DOI: https://doi.org/10.15276/hait.08.2025.32
UDC 004.8:658.8.012.12:005.21

Development of a product pricing algorithm using RFM strategy for user cohorts using machine learning methods

Vitaliy M. Kobets[1)]
ORCID: https://orcid.org/0000-0002-4386-4103; vkobets@kse.org.ua. Scopus Author ID: 56006224700
Ivan D. Shlyakhtenko[1)]
ORCID: https://orcid.org/0009-0007-1660-4334; ivan.shliakhtenko@university.kherson.ua
Serhii M. Zinchenko[2)]
ORCID: https://orcid.org/0000-0001-5012-5029; srz56@ukr.net. Scopus Author ID: 57214802895
[1)] Kherson State University, 27, Universitetska Str. Kherson, 73003, Ukraine
[2)] Kherson State Maritime Academy, 20 Ushakova Av. Kherson, 73000, Ukraine

ABSTRACT

This paper presents a comprehensive approach to developing a dynamic product pricing algorithm that integrates RFM (Recency, Frequency, Monetary) customer segmentation strategy with machine learning methods. The research addresses the critical challenge of personalizing pricing strategies for different user cohorts in competitive retail environments, particularly for supermarkets and online marketplaces. The study extends the classical RFM model by incorporating price elasticity of demand coefficients to create a cohort-based pricing framework. Using the K-Nearest Neighbors (KNN) algorithm, the authors developed an automated classification system that accurately segments customers based on their behavioral characteristics, achieving 100% classification accuracy for repeat purchasers. The methodology comprises five key stages: data collection and preprocessing, RFM score calculation using a quartile-based ranking, customer segmentation into five distinct groups (Champions, Loyal Users, Potential Loyalists, At-Risk, and Lost), training a machine learning model for segment prediction, and cohort formation based on the timing of the first purchase. The research implements a pricing algorithm that selectively applies discounts to target segments ("Potential Loyalists" and "At Risk" customers) who demonstrate inelastic demand and have maintained activity for at least six months. Experimental results demonstrate that strategic discount application not only reduces customer churn but also increases overall revenue through enhanced purchase volume. The proposed framework, implemented in the R programming language within the RStudio environment, provides businesses with a data-driven decision support tool for optimizing personalized pricing strategies while maintaining profitability. This approach enables companies to balance customer retention efforts with revenue maximization by utilizing sophisticated behavioral analytics and predictive modeling.

Keywords: RFM analysis; customer segmentation; machine learning; dynamic pricing algorithm; cohort analysis

***For citation*: Kobets V. M., Shlyakhtenko I. D., Zinchenko S. M. "Development of a product pricing algorithm using RFM strategy for user cohorts using machine learning methods". *Herald of Advanced Information Technology*. 2025; Vol.8 No.4: 497–509. DOI: https://doi.org/10.15276/hait.08.2025.32**

INTRODUCTION

In the current digital era, where competition between companies for consumer attention is constantly increasing, effective customer relationship management and personalized interaction strategies are gaining significant attention. The use of data analytics and the latest technologies, in particular machine learning, opens up new horizons for improving business processes, especially in the field of pricing. In such conditions, it is essential to develop an approach that allows not only to analyze customer behavior but also to adapt pricing policies in accordance with their characteristics.

In this context, the RFM strategy is relevant, which allows for segmenting customers according to three key criteria: the recency of the last purchase (Recency), the frequency of purchases (Frequency), and the monetary value (Monetary) [1]. Combining this approach with machine learning methods opens up the possibility of creating a flexible pricing algorithm tailored to specific user cohorts. This approach not only enables a better understanding of the needs of different segments but also allows for adapting sales strategies, thereby increasing the overall revenue of the business.

Thus, the development of a dynamic pricing algorithm based on RFM analysis and machine learning is not only theoretically important but also has practical significance for supermarkets, marketplaces, and other platforms that work with large amounts of customer data [2]. This approach enables informed pricing decisions, increases customer satisfaction, and enhances the overall efficiency of the business model.

Dynamic pricing is the process of flexibly adjusting prices in response to changes in demand or other factors. This approach is widely used in industries where demand fluctuates significantly over time, such as hotels and airlines, where prices increase during peak times and decrease during off-peak hours. Businesses can also use dynamic pricing to encourage sales during off-peak times, such as by offering early bird discounts at restaurants.

A strategy that takes into account the price elasticity of demand has a special place in the pricing system. With this strategy, it is possible to adjust the price formation according to the extent to which a price change will affect sales volume [3]. At the same time, many approaches include a restricted pricing strategy for customer segments, but do not propose an individual approach for different cohorts and personalized valuable propositions. This paper fills this gap.

The paper **aims** to develop a product pricing algorithm utilizing the RFM strategy for user cohorts, employing machine learning methods.

The structure of the paper is as follows: sect. 2 includes a literature review, sect. 3 describes RFM analysis, sect. 4 considers machine learning models for classifying user segments, sect. 5 investigates RFM analysis and ML model for user cohorts, sect. 6 proposes a cohort pricing algorithm and revenue discounts for users. The last section concludes.

RELATED WORKS

Developing accurate, personalized models requires collecting data that contains information about alternatives, context, customer characteristics, and customer behavior over time. The model has two types of personalization: preferences and price sensitivity at the individual-context level. Price sensitivity may differ, for example, depending on the contents of the consumer's basket, the day of delivery, or the customer's preferred time slot [4].

Recency, Frequency, Monetary analysis is a powerful marketing database technique widely used to rank customers based on their previous purchase history [5]. Using repeat purchase behavior based on the RFM model for customer classification and segmentation is a key area of retail research. This approach enables the identification of patterns in customer behavior by analyzing data on their transactions [6].

The customer segmentation model, based on the RFM (Recency, Frequency, and Monetary) model and utilizing the hierarchical Formal Concept Analysis (FCA) approach, enables the identification of not only customer segments but also the relationships between customer data in a comprehensible form [7].

Exploratory data analysis and pre-processing include [7]:

– Identification and collection of customer transaction data containing information on invoice number, product code, description, quantity, invoice date, price, customer ID, and country.

– Data cleansing includes removing missing values, canceled transactions (e.g., invoices starting with the letter 'c'), and transactions with zero or negative values.

– Calculation of RFM values: Three key variables are calculated for each unique customer: Recency: the number of days since the customer's last purchase; Frequency: the total number of purchases made by the customer; Monetary: the total amount of money spent by the customer in the store. Monetary value requires creating an aggregate variable, Total, by multiplying Quantity by Price for each transaction and then summing these values for each customer.

– RFM scores: the quintile method is used. This process enables the calculation of RFM scores for each customer, which serves as the basis for further developing a knowledge structure.

When segmenting customers, the monetary score is often the most important variable for dividing customers into groups.

