them with marketing methods.

Studies of the bank's corporate clients from the marketing point of view have hardly found their development in the literature. Furthermore, the ones that describe the peculiarities of working with corporate customers' data mainly focus on efforts to predict future processes by applying complex machine learning or neural network models. Approaches to analysing the current behaviour characteristics of corporate customers are almost not revealed in the available research.

The establishment of approaches to individual pricing in the corporate perimeter reduced the number of publications on developing marketing proposals to zero. Existing studies often rely on data unique to a specific case study, which is tricky to apply to other banks. The described problems determined the relevance of this study - the development of marketing segmentation of bank's corporate clients with the subsequent implementation of the pricing policy for each similarity group. The value of our research is the transfer of retail customer segmentation approaches to the corporate perimeter. Most of these algorithms have been known for a long time, do not require unique data, and are easily calculated and scaled to other companies. Due to its simplicity in implementation, most studies used the k-means technique, which led to the choice of this clustering algorithm in our study. At the same time, the small number of customers for analysis and the limited available characteristics of their activity significantly complicate the segmentation process using the k-means algorithm. Nevertheless, the results of our research prove that the available volumes of data are pretty enough to talk about its effectiveness for solving the tasks.

3. Aim of The Study

This study aims to develop a marketing segmentation model for corporate clients of a universal bank that will help optimize profitability and efficiency in today's competitive banking landscape.

The following tasks were defined to achieve the goal:

1. Describe and systematize the data clustering methodology and algorithm hyperparameters' setting.
2. Following the developed methodology, implement the clustering algorithm of the bank's corporate clients.
3. Analyse the results and propose pricing strategies for each identified similarity group.

This segmentation approach, connected with recommendations, is expected to optimize operating costs without introducing a uniform fee structure for all bank corporate clients. As a result, banks can remain attractive in a competitive environment while avoiding unnecessary expenses. Furthermore, the practical research findings can serve as a basis for the development of automated marketing campaigns and personalized communication strategies tailored to each marketing segment of corporate clients, enhancing overall customer satisfaction.

4. Materials and Methods

4.1. Methodology of banks' corporate client segmentation.

Data mining is the process of analysing large data sets to identify patterns and relationships that can help solve business problems. Data mining technologies make it possible to synthesize valuable information that is implicitly contained in the data. The large volumes of accumulated data in companies make data mining approaches helpful in finding insights that managers can use in decision-making. Data mining methods and tools allow enterprises to predict future trends and make more informed business decisions.

Clustering is one of the most common data mining approaches to gaining an intuitive understanding of data structure. Cluster analysis is a method of analysing data to find and identify similar data. This process helps to understand the differences and similarities between the data.

In our study, we will use only the k-means clustering method, which, due to its simplicity, is considered one of the most used clustering algorithms. Moreover, we used RStudio software and the R programming language for all calculations.

k-means clustering aims to divide objects into k clusters so that data points in one cluster are similar and data points in different clusters are as distant as possible. The distance between them determines the similarity of the two points.

The stages of implementation of the algorithm are:

1) determination of the required number of clusters - k;
2) determination of the initial value of the centroid;
3) calculation of the distance from existing objects to the centroid;
4) assignment of each object to the corresponding cluster depending on the calculated distances.

The first hyperparameter of the k-means clustering model is the number of similar groups that must be obtained as a result of the algorithm implementation. There are many criteria by which the statistically necessary number of groups can be pre-estimated. Our study used the NbClust R library (Dimitriadou 2002), which allows us to estimate the required number of clusters according to 26 criteria (Table 1). The number for which most criteria "vote" should be considered optimal. However, the final number of clusters is always chosen considering the organization's business needs.

The second hyperparameter of the algorithm is the number of times the algorithm is calculated, where a different initial centroid is randomly assigned each time. The number depends on the computing capabilities of the system. In our study, we used 25 iterations of the algorithm.

