

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ
ВАДИМА ГЕТЬМАНА МОИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ
ВАДИМА ГЕТЬМАНА МОИ УКРАЇНИ

Кваліфікаційна наукова праця
на правах рукопису

КЛОЧКО РОСТИСЛАВ СЕРГІЙОВИЧ

УДК 330.45:519.866:[339.138:005.346]](043.5)

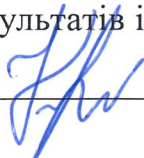
ДИСЕРТАЦІЯ

**МОДЕЛЮВАННЯ ПОВЕДІНКИ КЛІЄНТІВ У СИСТЕМІ
МАРКЕТИНГОВОГО МЕНЕДЖМЕНТУ**

Спеціальність 051 – Економіка

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело


_____ Р.С. Ключко

Науковий керівник: Піскунова Олена Валеріївна, доктор економічних наук,
професор

Київ – 2024

АНОТАЦІЯ

Клочко Р.С. Моделювання поведінки клієнтів у системі маркетингового менеджменту. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 051 «Економіка». – Київський національний економічний університет ім. Вадима Гетьмана, Київ, 2024.

Дисертаційну роботу присвячено розробці та реалізації нового концептуального підходу до вирішення наукової задачі моделювання поведінки клієнтів в системі маркетингового менеджменту. Цей підхід дозволяє перетворювати наявні первинні дані про клієнтів на знання про закономірності їх поведінки та надає практичні переваги при розробці маркетингових стратегій. Об'єктом дослідження є поведінка клієнтів з точки зору маркетингового менеджменту. Предметом дослідження є економетричні моделі та методи машинного навчання, що використовуються для моделювання поведінки клієнтів. Практична цінність дослідження полягає у розробці концептуальних положень щодо моделювання поведінки клієнтів, які можуть бути впроваджені у бізнес-процеси комерційних організацій для удосконалення маркетингового менеджменту, зокрема для підвищення ефективності процесу прийняття маркетингових рішень; формування дієвої маркетингової стратегії; оптимізації операційних витрат та маркетингових інвестицій; підвищення рівнів утримання та лояльності клієнтів; реалізації результативної комунікаційної стратегії; розробки та удосконалення продуктів і послуг.

У нинішньому конкурентному бізнес-середовищі розуміння поведінки клієнтів має вирішальне значення для комерційних організацій, які прагнуть розробляти ефективні маркетингові стратегії. Щоб задовольнити вимоги сучасної економіки, комерційні організації повинні завжди враховувати думку своїх клієнтів при прийнятті рішень. Проте це є складним завданням, за умови обслуговування тисяч клієнтів щодня. Визначено, що економіко-математичне моделювання має вирішальне значення для вирішення цієї задачі, оскільки воно пропонує системний підхід для розуміння складних закономірностей поведінки

клієнтів, які неможливо оцінити за допомогою традиційного аналізу. Економіко-математичне моделювання дозволяє комерційним організаціям швидко приймати рішення на основі даних, що є критично важливим у постійній боротьбі за лояльність клієнтів, підвищення ефективності маркетингової активності та збільшення продажів. Тим не менш, підходи до моделювання також повинні враховувати практичні перешкоди, накладені реальним бізнес-середовищем. Наприклад, деякі комерційні організації не мають достатніх ресурсів або якісної інфраструктури даних, що перешкоджає впровадженню підходів щодо моделювання поведінки клієнтів у системі маркетингового менеджменту. Сьогодні є актуальним дослідження, яке зосереджене на комплексному вивченні особливостей моделювання поведінки клієнтів як за умови обмежень в обсягах і якості маркетингових даних, так і за умови якісної ІТ інфраструктури.