We need to integrate an additional Diversity parameter into the classic RFM model to enhance customer analysis and predict customer responses. The Diversity parameter measures the number of different products purchased by a particular customer during the period under study. The Diversity parameter is introduced to identify customers who are open to new product offerings and are willing to test new products without hesitation, regardless of their overall budget. Such customers tend to have a high probability of purchasing the company's new products [8]. The RFM-D model aims to enhance the accuracy of customer behavior forecasting, enabling companies to predict which customers are likely to respond positively to new offers.

Segmentation using the RFM-D model yields better accuracy in predicting customers who are likely to respond positively to new offers compared to the classic RFM model. The "Variety" parameter helps identify customers who are "accustomed to shopping" and like to test new products.

Machine learning algorithms such as k-means and hierarchical clustering are used to identify customer segments. In the study, the k-means clustering algorithm was employed to construct consumer profiles, and the optimal number of clusters was identified using the elbow and silhouette coefficient methods [9]. The k-means

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

clustering based on a data-driven approach is used to characterize organizational identification using machine learning [10]. The optimal number of clusters is determined using the elbow method and silhouette scores [11]. After clustering the RFM-D model, a different number of clusters is generated compared to the classical RFM model. The RFM-D model shows a strong correlation between "Diversity" and "Frequency" / "Monetary value", which improves the analysis compared to the classic RFM model [8].

Fuzzy C-Means clustering allows a particular data point to belong to more than one cluster, which can be an advantage for large and similar datasets, as it increases the chances of customer retention by offering different offers for each segment [5].

After clustering, the behavior of each cluster is analyzed to identify groups of customers that generate the most profit for the company. A detailed analysis of these clusters helps to identify target customers and provide them with appropriate promotions and offers.

Segmentation of customers deepens relationships with them and enables the company to tailor its marketing strategies to individual customers based on their purchasing behavior. Retaining existing customers is considered more important than attracting new ones.

Research [9] develops a framework based on large language models (LLMs) for clustering and analyzing consumer survey data to study the effectiveness of LLMs in marketing research. Based on natural language processing (NLP) principles, the method converts aggregated text responses into vector representations that can then be subjected to clustering algorithms to generate refined consumer profiles. LLMs improve clustering accuracy compared to traditional models. Chatbots created using LLMs have achieved over 89% accuracy in simulating consumer preferences [9].

Machine learning methods are used to classify customers into different segments using a specially designed multilayer perceptron (MLP), as well as support vector machine (SVM) and decision tree (DTC) classification methods. Empirical analysis has shown that eight transactions are sufficient to classify a customer [6] accurately.

The Lifetime value (LTV) metric is used to compare the value of different customer segments [6]. LTV is a representative value for each customer, based on which the business determines its customer outreach strategies. The distribution of all RFM features characterizes the cluster with the highest LTV. The cluster with the lowest LTV usually has low or medium RFM values.

Development of targeted marketing strategies and recommendations means [7]: Companies should reward high-value customers, maintain close relationships, and promote new products and the brand; For customers at risk of losing them, understand the reasons for their inactivity and take measures to retain them; For customers with average frequency and sales, offer promotions and recommend products for cross-selling to increase the volume of purchases; Offer special offers for new or less active customers to increase frequency and purchases.

Tailored strategies enable luxury brands to develop more targeted and effective marketing approaches, taking into account the diverse motivations of each customer segment in the dynamic NFT market. The study allows us to identify clear customer segments in the luxury NFT market that differ in their characteristics and behavior [11].

Using RL, such as Q-Learning and Deep Q-Networks, to modify real-time interaction techniques is one of the most recent methods. Although these techniques are already being actively employed in other fields, such as robotics, their potential for marketing segmentation is expanding. This field's research demonstrates how to apply reinforcement methods to identify the most effective ways to communicate with customers in dynamic market environments [12]. Employing such flexible methods enables companies to focus on their customers' evolving requirements and preferences while also optimizing marketing techniques.

For customer segmentation in this research, we combine RFM analysis with ML techniques, which yields more precise and adaptable segmentation outcomes than conventional techniques. The primary objective is to apply an advanced segmentation technique that considers consumers' demographic and behavioral traits, utilizing ML training to classify further and refine marketing strategies, and provide customized offers to clients. The process involves multiple steps that combine contemporary ML techniques with traditional data analysis methodologies.

RFM STRATEGY AS A TOOL FOR FORMING USER COHORTS

In the current business environment, where competition is increasing and consumers expect a more personalized approach, segmentation and user cohort formation are becoming indispensable tools for effective personalized pricing. Understanding the diversity of customer behavior enables companies to tailor their pricing strategies, thereby enhancing customer satisfaction and boosting revenue [13].

For companies to effectively interact with their audience, it is essential not to view the customer base as a homogeneous set, but to segment it into distinct groups based on specific criteria. This approach enables a deeper understanding of the needs, motivations, and behaviors of different consumer types, allowing for the formation of personalized communication and pricing strategies [14].

One of the most effective tools for achieving this level of customer understanding is segmentation [15, 16]. Segmentation is the process of dividing a customer base into segments based on shared characteristics or behavioral traits. The basic principles of segmentation [17] include (Fig. 1):

- Demographic: age, gender, income, and education.
- Geographic: country, region, city.
- Psychographic: values, interests, lifestyle.
- Behavioral: purchase frequency, loyalty, reaction to discounts.
- Socio-economic: education, employment, income level.

These principles enable businesses to gain a deeper understanding of their customers and effectively target marketing campaigns

Fig. 1. **Principles of market segmentation**
***Source*: compiled by the authors**

.Performing RFM analysis manually is extremely time-consuming and inefficient. Thanks to this tool, companies can quickly determine which customers it is advisable to direct resources to, and which ones should be temporarily excluded from active interaction, since they do not bring the expected profit. The results of the RFM analysis enable us to understand how to best interact with each customer segment. For example, users who have recently made their first purchase with a high check have the potential to become regular customers. To increase the likelihood of this, send them personalized messages via email, SMS, or chatbot with an offer that emphasizes the value of the brand and the benefits of further cooperation.

Conducting an RFM analysis includes 5 key steps [17].

1. Data collection: The initial step is to collect information about customer transactions. This includes key parameters such as purchase dates, quantity, and cost of each purchase.

2. Determining Recency, Frequency, and Monetary values for each user. We calculate the recency (subtract the value of the last purchase of a specific customer from the last value in the selected time interval), frequency (find the total number of purchases made by the user during the selected period), and monetary value (determine the amount of customer spending over a specific time interval).

3. Customer Scoring. At this stage, each customer is given a rating score for each RFM value. Typically, a five-point scale is used, with 5 indicating the most desirable user behavior and 1 indicating the least active customer.