Table 1: Characteristics of indexes in the NbClust R library

Full name	Short name	Selection criterion
Hubert index. Hubert and Arabie 1985	Hubert	Graphical method
Dindex. Lebart et al. (2000)	Dindex	Graphical method
KL index. Krzanowski and Lai (1988)	KL	The maximum value of the index
CH index. Calinski and Harabasz (1974)	CH	The maximum value of the index
Hartigan index. Hartigan (1975)	Hartigan	The maximum difference between hierarchy levels of the index
Cubic Clustering Criterion (CCC). Sarle (1983)	CCC	The maximum value of the index
Scott index. Scott and Symons (1971)	Scott	The maximum difference between hierarchy levels of the index
Marriot index. Marriot (1971)	Marriot	Max. value of second differences between levels of the index
TraceCovW index. Milligan and Cooper (1985)	TrCovW	The maximum difference between hierarchy levels of the index
TraceW index. Milligan and Cooper (1985)	TraceW	The maximum value of absolute second differences between levels of the index
Friedman index. Friedman and Rubin (1967)	Friedman	The maximum difference between hierarchy levels of the index

Table 1: Characteristics of indexes in the NbClust R library (cont.)

Full name	Short name	Selection criterion
Silhouette index. Kaufman and Rousseeuw (1990)	Silhouette	The maximum value of the index
Ratkowsky index. Ratkowsky and Lance (1978)	Ratkowsky	The maximum value of the index
Ball index. Ball and Hall (1965)	Ball	The maximum difference between hierarchy levels of the index
PtBiserial index. Examined by Milligan (1980,1981)	Ptbiserial	The maximum value of the index
Dunn index. Dunn (1974)	Dunn	The maximum value of the index
Rubin index. Friedman and Rubin (1967)	Rubin	The minimum value of second differences between levels of the index
C-index. Hubert and Levin (1976)	Cindex	The minimum value of the index
DB index. Davies and Bouldin (1979	DB	The minimum value of the index
Duda index. Duda and Hart (1973)	Duda	The smallest number of clusters such that index > criticalValue
Pseudot2 index. Duda and Hart (1973)	Pseudot2	The smallest number of clusters such that index < criticalValue
Beale index. Beale (1969)	Beale	The number of clusters such that the critical value of the index >= alpha
Frey index. Frey and Van Groenewoud (1972)	Frey	The cluster level before that index value < 1.00
Mcclain index. McClain and Rao (1975)	McClain	The minimum value of the index
SDindex. Halkidi et al. (2000)	SDindex	The minimum value of the index
SDbw. Halkidi et al. (2001)	SDbw	The minimum value of the index

The criterion for choosing the optimal initial centroid is the parameter Within Cluster Sum of Squares (W_total) (1). The algorithm assigns data points to a cluster so that the sum of the squares of the distances between the data points and the cluster's centroid is minimal. The less variation we have within the clusters, the more similar the data points in the same cluster.

$$W_{total} = \sum_{l=1}^{k} \frac{\rho(C^l)}{n^l}, \quad (1)$$

where $\rho(C^l)$ - the sum of Euclidean distances between points within the cluster l ; n^l – the number of points in cluster l; k– the number of clusters.

The final step of our research is the description of the constructed clusters. Each segment differs in its average indicators of activity, which must be calculated. Calculating the average values of the model's parameters makes it possible to provide recommendations regarding the need to carry out specific work with a particular group of clients to reduce bank operational costs.

4.2. Initial data set features

The basis of the study is data on the activity of euro current account users for the period from April 2020 to April 2021. The sample includes 4,500 corporate clients of the Ukrainian bank (Private entrepreneurs and Legal entities). The complete list of initial data fields is in Table 2.

In Ukraine, the economic crises of 2008-2010 and 2014-2015 resulted in stopping bank lending in foreign currency. As a result, euro deposits have become a loss-making type of service for banks (service cost is 1% per year). Accordingly, if a client has only one product - a euro account, and does not perform any banking operations for which the bank would receive commission income - such a client is entirely unprofitable. Given this, there is a need to minimize operating costs by the following:

- decrease in clients' euro account balances;
- increase in the number of operations with euro accounts;
- introduction of a new service fee for euro accounts.

Table 2: Characteristics of the initial data sample fields

Model Variables	Description
CL_ID	Client internal ID
NOM_nb	The number of months when the client's euro account had positive account balances
Oper_nb	The number of months when the client performed transactions with euros
NIM_nb	The number of months when the net interest income for this client was negative (i.e. net marginal loss)
NBI_nb	The number of months when the net banking income for this client was negative (i.e. net banking loss). Net banking income is the sum of net interest income and commission income.