У роботі проаналізовано розвиток поняття маркетингового менеджменту. Розглянуто сутність та фундаментальні принципи реалізації маркетингового менеджменту, а також виокремлено його основні сучасні тенденції, зокрема: впровадження клієнтоорієнтованого підходу в організаціях; увага до персоналізованих маркетингових стратегій; використання аналізу на основі даних для отримання уявлення про потреби та поведінку клієнтів; підвищення ролі ефективного управління витратами в маркетинговому менеджменті. Обґрунтовано, що усі дії, що їх здійснює комерційна організація, підпадають під маркетинг і відіграють вирішальну роль у визначенні загального клієнтського досвіду. Проведений аналіз показав, що в сучасному маркетинговому менеджменті відбувся перехід від фокусу на товарах до фокусу на клієнтах. Обґрунтовано зниження ефективності масового маркетингу та важливість процесів управління гетерогенністю клієнтів в сучасному бізнес-середовищі. На основі детального аналізу наукової літератури сформовано бачення, що комерційні організації можуть досягти справжнього успіху і встановити міцні зв'язки з клієнтами, якщо надаватимуть пріоритет розумінню їх поведінки і

персоналізації їхнього досвіду. Автором обґрунтовано важливість обміну цінностями між клієнтами та комерційними організаціями.

Розглядаючи значення поняття «інтелектуальний маркетинг», було зроблено висновок, що технології моделювання відіграють важливу роль у маркетингу, оскільки вони можуть виявити неефективні маркетингові активності комерційних організацій та підвищити їхню загальну результативність за допомогою вдосконалення процесів прийняття рішень. Розглянуто останні тренди в галузі інтелектуального маркетингу, які включають великі дані, діджиталізацію, життєву цінність клієнта, користувацький досвід та персоналізацію. Обґрунтовано, що прийняття комплексної стратегії збору даних є життєво важливим для комерційних організацій. Запропоновано не обмежуватися збором лише поведінкових даних, а також додавати до аналізу дані третіх сторін, включаючи відкриті дані про організації, клієнтів і конкурентів. Додатково проаналізовано бар'єри, що перешкоджають потенціалу впровадження процесів моделювання в маркетинговий менеджмент, включаючи суб'єктивні судження осіб, що приймають управлінські рішення, суперечливі короткострокові та довгострокові результати впровадження, а також недостатній рівень володіння методологіями моделювання. Обґрунтовано, що рівень складності алгоритму прямо корелює з рівнем складності його пояснення. Пріоритетність обчислювальної гнучкості та здатності швидко отримувати та впроваджувати результати моделювання часто домінує над перевагами, які дає більш багата на ознаки модель.

Вперше розроблено концептуальні положення щодо моделювання поведінки клієнтів у системі маркетингового менеджменту з використанням економетричних моделей та алгоритмів машинного навчання, що, на відміну від існуючих підходів, дозволяє виявляти закономірності клієнтської поведінки, як за умови обмежень в обсягах і якості маркетингових даних, так і за умови якісної IT інфраструктури даних.

Другий розділ дисертаційної роботи присвячено розробці загального підходу до моделювання поведінки клієнтів при обмеженості даних. Автором

обґрунтовано, що низька якість даних не тільки ускладнює початковий аналіз, але й може призвести до помилкових інтерпретацій результатів моделювання та неефективної розробки маркетингової стратегії. Запропоновано типологію даних, яка включає поділ на комерційні та відкриті. Наведено як плюси, так і мінуси кожного типу. Машинне навчання визнано найкращим методом для роботи з обмеженими даними. Розроблено підхід щодо моделювання поведінки клієнтів в умовах обмеженості даних, що складається з дев'яти етапів. Автором удосконалено модель прогнозування поведінки бізнес-клієнтів консалтингової агенції на підґрунті методів машинного навчання, шляхом систематизації підходів щодо роботи з неякісними даними та підвищення ефективності моделювання, за рахунок реалізації процесу автоматичного налаштування гіперпараметрів та проведення множинних перехресних перевірок метрик оцінки якості моделей. На основі розробленого підходу було проведено комп'ютерну реалізацію та впровадження комплексу прогнозних моделей, які дозволяють передбачати ймовірність фінансового успіху бізнес-клієнтів консалтингової агенції. Було використано три методи: дерево класифікації та регресії, логістична регресія та випадковий ліс. Дерево класифікації та регресії було визначено найефективнішим алгоритмом. Розроблений підхід є корисним для маркетологів та консультантів, які можуть використовувати побудовані моделі для розробки ефективної маркетингової стратегії та підвищення ефективності діяльності бізнес-клієнтів.