4. Segmentation of the customer base. The obtained scores for each parameter are analyzed to classify customers into segments. Different combinations of RFM values enable the strategic highlighting of key customer groups. At this stage, it is essential to consider the specific details of the business to ensure an accurate interpretation of the data. One method of forming segments based on previously obtained estimates involves determining the sum of each RFM value, followed by the definition of user groups.

5. Developing personalized marketing strategies. The final stage involves creating targeted marketing approaches for each user segment. Communication, offers, and discounts should meet the needs and motivations of a specific consumer group.

Typical customer segments and engagement strategies are defined in Table 1.

In addition to the classic RFM analysis, there are its modifications that can take into account different business specifics. Additionally, one of the following steps after conducting RFM segmentation is to form user cohorts based on the resulting segments. This allows us to group customers not only by behavioral indicators but also by periods, which enables a deeper analysis of their behavior and the adaptation of marketing strategies.

A cohort is a group of users who share a common characteristic, typically associated with a specific event that occurred during a particular period, such as the date of their first purchase or

Table 1. **Typical customer segments and engagement strategies**

Segment name	Segment description	Method of promotion
Champions	The most active and profitable customers are those who have recently made purchases, make them regularly, and spend significant amounts	Exclusive offers, priority access to new products, and personalized communication
Loyal users	Make purchases regularly and demonstrate consistent behavior, although their average check or the recency of their last purchase may be lower than that of champions	Maintain good relationships by forming loyalty programs, offering cross-selling and bonuses for recommending friends, so that with proper attention, they can turn into champions
Potential loyalists	May be new customers or those who have made only a few purchases, but for a reasonably large amount. Their order frequency is low, but they have the potential to become loyal customers	Adaptive communications: a series of greeting cards, discounts for second purchases, informational materials about the product or service that will help increase brand trust
Clients at risk	Buyers who previously demonstrated stable activity and could even spend significant amounts, but have not made any purchases recently	To reactivate them, use personalized reminders, repeated offers, and surveys to identify reasons for the decreased interest, and create time-limited promotions that encourage a return
Lost customers	Users who have not made purchases for a long time and do not show any activity (casual buyers, tourists)	it is More expedient for companies not to waste resources on them, but to focus on maintaining more promising segments

***Source*: compiled by the authors**

registration. Unlike segmentation, which can be dynamic and change over time, cohorts are fixed groups that allow for the tracking of changes in customer behavior over a specific period.

The primary difference between a segment and a cohort is that a segment is a group of users defined by specific characteristics that can change over time (e.g., age, income, behavior). In contrast, a cohort is a group formed based on an event that occurred at a particular point in time and remains unchanged.

Thus, RFM segmentation and cohort formation allow businesses to more accurately analyze customer behavior, evaluate the effectiveness of marketing activities, and develop personalized strategies for interacting within different user groups [3].

MACHINE LEARNING MODELS FOR CLASSIFYING USER SEGMENTS

The next step after conducting RFM analysis is the need for automated segmentation for new customers. Machine learning models, such as logistic regression, the KNN (k-Nearest Neighbors) algorithm, and decision trees, can effectively solve this problem. The choice of parameter k plays a crucial role in the model's accuracy. If k is too small, the model may be too sensitive to noise, while if k is too large, its ability to detect meaningful patterns may be reduced, leading to overfitting [18].

There are three main statistical approaches to determining the optimal value of k: cross-validation, the elbow method, and avoiding odd values in classification problems to prevent situations where neighbors with the same number of votes belong to different classes, which can complicate the final decision.

Decision tree regression is a ML method used to build predictive models. Its essence lies in the hierarchical division of the input space into subsets based on criteria, determined by the feature values. During the model-building process, a tree is created, where the internal nodes correspond to the data partitioning conditions, and the leaf nodes contain the predicted numerical values. This approach provides high interpretability of the results and allows modeling complex dependencies between variables.

Based on a review of previous data, we present a comparative Table 2 of the advantages and disadvantages of the considered machine learning models for classifying user segments.

The choice of a ML model for classifying users by RFM segments depends on the specifics of the task, the nature and volume of data, as well as the requirements for interpreting the results [19]. Such a

Table 2. **Comparative table of advantages and disadvantages of ML models**

Model	Advantages	Disadvantages
Logistic regression	Ease of implementation, interpretability, and efficiency on big data	Limitations in cases of nonlinear dependencies
KNN (k nearest neighbors)	Flexibility, no assumptions about data distribution	High computational complexity, sensitivity to the choice of metric, and parameter k
Decision tree regression	Interpretability, the ability to process different types of data	Prone to overfitting, need for tree pruning to avoid overfitting

***Source*: compiled by the authors**

comparison allows us to understand which model is appropriate to use in a particular case: logistic regression is optimal for tasks with linear dependencies and the need for fast and straightforward interpretation; the KNN algorithm provides flexible analysis of behavioral patterns, especially in the presence of clearly expressed clusters [20], although it requires significant computational resources; decision tree regression is distinguished by its clarity, the ability to work with different types of data and formulate visual classification rules, but requires control of the depth of the tree to avoid overfitting. In practice, ensemble methods are often used that combine the advantages of several models to achieve higher accuracy, stability, and scaling of results.

RFM ANALYSIS AND ML MODEL FOR USER COHORTS

We will conduct RFM analysis, calculate R, F, and M, and build ML models to classify and isolate user cohorts using open data set from Kaggle. The next step in conducting RFM analysis is assigning RFM scores to users. There are different ways to perform this procedure; in the paper, a division into four equal parts (quartiles) is used, resulting in a scale of scores from 1 to 4, where 1 is the worst score and 4 is the best. Also, we additionally calculate the summary indicators rfm_score (to construct a numerical segment code) and the overall score rfm_total (the sum of the three scores).

The next stage of RFM analysis is user segmentation. In the research, a customer with a value of 12 is assigned the “Champions” segment, from 9 to 11 – “Loyal users”, from 6 to 8 – “Potential loyalists”, from 4 to 5 – “At risk”, and users with a value of 3 – “Lost”. Let us build a distribution of the number of customers (Fig. 2) and purchases by segments (Fig. 3).