5. Results

As mentioned above, the first step in implementing the bank's corporate clients clustering is determining the optimal number of similarity groups. Table 3 presents the results of calculating 26 evaluation criteria in the NbClust package (the optimal values for each criterion are highlighted). Table 4 shows the "voting" results for the optimal number of clusters.

Table 3: Values of calculated indices in the NbClust package

Index	Number of clusters								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
KL	1.293	0.318	1.624	**28.729**	0.030	20.489	3.749	0.189	1.322
CH	494	620	979	1238	1133	**1682**	1601	1510	1482
Hartigan	586	1006	765	190	1066	**169**	121	163	138
CCC	-19	-14	12	32	31	**71**	70	69	70
Scott	3272	4302	6233	**9692**	10997	13060	13561	14099	14956
Marriot	3.52E+12	4.47E+12	2.72E+12	**6.22E+11**	4.33E+11	1.88E+11	1.85E+11	1.74E+11	1.33E+11
TrCovW	1892585	1222424	**525480**	282585	256320	69466	62513	57201	49881
TraceW	5644	4257	**2730**	1914	1731	1085	992	929	852
Friedman	5	5	8	**17**	24	28	29	32	38
Rubin	1.275	1.690	2.636	3.759	4.158	**6.629**	7.256	7.745	8.449
Cindex	0.179	0.134	0.146	0.164	0.235	0.180	0.167	0.167	**0.134**
DB	1.721	1.471	1.050	0.901	0.911	**0.790**	0.904	0.990	1.055
Silhouette	0.328	0.382	0.375	0.435	0.436	**0.513**	0.511	0.511	0.446
Duda	**0.814**	4.857	2.629	0.428	5.324	1.175	1.486	2.098	1.780
Pseudot2	**293**	-1058	-442	76	-604	-90	-128	-274	-27
Beale	**0.551**	-1.915	-1.494	3.197	-1.905	-0.360	-0.788	-1.225	-1.053
Ratkowsky	0.213	0.347	**0.392**	0.381	0.353	0.348	0.328	0.311	0.297
Ball	2822	**1419**	682	383	288	155	124	103	85
Ptbiserial	0.266	0.413	0.457	0.509	0.513	**0.537**	0.528	0.527	0.476
Frey	-0.024	-0.198	-0.073	0.034	0.175	0.567	0.367	1.355	2.317
McClain	**0.570**	0.948	0.922	0.891	0.889	0.893	0.931	0.935	1.163
Dunn	0.019	0.018	0.022	0.030	0.044	0.044	0.044	**0.045**	0.040
Hubert	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SDindex	3.393	3.277	4.310	3.234	2.807	**2.501**	2.539	2.713	3.748
Dindex	1.398	1.152	1.005	0.871	0.847	0.655	0.634	0.619	0.560
SDbw	1.140	1.028	1.247	0.897	0.579	0.466	0.437	0.415	**0.400**

Table 4: "Voting" results for the optimal number of clusters

Number of clusters	**2**	**3**	**4**	**5**	**7**	**9**	**10**
The number of indices that voted for the optimum	4	1	3	4	8	1	2

Most of the indices voted for 7 clusters. Next is calculating the average model parameter values for the optimal number of clusters (Table 5).

Table 5: Average characteristics of clusters when clients are divided into seven marketing segments

Cluster	Number of customers	NOM_nb	Oper_nb	NIM_nb	NBI_nb
1	663	12	11	0	0
2	788	2	11	0	0
3	1073	11	1	0	0
4	**92**	**10**	**3**	**8**	**5**
5	**53**	**12**	**0**	**11**	**11**
6	1578	1	2	0	0
7	253	11	4	8	0

The determination of the initial centroid is performed randomly among all available objects. The number of repetitions of the algorithm calculation is pre-set to optimize the quality of the clustering algorithm, considering that a different centroid will be assigned each time. In our study, we used a 25-fold repetition of the calculation, which is sufficient to avoid the problem of an incorrectly chosen centroid.

The optimality of the initial centroid selection and the implementation of the entire algorithm depends on one more hyperparameter - the choice of the method of calculating the distance from existing objects to the centroid. Our study used the best-known calculation algorithm - the Euclidean distance between points. Accordingly, the optimal algorithm, and the optimal initial centroid, is the algorithm iteration in which the index of the total distance from the existing objects to the group's centroid is the smallest (equation 1).