Ще одним експериментальним дослідженням у Розділі 2 була побудова системи моделей, яка дозволила передбачити час, необхідний для здійснення наступного замовлення клієнта в інтернет-магазині. Автором удосконалено модель прогнозування поведінки клієнтів в електронній комерції за рахунок використання в якості фактору моделі результатів сегментації клієнтів, що були отримані шляхом попередньої кластеризації на підґрунті методів машинного навчання, що дозволило більш точно спрогнозувати майбутню купівельну активність. Спочатку було використано метод кластеризації k-середніх для сегментації клієнтів інтернет-магазину відповідно до показників купівельної

активності RFM. На основі розрахованих значень показників кластеризації було визначено шість груп клієнтів: 1. втрачені клієнти; 2. новий оптовий покупець; 3. клієнти, яких ми скоро втратимо; 4. активний роздрібний покупець; 5. новий роздрібний покупець; 6. активний оптовий покупець. На другому етапі було реалізовано п'ять алгоритмів класифікації: лінійний дискримінантний аналіз, дерево рішень, метод опорних векторів, випадковий ліс і k-найближчих сусідів, для прогнозування часу до наступного замовлення. Алгоритм випадкового лісу показав найкращу прогнозну силу. Результати моделювання та подальшого оцінювання дозволили зробити висновки, що запропонований підхід щодо моделювання дозволяє отримати цінні знання про закономірності поведінки клієнтів з первинних даних, незалежно від їх розміру, повноти та якості.

Розділ 3 дисертаційної роботи присвячений експериментальному дослідженню моделювання поведінки клієнтів банку. Обґрунтовано вигідне становище банків порівняно з іншими сферами економічної активності, за рахунок оперування великими об'ємами даних. Автором виокремлено зовнішні обставини, які мають значний вплив на банківську сферу, і можуть змінити особливості співпраці між банком і клієнтами, включаючи регуляторні зміни, економічні та ринкові тенденції. Проаналізовано поведінку бізнес-клієнтів з рахунками в євро та роздрібних клієнтів з автокредитом. Встановлено, що ці сегменти є найбільш прибутковими для банку, однак через особливості їхньої поведінки, ефективний маркетинговий менеджмент стосовно цих клієнтських груп, є складним завданням. Вперше було побудовано економетричні моделі виживання Кокса та Вейбулла для аналізу тривалості клієнтської активності, які, на відміну від існуючих підходів, дозволяють визначати фактори, що впливають на рівень утримання клієнтів, за умови цензуруваних спостережень. На першому етапі для кожної соціально-демографічної характеристики клієнта було створено окремі моделі виживання Вейбулла. Функція stepAIC була використана для пошуку найкращої комбінації характеристик клієнта та рівнів цих характеристик, щоб покращити якість побудованих моделей. Клієнти, які показали найвищий ризик передчасного закриття, включали людей старше 55

років, які проживають на заході України, і заробляють більше 150 тисяч гривень на місяць. Вплив продажів кредитних карток на виживання клієнтів було визначено на другому етапі, за допомогою моделі пропорційних небезпек Кокса. Згідно з аналізом кривих виживання, клієнти з кредитними картками мають статистично довший період активності. Однак припущення про пропорційність небезпек для цієї змінної не було підтверджено. Таким чином, було розроблено модель виживання для фактору «активована кредитна картка». На основі запропонованого підходу до моделювання поведінки клієнтів визначено, що продаж кредитної картки дозволяє пролонгувати активність клієнтів з кредитом на авто. Однак критичним фактором успіху є продаж кредитної картки та її активація.

Дістали подальшого розвитку практичні аспекти побудови цінової стратегії для бізнес-клієнтів банку за допомогою розробки маркетингової сегментації, яка, на відміну від існуючих підходів, базується на алгоритмі кластеризації k-середніх. У процесі розрахунку алгоритму було виявлено шість груп схожості, які можна охарактеризувати як: 1. недохідні транзактори з накопиченням євро; 2. збиткові накопичувачі євро; 3. прибуткові накопичувачі євро; 4. прибуткові транзактори з євронакопиченням; 5. клієнти без євронакопичення; 6. прибуткові транзактори без євронакопичення. Для кожного сегмента була розроблена стратегія ціноутворення для обслуговування рахунків в євро. На основі аналізу даних було виявлено, що впровадження розроблених рекомендацій дозволить оптимізувати операційні витрати банку без встановлення комісії для всіх бізнес-клієнтів.