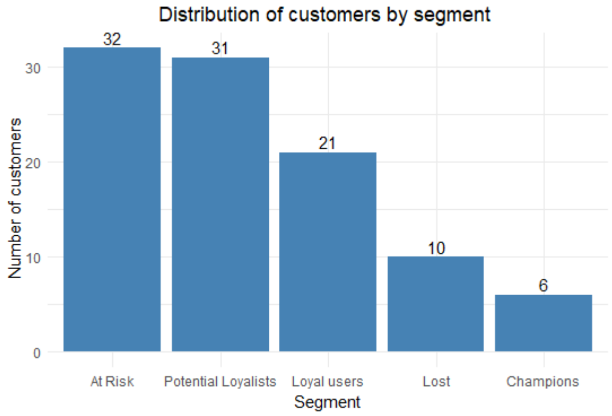

Fig. 2. **Distribution bar of the number of customers by segments**
***Source*: compiled by the authors**

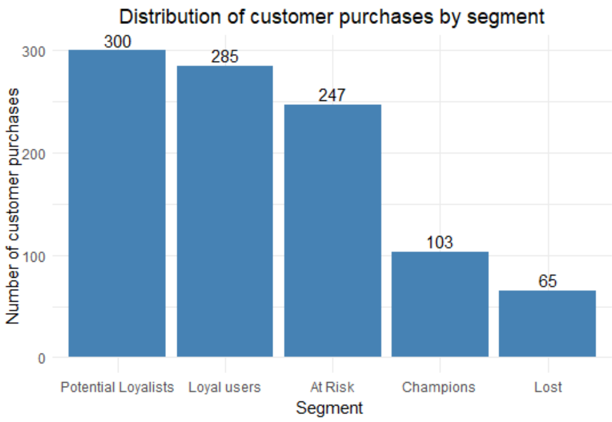

Fig. 3. **Distribution bar of customer purchases by segment**
***Source*: compiled by the authors**

From these figures, we can see that 84% of our customers belong to segments such as "Loyal users", "Potential loyalists", and "At risk". It is from these types of customers that we receive the most significant number of purchases. Additionally, we will create graphs of the distribution of customers by each RFM indicator (Fig. 4, Fig. 5, and Fig. 6).

After analyzing these graphs, we find that the most common recency indicators for customers are between 0 and 36 days and approximately 50 to 60 days. The most common frequency values for customers are between 7 and 13, and the most common monetary value indicators for customers are between 200 and 650 euro. We will also construct graphs of the distribution of total income (Fig. 7), average check (Fig. 8), and average spending by customer segments (Fig. 9).

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

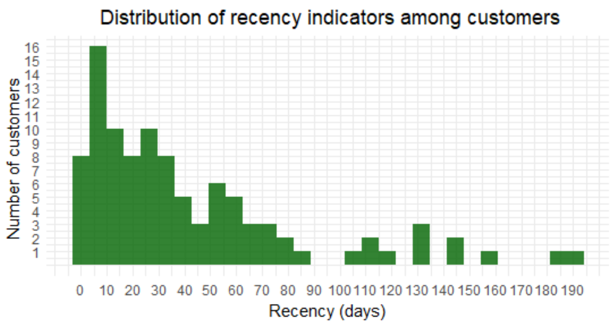

Fig. 4. **Distribution histogram of recency indicators among customers**
***Source*: compiled by the authors**

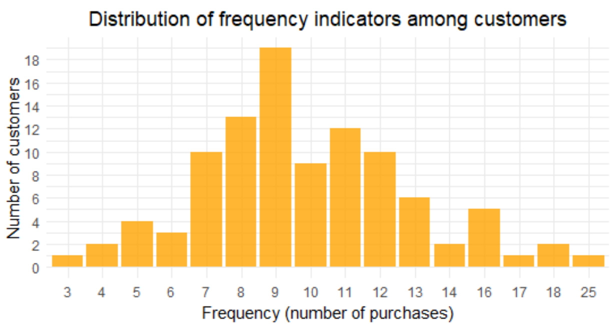

Fig. 5. **Frequency distribution histogram among customers**
***Source*: compiled by the authors**

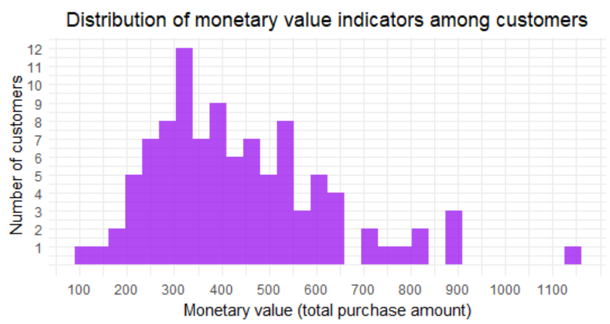

Fig. 6. **Distribution histogram of monetary indicators among customers**
***Source*: compiled by the authors**

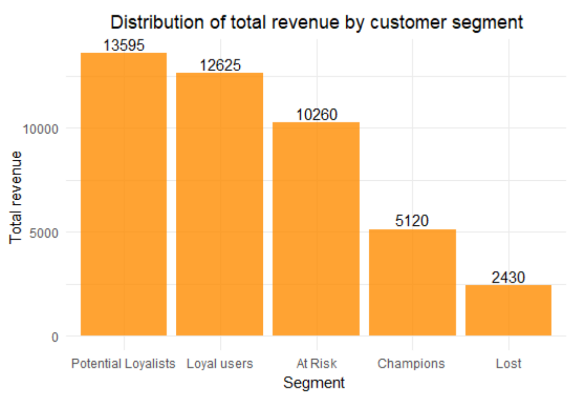

Fig. 7. **Total revenue distribution chart by client segments**
***Source*: compiled by the authors**

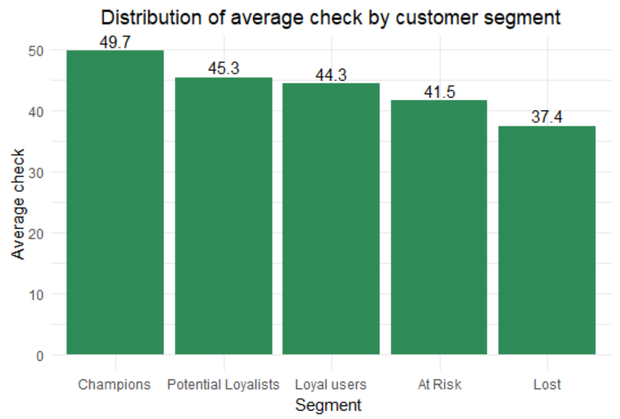

Fig. 8. **Average check distribution chart by client segments**
***Source*: compiled by the authors**

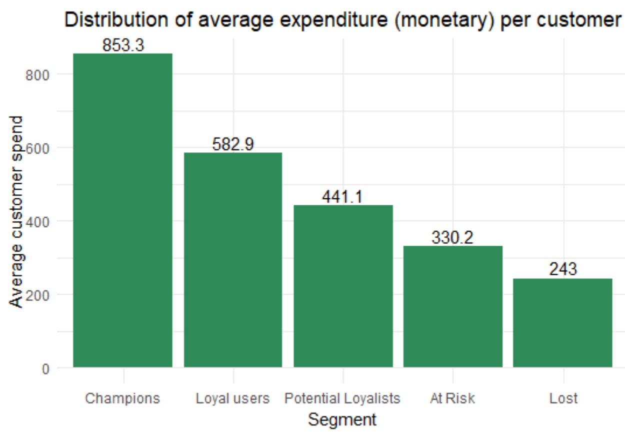

Fig. 9. **Distribution chart of average monetary expenditure per customer**
***Source*: compiled by the authors**

After analyzing these graphs, the following conclusions can be drawn:

1) The distribution of total revenue by customer segments is dominated by segments such as “Potential Loyalists”, “Loyal Users”, and “At Risk”. This once again confirms that it is from these customers that we receive the highest revenue.