Clusters 4 and 5 characterize the same group of clients. The only difference is the intensity of euro account usage. Because of this, we decided to implement an algorithm with only six groups. The characteristics of the distribution into six segments are presented in Table 6.

Table 6: Average characteristics of clusters when clients are divided into six marketing segments

Cluster	Number of customers	NOM_nb	Oper_nb	NIM_nb	NBI_nb
1	284	11	4	8	1
2	110	11	2	10	8
3	1 071	11	1	0	0
4	665	12	11	0	0
5	1 581	1	2	0	0
6	789	2	11	0	0

To facilitate the characterization of each defined marketing segment and to provide personalized recommendations for reducing the bank's operating costs, data on clients' activity in each segment were displayed as a boxplot (Fig. 1-6).

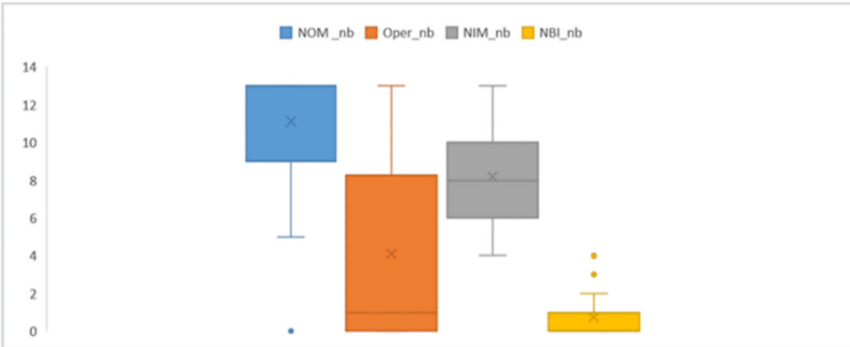

Figure 1: Activity in the first segment of corporate clients

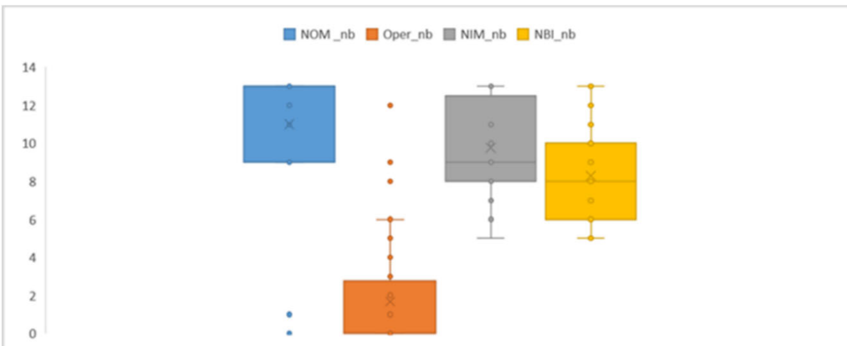

Figure 2: Activity in the second segment of corporate clients

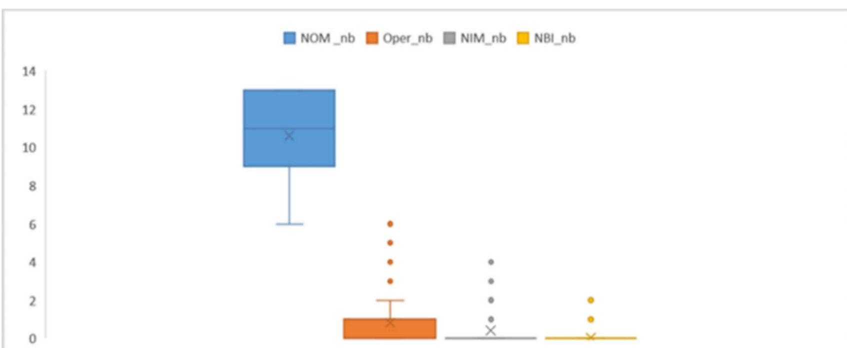

Figure 3: Activity in the third segment of corporate clients

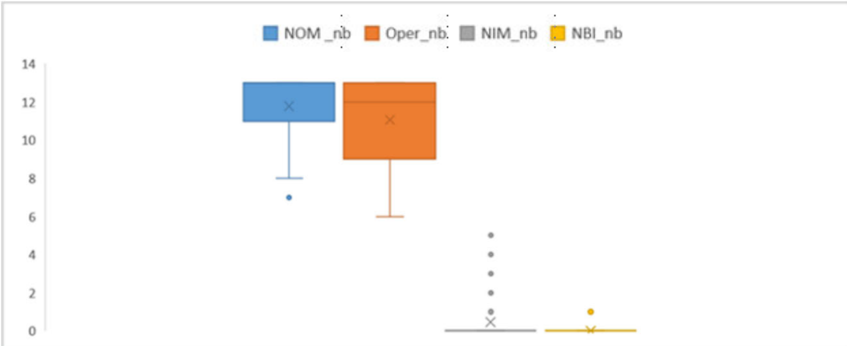

Figure 4: Activity in the fourth segment of corporate clients

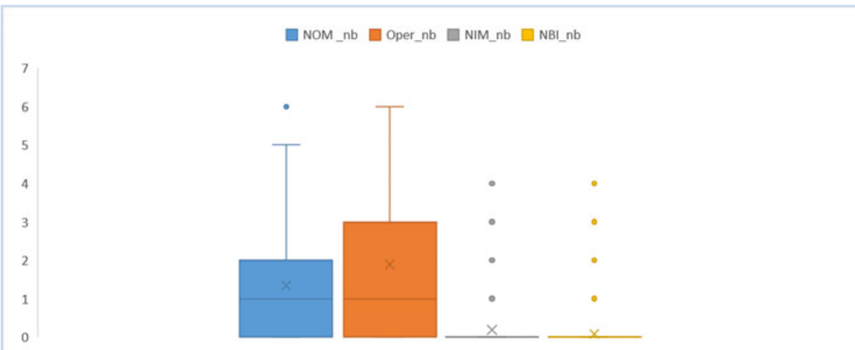

Figure 5: Activity in the fifth segment of corporate clients

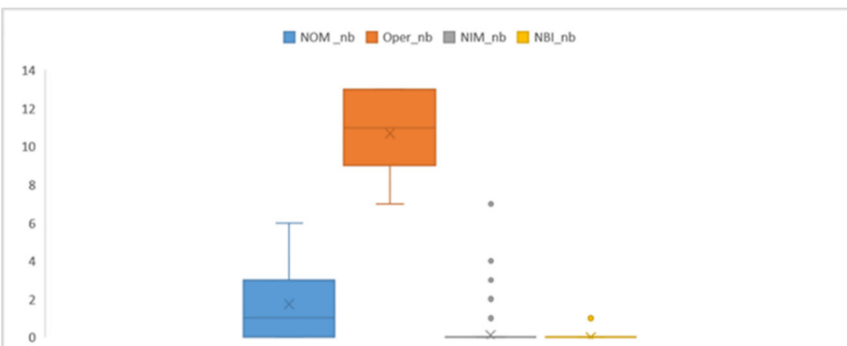

Figure 6: Activity in the sixth segment of corporate clients

6. Discussion

This study reveals the entire methodology of clustering the customer base, considering their behaviour's peculiarities. The object of the study was corporate clients who are active euro account users in one of the Ukrainian banks. Since, in Ukraine, there is no possibility to earn on foreign currency (due to insufficient lending), maintenance of euro accounts is a net operating loss for the bank.

After the customer activity analysis, each marketing segment can be characterized as follows (Table 6):

1) unprofitable transactors with euro accumulation;
2) unprofitable euro accumulation;
3) profitable euro accumulation;
4) profitable transactors with euro accumulation;
5) without euro accumulation;
6) profitable transactors without euro accumulation.

The last task is to develop a pricing strategy for the euro account service. Six graphs showing each parameter's average in a boxplot were created to understand better the customers' activity in each cluster (Figures 1-6).

Figure 1 displays the clients' activity in cluster 1:

- constant positive balance on euro accounts;
- euro transactions are performed only 4-6 times a year;
- these clients show a net marginal loss most of the time due to insufficient outstanding in the national currency or loans;
- euro transactions bring additional commission income, mainly covering net marginal loss.