Ключові слова: маркетинг, маркетингова активність, маркетинговий менеджмент, клієнт, поведінка клієнтів, сегментація, утримання клієнтів, стратегія, підтримка прийняття рішень, моделювання, прогнозування, регресія, економіко-математичне моделювання, економетричні моделі, машинне навчання, класифікація, кластеризація.

ABSTRACT

Klochko R. S. Modeling of client behavior in the marketing management system.

– Qualifying scientific work on manuscript rights.

Thesis for the degree of Doctor of Philosophy in the specialty 051 «Economics».

– Kyiv National Economic University named after Vadym Hetman, Kyiv, 2024.

The thesis is devoted to the development and implementation of a new conceptual approach to solving the scientific problem of modeling client behavior in the marketing management system. This approach allows to transform the available primary data about clients into knowledge about the patterns of their behavior and provides practical advantages in the development of marketing strategies. The object of research is client behavior from the marketing management point of view. The subject of the study is econometric models and machine learning methods used to model client behavior. The practical value of the study is the development of conceptual frameworks for modeling client behavior that can be implemented in business processes of commercial organizations to improve marketing management, in particular, to increase the efficiency of marketing decision-making; develop an effective marketing strategy; optimize operating costs and marketing investments; increase client retention and loyalty; design an successful communication strategy; develop and improve products and services.

In today's competitive business environment, understanding client behavior is crucial for commercial organizations aiming to develop effective marketing strategies. To meet the demands of today's economy, businesses must always take their clients' opinions into account in their decision-making process. However, this is a challenging task, while serving thousands of clients every day. It is determined that economic and mathematical modeling is crucial for solving this problem, as it offers a systematic approach to understanding complex patterns of client behavior that cannot be evaluated using traditional analysis. Economic and mathematical modeling allows commercial organizations to make data-driven decisions quickly, which is critical in the constant struggle for client loyalty, for increasing the effectiveness of marketing activities, and for increasing sales. Nevertheless, modeling approaches must also take into account

the practical obstacles imposed by the real business environment. For example, some commercial organizations do not have sufficient resources or a good data infrastructure, which hinders the implementation of client behavior modeling approaches in the marketing management system. Today, a research that focuses on a comprehensive study of the features of client behavior modeling both under the conditions of limitations in the volume and quality of marketing data and under the condition of a high-quality IT infrastructure is highly actual.

In the work the development of the marketing management concept is analyzed. The essence and fundamental principles of marketing management are reviewed, and its main current trends are highlighted, in particular: the implementation of a client-oriented approach in organizations; attention to personalized marketing strategies; the use of data-driven analysis to generate insights about client needs and behavior; and the increasing role of effective cost management in marketing management. It is substantiated that all actions performed by a commercial organization are considered as marketing and play a crucial role in determining the overall client experience. The analysis has shown that in modern marketing management there has been a shift from a focus on products to a focus on clients. The decrease in the effectiveness of mass marketing and the importance of client heterogeneity management processes in the modern business environment are substantiated. Based on a detailed analysis of the scientific literature, the author suggests that commercial organizations can achieve real success and establish strong relationships with clients if they prioritize understanding their behavior and personalizing their experience. The author substantiates the importance of value exchange between clients and commercial organizations.

By examining the meaning of the concept of "intelligent marketing", it is concluded that modeling technologies play an important role in marketing, as they can identify ineffective marketing activities of a commercial organizations and increase their overall performance by improving decision-making processes. The latest trends in the field of intelligent marketing are reviewed, including big data, digitalization, customer lifetime value, client experience, and personalization. It is substantiated that the implementation of a comprehensive data collection strategy is vital for commercial

organizations. It is proposed not to be limited to collecting only behavioral commercial data, but also to add third-party data to the analysis, including open data on organizations, clients and competitors. Additionally, the barriers that limit the potential for implementing modeling processes in marketing management are analyzed, including subjective judgments of decision makers, contradictory short-term and long-term results of implementation, as well as insufficient knowledge of modeling methodologies. It is justified that the level of model complexity is directly correlated with the level of complexity of their explanation. The priority of computational flexibility and the ability to quickly obtain and implement modeling results often dominates the advantages provided by a more feature-rich model.