2) The distribution of the average check by client segments does not differ significantly, but overall, the distribution hierarchy remains.

3) The distribution of the average monetary amount spent per customer shows that buyers from the "Champions" segment spend the most money, while spending gradually decreases from “Loyal Users” to “Lost”. This graph effectively illustrates the hierarchy of segments.

Elasticity indicators for all consumer segments are specified in the table below (Table 3).

Table 3 shows a clear pattern: the lower the RFM scores of user segments, the lower their elasticity coefficients determined in equation (1).

After conducting RFM segmentation and analyzing the previous graphs, we can see that calculating RFM indicator estimates by quartiles and forming segments using the rfm_total indicator, which is determined by the total sum of estimates, is appropriate for user segmentation.

Table 3. **Elasticity indicators for all consumer segments**

Segment	Average elasticity	Median elasticity	Users count
Champions	-2.40	-2.39	6
Loyal users	-2.57	-2.51	23
Potential loyalists	-2.74	-2.72	32
Clients at risk	-2.85	-2.77	31
Lost customers	-3.08	-2.88	10

***Source*: compiled by the authors**

The next step is to build a machine learning model to classify user segments. For this, we use the KNN algorithm, for which we create a separate data frame that will include the following data: a unique user ID, its age, frequency, and monetary value estimates, and a class whose value will be based on belonging to the corresponding segment, where 1 is "Lost" and 5 is “Champions”. Next, we create a training and test sample, in a ratio of 70 to 30 percent, respectively, using the function. The next step is to build a KNN model based on the training data, which includes the RFM estimates, the test data to be classified, the class vector for the training sample, and the number of nearest neighbors used to determine the class.

Let us create a graph based on the Elbow method to look at the dependence of the classification error rate on the value of the parameter k (Fig. 10).

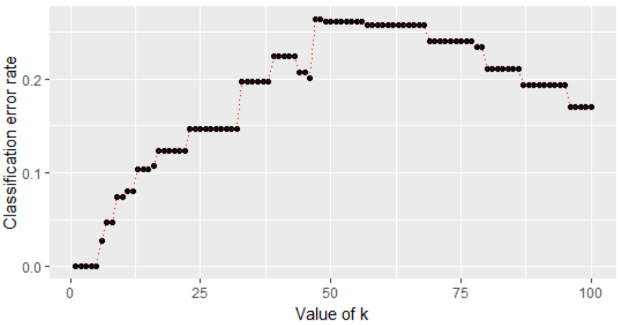

Fig. 10. **Elbow graph of the classification error rate on the value of the parameter k**
***Source*: compiled by the authors**

Analyzing this graph, it is clear that for parameter k values from 1 to 5, the classification error rate is zero. Starting from parameter k = 6, the error rate begins to increase. In cases where a person has already made multiple purchases, the KNN algorithm can easily classify them by segment. Additionally, when we have different observations for the same customer, this ML model provides a forecast with 100% classification accuracy (Table 4). Therefore, the KNN algorithm is suitable for use when customers perform repeated actions. Having performed these actions once, we can quickly determine which segment they belong to based on their actions.

Table 4. **Confusion matrix for test set of KNN model (k=2)**

	Actual				
Predicted	Cham-pions	Loyal users	Potential loyalists	Client at risk	Lost custo-mers
Champions	31	0	0	0	0
Loyal users	0	85	0	0	0
Potential loyalists	0	0	90	0	0
Clients at risk	0	0	0	74	0
Lost custo-mers	0	0	0	0	19

***Source*: compiled by the authors**

The next step is to form a cohort of users. Next, we form a cohort of users who made their first purchase of products in January 2023. After that, we will conduct RFM segmentation of this cohort of users for each month of 2023. The primary objective of this approach is to monitor the evolution of the customer segment and document these changes in a comprehensive matrix of segments by month.

Now, let us create a graph of the number of active users from the first cohort by month, as well as graphs of the number (Fig. 11) and share (Fig. 12, Fig. 13) of active users by segments from the first cohort by month.

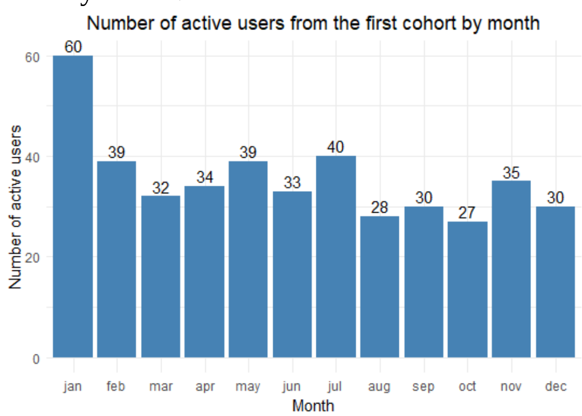

Fig. 11. **Graph of the number of active users from the first cohort by month**
***Source*: compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

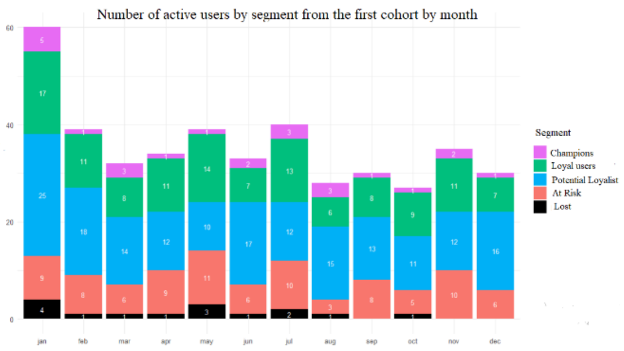

Fig. 12. **Graph of the number of active users by segments from the first cohort by month**
***Source*: compiled by the authors**

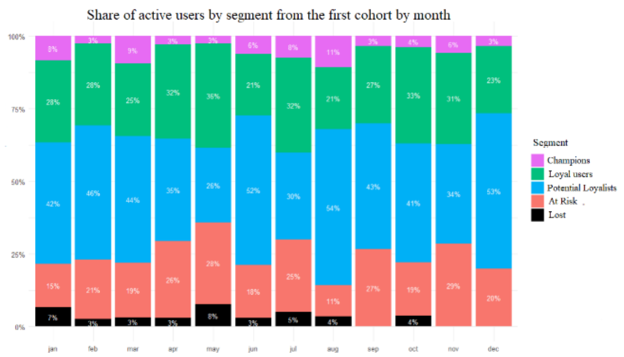

Fig. 13. **Graph of the share of active users by segments from the 1st cohort by month**
***Source*: compiled by the authors**

If we compare the January and February cohorts, we can see that there are no significant differences between these cohorts (Fig. 14).