Given the above characteristic and taking into account the small number of customers - 284 (Table 6), the following recommendations were developed to improve the profitability of the marketing segment:

- to communicate with clients and recommend purchasing euros at the time of need only;
- introduce a 1% service fee on euro accounts.

Figure 2 displays the clients' activity in cluster 2:

- constant positive balance on euro accounts;
- euro transactions are performed only a few times a year;
- these customers always show a net banking loss.

Given the above characteristic and taking into account the small number of customers - 110 (Table 6), the following recommendations were developed to improve the profitability of the marketing segment:

- introduce a 1% service fee on euro accounts;
- it is recommended to survey to determine the reason for the constant euro accumulation without further use.

Figure 3 displays the clients' activity in cluster 3:

- constant positive balance on euro accounts;
- euro transactions are performed only a few times a year;
- these clients, thanks to other banking products, are currently profitable for the bank.

Because of the above characteristic and taking into account a large number of customers - 1,071 (Table 6), the following recommendations were developed to improve the profitability of the marketing segment:

- it is recommended to survey to determine the reason for the constant euro accumulation without further use;
- to communicate with clients about introducing a 1% service fee on euro accounts in the second stage of the reduction operating costs exercise (to give time to prepare for this).

Figure 4 displays the clients' activity in cluster 4:

- constant positive balance on euro accounts;
- accumulation is constantly used for euros transactions;
- current activity allows clients to remain fully profitable for the bank.

Because of the above characteristic and taking into account a large number of customers - 665 (Table 6), the following recommendations were developed to improve the profitability of the marketing segment:

- to communicate with clients and recommend purchasing euros at the time of need only;
- to communicate with clients about introducing a 1% service fee on euro accounts in the second stage of the reduction operating costs exercise (to give time to prepare for this).

Figure 5 and Figure 6 display the clients' activity in clusters 5 and 6:

- several times a year show a positive balance on euro accounts;
- customers of the sixth segment usually buy currency at the time of the transaction, which allows them not to accumulate euro on current accounts;
- accumulation is constantly used for euros transactions (difference between clusters is in the number of transactions);
- current activity allows clients to remain fully profitable for the bank.

Given the above characteristic and taking into account the most significant number of customers - 1,581 and 789 (Table 6), the following recommendations were developed to improve the profitability of the marketing segment:

- do not introduce any commissions for maintenance of euro accounts.

As we can see, clusters 1 and 2 are a concentration of unprofitable customers for whom an introduction of a service fee is urgently needed. Marketing segments 3 and 4 are customers who do not show net losses but with whom it is necessary to work to improve their profitability. Furthermore, to prevent possible losses, introduce service fees for these customers in the future. The last segments are examples of the healthy use of euro accounts. Customers do not accumulate euro but buy it at the moment of need without bringing losses to the bank. These customers are recommended not to set additional service fees, which saves human resources since these customers make up more than half of all users of euro accounts.

The effectiveness of the developed algorithm can be evidenced by the case of description and characteristics of similarity groups, as each segment has clearly defined differences that allow for the development of individual marketing recommendations for pricing. This segmentation and recommendations will optimize operating costs without setting a fee for all bank corporate clients. In the first wave, only 9% of clients were processed (an individual service fee was introduced), which also affected the bank's human resources optimization. For the first time, segmentation marketing approaches inherent in retail were applied for a personalized approach to corporate customers. This methodology allows the bank to remain attractive in a competitive environment while not incurring unnecessary costs.

Research that concerns the bank's corporate clients mainly examines improving the operational efficiency of doing business. In contrast to works (Oleynik & Formánek 2020) and (Hamal & Senvar 2021), this study is one of the first attempts to consider the bank's corporate clients from the point of view of the marketing component, especially with intelligent modelling technologies. The approach's basis was practiced primarily developed for retail clients, which we projected for the bank's corporate clients.

k-means is the basis of most of these studies, but compared to the algorithms Yanık & Elmorsy (2019) and Djurisic et al. (2020), we have improved the process of selecting hyperparameters of the model. A 25-fold selection of the initial centroid was applied according to the Within Cluster Sum of Squares smallest value (1). Also, based on the calculation of 26 indices, the optimal number of clusters was selected (Table 3).

Also, in our study, compared to research (Hung et al. 2019) and (Calvo-Porral & Lévy-Mangin 2020), more attention is paid to characterizing marketing segments and the development of personalized marketing strategies.