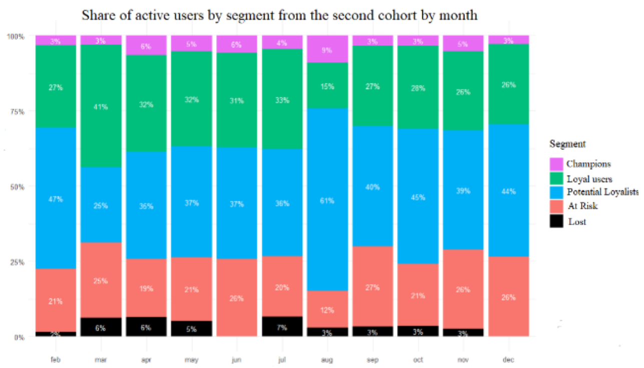

Fig. 14. **Graph of the share of active users by segments from the 2nd cohort by month**
***Source*: compiled by the authors**

The cohort under study has a relatively stable share of users in the "Loyal Users" and "Potential Loyalists" segments, but a noticeable churn is observed in the second month. To increase the effectiveness of cohort pricing, it is advisable to stimulate repeat purchases, thereby moving users to higher loyalty segments and reducing the probability of churn through personalized communications and targeted activities.

Increase the share of users in the “Champions” and “Loyal Users” segments by implementing a reward system, offering premium content, or providing exclusive terms of interaction. These segments have high monetary value indicators, which means that additional costs for attracting customers will be justified.

The application of KNN classification model demonstrated the ability to predict the segment of a new or existing customer based on their RFM scores.

Special attention was paid to the cohort analysis of users who made their first purchase in January 2023. For this cohort, a matrix of segment changes in dynamics by month was constructed, which allowed us to identify patterns of transitions between segments and assess the stability of customer loyalty throughout the year.

COHORT PRICING ALGORITHM AND REVENUE DISCOUNTS FOR USERS

After conducting RFM segmentation and building a user classification model, the next logical step is to develop a cohort pricing model. This approach not only allows for differentiating price offers according to the behavioral characteristics of users, but also for assessing the impact of price changes on total revenue. Using the price elasticity of demand coefficient opens up opportunities for more effective pricing for individual user segments, particularly by offering personalized discounts.

This section implements a pricing algorithm for Cohort 1, as previously defined, which involves providing price incentives to individual user segments to increase their activity. Additionally, a comparative analysis of actual and updated revenue before and after the implementation of changes in pricing policy is conducted, which enables an assessment of the effectiveness of the decisions made.

For further pricing, the price elasticity of demand coefficient is determined in equation (1), which shows how much the annual budget of the buyer's expenses on supermarket products will change if the share of his budget for one trip to the store changes by 1%. That is, instead of the percentage change in the number of products (Q) in the numerator, there is a change in the budget of the buyer's annual expenses on supermarket products, and instead of the percentage change in the price (P) in the denominator, there is a change in the share of expenses for purchases per one trip to the store in the buyer's annual budget.

$$e_{r,avr} = \frac{\ln(total\_revenue)}{\frac{\ln(avg\_order\_value)}{total\_revenue}} \quad (1)$$

For example, if the indicator equals -2, this means that a 1% increase in the share of expenses for one trip to the store will result in a 2% decrease in total annual expenses. The next step in determining one of the criteria by which to decide whether to provide a discount to the corresponding client is to determine the value of the segment's order for each user during the year.

Another criterion for providing a discount to eligible customers is the number of months in which users have made at least 1 purchase. Next, we combine the data into a data frame that will accept the values user_id, elasticity, segment_mode, and month_active.

We will select customers who will be given a discount based on the appropriate criteria, namely.

1. The number of months in which purchases were made must be greater than or equal to 6 within one year.

2. The segment's fashion value should correspond to categories such as "At Risk" and "Potential Loyalists."

3. The value of the price elasticity of demand must be less than or equal to -1 (inelastic demand).

Next, we calculate statistical indicators for the transactions of customers who will receive a discount, including the number of transactions, the average check, and the total revenue. Thus, we model the change in user behavior following the implementation of the updated pricing policy, enabling us to assess its financial impact. The next part of the program performs the final stage of estimating the impact of discounts on total revenue across the user cohort. First, the updated revenue, which consists of two parts.

1. Amounts of new revenue from users who were given discounts, taking into account changes in average check and number of transactions.

2. Initial revenue of users in the same cohort who were not given discounts.

Next, the actual (initial) revenue is calculated as the total of all orders from users in the cohort. This is followed by the absolute change in revenue, which is the difference between the updated and original revenue, as well as its relative change expressed as a percentage.

The final step will be to plot the change in revenue after applying discounts (Fig. 15).

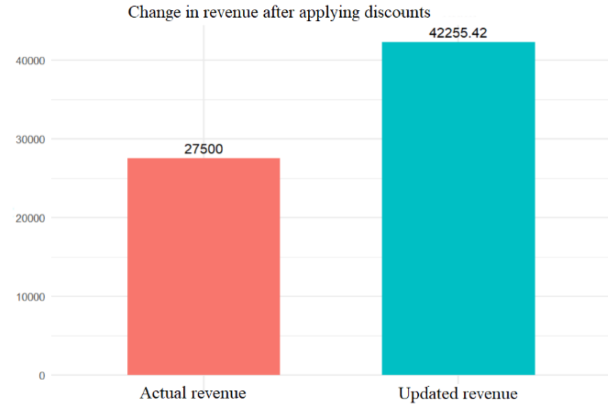

Fig. 15. **Revenue change schedule after applying discounts**

***Source*: compiled by the authors**

In this section, a cohort pricing model was developed, which enables the assessment of the impact of introduced discounts on revenue, taking into account the behavioral characteristics of users. Based on the previously formed cohort of users and the results of RFM analysis, target segments for providing discounts were identified (in particular, "Potential Loyalists" and "At Risk"). Using the price elasticity of demand coefficient, the change in the number of transactions and the average check after applying the discount was modeled for each user. As a result of the calculations, the expected change in total revenue for the specified cohort was estimated, allowing for the verification of the pricing policy's effectiveness and the formulation of recommendations for its improvement. The model demonstrated its practical feasibility as a decision-making support tool in the field of personalized pricing.

CONCLUSIONS

The paper analyzed modern approaches to pricing policy formation. The role of price personalization and the use of consumer data analytics in developing effective pricing models was assessed. This approach allows taking into account not only market factors, but also user behavioral characteristics, which is an important prerequisite for a more accurate impact on the company's revenue.

The effectiveness of RFM strategies for customer segmentation was investigated, which allowed us to identify groups of users based on the age of their last purchase, transaction frequency, and total spending. Forming cohorts based on these three

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

parameters provided a structured picture of the consumer base and allowed us to prepare data for subsequent cohort pricing.

Applying a KNN model enabled us to automate the process of classifying users by RFM segments, facilitating the formation of cohorts based on this segmentation. This model demonstrated high accuracy, confirming the feasibility of using this ML model for analyzing customer behavior and developing personalized approaches to working with them.

Additionally, an algorithm and product pricing model were implemented in the R programming language within the RStudio environment, enabling us to effectively process input data, build segments, visualize behavioral patterns, and implement ML models.

At the final stage, a cohort pricing model was built, and its impact on revenue was assessed based on input data. As a result of experimental modeling, a scenario of changes in the average check and the number of transactions was simulated under the influence of price incentives for individual user segments, using the price elasticity of demand. The analysis revealed that, with the correct application of discounts, it is not only possible to retain customers at high risk of churn, but also to increase overall revenue through increased purchase volume.

In our investigation, the RFM analysis is limited to the available data from the dataset. It can be expanded if more characteristics of purchase diversity are available, from RFM to RFM-D. Adding Diversity and Variety predictors makes it possible to define customer segments and cohorts in more detail, form a more personalized customer engagement strategy, and increase consumer engagement in purchases.

REFERENCES

1. Khajvand, M., Zolfaghar, K., Ashoori, S. & Alizadeh, S. “Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study”. *Proceedings of Computer Science.* 2011; 3: 57–63, https://www.scopus.com/pages/publications/79952519420?origin=resultslist.
DOI: https://doi.org/10.1016/j.procs.2010.12.011.

2. Kobets, V. & Yatsenko, V. “What can economic experiments reveal about the evolutionary effectiveness of supermarket strategies?” *Marketing and Management of Innovations.* 2018; 1: 152–168. DOI: https://doi.org/10.21272/mmi.2018.1-11.

3. Kobets, V., Yashyna, I. “Development of RFM strategies for customer segmentation using algorithmic methods”. *Marketing and Digital Technologies* 2025; 9 (3): 32–43. DOI: https://doi.org/10.15276/mdt.9.3.2025.3.

4. Gür, Ö. A. & Amorim, P. “Personalized choice model for demand forecasting under pricing scenarios with observational data-The case of attended home delivery”. *International Journal of Forecasting.* 2024; 40 (2): 706–720, https://www.scopus.com/pages/publications/85161032474?origin=resultslist. DOI: https://doi.org/10.1016/j.ijforecast.2023.04.008.

5. Christy, J., Umamamakeswari, A., Priyatharsini, L. & Neyaa, A. “RFM Ranking - an effective approach to customer segmentation”. *Journal of King Saud University - Computer and Information Sciences* 2021; 33 (10): 1251–1257, https://www.scopus.com/pages/publications/85052988327?origin=resultslist. DOI: https://doi.org/10.1016/j.jksuci.2018.09.004.

6. Rahim, M. A., Mushafiq, M., Khan, S.& Arain, Z. A. “RFM-based repurchase behavior for customer classification and segmentation”. *Journal of Retailing and Consumer Services.* 2021; 61: 102566, https://www.scopus.com/pages/publications/85104140032?origin=resultslist.
DOI: https://doi.org/10.1016/j.jretconser.2021.102566.

7. Rungruang, C., Riyapan, P., Intarasit, A., Chuarkham, K.& Muangprathub, J. “RFM model customer segmentation based on hierarchical approach using FCA”. *Expert Systems with Applications.* 2024; 237: 121449–121449, https://www.scopus.com/pages/publications/85171617109?origin=resultslist. DOI: https://doi.org/10.1016/j.eswa.2023.121449.

8. Smaili, M. Y.& Hachimi, H. “New RFM-D classification model for improving customer analysis and response prediction”. *Ain Shams Engineering Journal.* 2023, 14 (12): 102254, https://www.scopus.com/pages/publications/85152549205?origin=resultslist.
DOI: https://doi.org/10.1016/j.asej.2023.102254.

9. Li, Y., Liu, Y. &, Yu, M., "Consumer segmentation with large language models". *Journal of Retailing and Consumer Services.* 2025; 82: 104078, https://www.scopus.com/pages/publications/85203416745?origin=resultslist. DOI: https://doi.org/10.1016/j.jretconser.2024.104078.

10. Ybañez, A. et al. "How can we use machine learning for characterizing organizational identification - a study using clustering with Picture fuzzy datasets". *International Journal of Information Management Data Insights* 2023; 3 (1): 100157–100157, https://www.scopus.com/pages/publications/85147018032?origin=resultslist. DOI: https://doi.org/10.1016/j.jjimei.2023.100157.

11. Chen, Q., Choi, B.-J. & Lee, S.-J. "Tailoring customer segmentation strategies for luxury brands in the NFT market - The case of SUPERGUCC". *Journal of Retailing and Consumer Services.* 2024; 82: 104121, https://www.scopus.com/pages/publications/85206268017?origin=resultslist.
DOI: https://doi.org/10.1016/j.jretconser.2024.104121.

12. Alzubaidi, L. et al. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions". *Journal of Big Data.* 2021; 8 (1): 53, https://www.scopus.com/pages/publications/85103843272?origin=resultslist. DOI: https://doi.org/10.1186/s40537-021-00444-8.

13. Verma, R., Rathor, D., Kumar, S., Mishra, M. & Baranwal, M. "enhancing customer repurchase prediction: integrating classification algorithms with RFM analysis for precision and actionable insights". *IIMB Management Review.* 2025; 37 (2): 100574, https://www.scopus.com/pages/publications/105008536846?origin=resultslist. DOI: https://doi.org/10.1016/j.iimb.2025.100574.

14. Yang, C., Dong, W., Yuan, Q., Cao, R. & Yu, C. "Integrating user values assessment with the augmented RFM model to enhance bus transit systems' market segmentation". *Transport Policy.* 2025; 171: 50–62, https://www.scopus.com/pages/publications/105006875156?origin=resultslist.
DOI: https://doi.org/10.1016/j.tranpol.2025.05.028.

15. Kobets, V. & Novak, O. "EU countries were clustering for the state of food security using machine learning techniques". *Neuro-Fuzzy Modeling Techniques in Economics.* 2021; 10: 86–118, https://www.scopus.com/pages/publications/85161960584?origin=resultslist.
DOI: https://doi.org/10.33111/nfmte.2021.086.