If we discuss the research of corporate clients from the point of view of the marketing component, the advantage of our methodology over (Formisano et al. 2020) is that we used data that are publicly available for each bank, and the modelling algorithm can be easily tested on the client base of other banks.

The current study is focused on the characteristics of the current activity of banks' corporate clients and the development of personalized recommendations that can improve business efficiency in the nearest time. In contrast, most studies that describe the features of working with corporate clients' behaviour data (Osowski & Sierenski (2020)) are mainly focused on efforts to predict future processes by applying complex models of machine learning or neural networks.

Attempts to develop personalized marketing activities are found in (Tungjitnob et al. 2021), but for the first time, we proposed an approach specifically to personalized pricing based on the bank's corporate clients' marketing segments.

Implementation of the developed methodology will be valuable only to those banks ready to partially move from personal services to retail approaches with their segmentation practices. Very often, key stakeholders are not ready to trust intelligent modelling technologies, so they make decisions based on their intuition and understanding of the business. Also, cooperation with corporate clients is often associated with previous long-term agreements, which makes it impossible to change specific price offers in the short term. The results of our research are not relevant for countries whose banks do not incur operational losses from servicing accounts in foreign currency. Also, for the algorithm's efficiency, the bank must have a large number of corporate clients as the algorithm is sensitive to the number of observations it must divide into similarity groups.

The main drawback of the study is the use of a small number of customer activity characteristics. Accordingly, further development may include additional criteria for evaluating customer activity to improve the quality of clustering of the customer base.

7. Conclusions

For most universal banks that provide a wide range of services for the retail and corporate segments, the share of profit received from the corporate is system-forming. At the same time, the traditional sales model in corporate banking is associated with personal interaction between bank employees and business owners. It is often ineffective when the bank faces the issue of maximizing the efficiency of business activities, especially when they have thousands of corporate clients in their portfolios. It is challenging to provide an individual approach for every client, not lose the effectiveness of clients' value management, and not spend many human resources simultaneously. Accordingly, marketing segmentation and subsequent personalized pricing are the main options for solving this problem in the current highly competitive banking environment.

The basis of the study was 4,500 corporate clients of the Ukrainian bank who are active users of euro accounts. The k-means data mining algorithm was used to develop marketing segments. Two hyperparameters were previously set before the implementation: the number of iterations of the algorithm to determine the initial centroid and the number of clusters to be selected. The optimal initial centroid was selected among 25 algorithm iterations based on the minimum value of the Within Cluster Sum of Squares. The optimal number of clusters was determined by weighing the results of calculating 26 indices from the NbClust package and the bank's business requirements. Most indices for evaluating the optimal number of clusters voted for seven similarity groups. Nevertheless, after conducting a preliminary analysis of the average activity of each segment, it was determined that clusters 4 and 5 are representatives of the same group. Accordingly, a decision was made to merge these segments. Six similarity groups were found during the calculation of the algorithm:

1) unprofitable transactors with euro accumulation;
2) unprofitable euro accumulation;
3) profitable euro accumulation;
4) profitable transactors with euro accumulation;
5) without euro accumulation;
6) profitable transactors without euro accumulation.

During the analysis of the average activity indicators of each cluster, it was found that clusters 1 and 2 are a concentration of unprofitable customers for whom an introduction of a service fee is urgently needed. Marketing segments 3 and 4 are customers who do not show net losses but with whom it is necessary to work to improve their profitability. Furthermore, to prevent possible losses, introduce service fees for these customers in the future. The last segments are examples of the healthy use of euro accounts. Customers do not accumulate euro but buy it at the moment of need without bringing losses to the bank. These customers are recommended not to set additional service fees, which saves human resources since these customers make up more than half of all users of euro accounts.

This segmentation and recommendations will optimize operating costs without setting a fee for all bank corporate clients. In the first wave, only 9% of clients were processed (an individual service fee was established), which also affected the optimization of the bank's human resources. Accordingly, for the first time, segmentation marketing approaches inherent in retail were applied for a personalized approach to corporate customers. This methodology allows the bank to remain attractive in a competitive environment while not incurring unnecessary costs.