16. Kobets, V., Yatsenko, V. & Voynarenko, M. "Cluster analysis of countries inequality due to IT development through macros application". *Communications in Computer and Information Science.* 2020; 1175: 415–439, https://www.scopus.com/pages/publications/85079094124?origin=resultslist.
DOI: https://doi.org/10.1007/978-3-030-39459-2_19.

17. Zaghloul, M., Barakat, S. & Rezk, A. "Enhancing customer retention in online retail through churn prediction: A hybrid RFM, K-means, and deep neural network approach". *Expert Systems With Applications.* 2025; 290: 128465, https://www.scopus.com/pages/publications/105007418868?origin=resultslist. DOI: https://doi.org/10.1016/j.eswa.2025.128465.

18. Tonmoy, A. K., Mahmood, T., Sevart, C., Ling, Y. & Wang, Y. "Explainable and Reduced-Feature Machine learning models for shape and drag prediction of a freely moving drop in the sub-critical Weber number regime". *Chemical Engineering Journal.* 2024; 501: 157476, https://www.scopus.com/pages/publications/85209408394?origin=resultslist.
DOI: https://doi.org/10.1016/j.cej.2024.157476.

19. Tang, Z., Jiao, Y. & Yuan, M. "RFM user value tags and XGBoost algorithm for analyzing electricity customer demand data". *Systems and Soft Computing.* 2024; 6: 200098, https://www.scopus.com/pages/publications/85191494563?origin=resultslist.
DOI: https://doi.org/10.1016/j.sasc.2024.200098.

20. Kobets, V. M., Gulin, O. V. & Nosov, P. S. "A cluster approach to matching the competencies of data specialists with skills in demand on the labor market". *Applied Aspects of Information Technology.* 2024; 7 (3): 231–241. DOI: https://doi.org/10.15276/aait.07.2024.16.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 06.10.2025
Received after revision 28.11.2025
Accepted 04.12.2025

DOI: https://doi.org/10.15276/hait.08.2025.32
УДК 004.8:658.8.012.12:005.21

Розробка алгоритму ціноутворення продуктів засобами RFM-стратегії для когорт користувачів методами машинного навчання

Кобець Віталій Миколайович[1)]
ORCID: https://orcid.org/0000-0002-4386-4103; vkobets@kse.org.ua. Scopus Author ID: 56006224700
Шляхтенко Іван Денисович[1)]
ORCID: https://orcid.org/0009-0007-1660-4334; ivan.shliakhtenko@university.kherson.ua
Зінченко Сергій Михайлович[2)]
ORCID: https://orcid.org/0000-0001-5012-5029; srz56@ukr.net. Scopus Author ID: 57214802895
[1)] Херсонський державний університет, вул. Університетська, 27. Херсон, 73003, Україна
[2)] Херсонська державна морська академія, пр-т Ушакова, 20. Херсон, 73000, Україна

АНОТАЦІЯ

У статті представлено комплексний підхід до розробки алгоритму динамічного ціноутворення на продукцію, який інтегрує стратегію сегментації клієнтів за RFM (Recency, Frequency, Monetary) з методами машинного навчання. Дослідження присвячене вирішенню важливої проблеми персоналізації стратегій ціноутворення для різних груп користувачів у конкурентному середовищі роздрібної торгівлі, зокрема для супермаркетів та онлайн-маркетплейсів. Дослідження розширює класичну модель RFM, включаючи коефіцієнти цінової еластичності попиту для створення системи ціноутворення на основі груп. Використовуючи алгоритм K-найближчих сусідів (KNN), розроблена автоматизована система класифікації, яка точно сегментує клієнтів на основі їх поведінкових характеристик, досягаючи 100% точності класифікації для постійних покупців. Методологія складається з п'яти ключових етапів: збір і попередня обробка даних, розрахунок балів RFM з використанням рейтингу на основі квартилів, сегментація клієнтів на п'ять окремих груп (чемпіони, лояльні користувачі, потенційні лояльні клієнти, клієнти в зоні ризику та втрачені клієнти), навчання моделі машинного навчання для прогнозування сегментів та формування когорт на основі часу першої покупки. Дослідження реалізує алгоритм ціноутворення, який вибірково застосовує знижки до цільових сегментів (клієнтів «Потенційні лояльні клієнти» та «Клієнти в зоні ризику»), які демонструють нееластичний попит і підтримують активність протягом щонайменше шести місяців. Результати експерименту показують, що стратегічне застосування знижок не лише зменшує відтік клієнтів, але й збільшує загальний дохід за рахунок збільшення обсягу покупок. Запропонована структура, реалізована на мові програмування R у середовищі RStudio, надає підприємствам інструмент підтримки прийняття рішень на основі даних для оптимізації персоналізованих стратегій ціноутворення, спрямованих на збереження прибутковості. Такий підхід дозволяє компаніям збалансувати зусилля з утримання клієнтів та максимізацію доходу за допомогою використання складних поведінкових аналітичних даних та прогнозного моделювання.

Ключові слова: RFM аналіз; сегментація клієнтів; машинне навчання; алгоритм динамічного ціноутворення; когортний аналіз

ABOUT THE AUTHORS

Vitaliy M. Kobets - Doctor of Economic Science, Professor, Department of Computer Science and Software Engineering. Kherson State University. 27, Universitetska Str. Kherson, 73003, Ukraine
ORCID: https://orcid.org/0000-0002-4386-4103; vkobets@kse.org.ua. Scopus Author ID: 56006224700
Research field: Data Science in Economics; Evolutionary Microeconomics; Robo-Advisor

Кобець Віталій Миколайович - доктор економічних наук, професор, професор кафедри Комп'ютерних наук та програмної інженерії. Херсонський державний університет, вул. Університетська, 27. Херсон, 73003, Україна

Ivan D. Shlyakhtenko - student of the bachelor's degree program on Information systems and technology. Kherson State University. 27, Universitetska Str. Kherson, 73003, Ukraine
ORCID: https://orcid.org/0009-0007-1660-4334; ivan.shliakhtenko@university.kherson.ua
Research field: Business analysis; data science; machine learning

Шляхтенко Іван Денисович - студент бакалаврської освітньої програми Інформаційні системи та технології кафедри Комп'ютерних наук та програмної інженерії. Херсонський державний університет, вул. Університетська, 27Херсон, 73003, Україна

Sergiy M. Zinchenko - Doctor of Engineering Science, Associated Professor, Department of Ship Handling at Sea. Kherson State Maritime Academy. 20 Ushakova Ave. Kherson, 73000, Ukraine
ORCID: https://orcid.org/0000-0001-5012-5029; srz56@ukr.net. Scopus Author ID: 57214802895
Research field: automatic control systems for moving objects

Зінченко Сергій Михайлович - доктор технічних наук, доцент кафедри Упраління судном. Херсонська державна морська академія, пр-т Ушакова, 20. Херсон, 73000, Україна