Practical research results can also be the basis for building automatic marketing campaigns and communications with individual offers for each marketing segment separately. Further research could test an expanded number of customer activity characteristics to improve the quality of the segmentation model. Also, based on the results of the development of marketing segmentation of the bank's corporate clients, it is possible to develop a model for forecasting client needs for a specific product or service.

References

Aghaei, M., (2021). Market segmentation in the banking industry based on customers' expected benefits: A study of Shahr Bank. *Interdisciplinary Journal of Management Studies*, 14(3), pp. 629–648.

Aysha Fathima, Y. and Muthumani, S., (2019). Client cluster identification of internet bank services. *Journal of Computational and Theoretical Nanoscience*, 16(8), pp. 3554–3559.

Barman, D., Chowdhury, N., (2019). A novel approach for the customer segmentation using clustering through self-organizing map. *International Journal of Business Analytics*, 6(2), pp. 23–45.

Calvo-Porral, C. and Lévy-Mangin, J., (2020). An emotion-based segmentation of bank service customers. *International Journal of Bank Marketing*, 38(7), pp. 1441–1463.

Chawla, D., Joshi, H., (2021). Segmenting mobile banking users based on the usage of mobile banking services. *Global Business Review*, 22(3), pp. 689–704.

Dang Tran, H., Le, N. and Nguyen, V.-H., (2023). Customer churn prediction in the banking sector using machine learning-based classification models. *Interdisciplinary Journal of Information, Knowledge, and Management*, 18, pp. 087–105.

Dimitriadou, E., Dolničar, S. and Weingessel, A., (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika*, 67(1), pp. 137–159.

Djurisic, V., Kascelan, L., Rogic, S. and Melovic, B., (2020). Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method. *Applied Artificial Intelligence*, 34(12), pp. 941–955.

Firdaus, U., Utama, D.-N., (2021). Development of bank's customer segmentation model based on rfm+b approach. ICIC Express Letters, *Part B: Applications*, 12(1), pp. 17–26.

Formisano, V., Moretta Tartaglione, A., Fedele, M. and Cavacece, Y., (2020). Banking services for SMEs' internationalization: evaluating customer satisfaction. *The TQM Journal*, 33(3), pp. 662–680.

Hamal, S., Senvar, O., (2021). Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for Turkish SMEs. *International Journal of Computational Intelligence Systems*, 14(1), p. 769.

He, W., Hung, J.-L. and Liu, L., (2022). Impact of big data analytics on banking: A case study. *Journal of Enterprise Information Management*.

Hung, P., Lien, N. and Ngoc, N., (2019). Customer Segmentation Using Hierarchical Agglomerative Clustering. *Proceedings of the 2019 2nd International Conference on Information Science and Systems*.

Kovács, T., Ko, A. and Asemi, A., (2021). Exploration of the investment patterns of potential retail banking customers using two-stage cluster analysis. *Journal of Big Data*, 8(1).

Oleynik, M., Formánek, T., (2020). Predicting Default Probability of Bank's Corporate Clients in the Czech Republic. Comparison of Generalized Additive Models and Support Vector Machine Approaches. *Software Engineering Perspectives in Intelligent Systems*, pp. 709–722.

Osowski, S., Sierenski, L., (2020). Prediction of Customer Status in Corporate Banking Using Neural Networks. *2020 International Joint Conference on Neural Networks* (IJCNN).

Rajaobelina, L., Brun, I. and Ricard, L., (2019). A classification of live chat service users in the Banking Industry. *International Journal of Bank Marketing*, 37(3), pp. 838–857.

Tungjitnob, S., Pasupa, K. and Suntisrivaraporn, B., (2021). Identifying SME customers from click feedback on mobile banking apps: Supervised and semi-supervised approaches. *Heliyon*, 7(8).

Umuhoza, E., Ntirushwamaboko, D., Awuah, J. and Birir, B., (2020). Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in Africa. *SAIEE Africa Research Journal*, 111(3), pp. 95–101.

Vijayalakshmi, M., Gupta, S.-S. and Gupta, A., (2020). Loan approval system through customer segmentation using big data analytics and machine learning. *International Journal of Advanced Science and Technology*, 29(6), pp. 2374–2380.

Yanık, S., Elmorsy, A., (2019). SOM approach for clustering customers using credit card transactions. *International Journal of Intelligent Computing and Cybernetics*, 12(3), pp. 372–388.