For the first time, the conceptual framework for modeling client behavior in the marketing management system is developed using modern econometric models and machine learning algorithms, which, unlike existing approaches, allows identifying patterns of client behavior, both under conditions of limitations in the volume and quality of marketing data and under conditions of high-quality IT infrastructure.

The second chapter of the thesis is devoted to the development of a general approach to modeling client behavior with limited data. The author proves that poor data quality not only complicates the initial analysis, but can also lead to misinterpretations of modeling results and ineffective development of a marketing strategy. A typology of data is proposed, which includes a division into commercial and open data. Both the pros and cons of each type are presented. Machine learning is recognized as the best method for working with limited data. An approach to modeling client behavior in the case of limited data, consisting of nine stages, is developed. The author has improved the model for predicting the behavior of business clients of a consulting agency based on machine learning methods by systematizing approaches to working with low-quality data and improving modeling efficiency by implementing the process of automatic hyperparameters tuning and conducting multiple cross-validation of model quality measurement metrics. Based on the developed approach, a set of forecasting models was programmed and implemented to predict the probability of financial success of business clients. Three methods were used: classification and

regression tree, logistic regression, and random forest. The classification and regression tree was chosen as the most efficient algorithm. The developed approach is useful for marketers and consultants who can use the built models to develop an effective marketing strategy and improve the performance of business clients.

Another experimental study in Chapter 2 was developing a system of models that allowed predicting time before the next order in the online store. The author improved the model for predicting client behavior in e-commerce by using the results of client segmentation obtained by preliminary clustering of clients based on machine learning methods as model variable, which allowed for higher accuracy in predicting future purchase activity. Initially, the k-means clustering method was used to segment online store clients according to the RFM purchase activity rate. Based on the calculated values of clustering parameters, six groups of clients were identified: 1. lost clients; 2. new wholesale buyer; 3. clients we will soon lose; 4. active retail buyer; 5. new retail buyers; 6. active wholesale buyer. At the second stage, five classification algorithms were implemented: linear discriminant analysis, decision tree, support vector machine, random forest, and k-nearest neighbors to predict the time to the next order. The random forest algorithm showed the best predictive power. The results of the modeling and subsequent evaluation allowed us to conclude that the proposed modeling approach allows us to gain valuable knowledge about client behavior patterns from primary data, regardless of its size, completeness, and quality.

Chapter 3 of the thesis is devoted to an experimental study of modeling bank client behavior. The author substantiates the advantageous position of banks, compared to other areas of economic activity, due to the management of large volumes of data. The author identifies external conditions that have a significant impact on the banking sector and can change the peculiarities of cooperation between the bank and its clients, including regulatory changes, economic and market trends. The behavior of business clients with accounts in euros and retail clients with car loans is analyzed. It is found that these segments are the most profitable for the bank, but due to the specifics of their behavior, effective marketing management in relation to these client groups is a difficult task. For the first time, Cox and Weibull survival econometric models were

built to analyze the duration of client activity, which, unlike existing approaches, allow determining the factors that affect the level of client retention, under conditions of censored observations. At the first stage, separate Weibull survival models were built for each client's socio-demographic characteristic. The stepAIC function was used to find the best combination of client characteristics and levels of these characteristics to improve the quality of the models. Clients who showed the highest risk of premature closure included people over 55 years old, living in western Ukraine, and earning more than UAH 150 thousand per month. The impact of credit card sales on client survival was estimated at the second stage using the Cox proportional hazards model. According to the analysis of survival curves, clients with credit cards have a statistically longer period of activity. However, the assumption of proportional hazards for this variable was not confirmed. Thus, a survival model was developed for the factor of «active credit card». Based on the proposed approach to modeling client behavior, it was concluded that the sale of a credit card prolongs the activity of clients with a car loan. However, the critical success factor is the sale of a credit card and its activation.

The practical aspects of building a pricing strategy for the bank's business clients were further improved by developing a marketing segmentation, which, unlike existing approaches, is based on the k-means clustering algorithm. In the process of calculating the algorithm, six groups of similarities were identified, which can be characterized as: 1. unprofitable transactions with euro savings; 2. unprofitable euro savings; 3. profitable euro savings; 4. profitable transactions with euro savings; 5. customers without euro savings; 6. profitable transactions without euro savings. For each segment, a pricing strategy was developed to maintain accounts in euros. Based on the data analysis, it was found that the implementation of the developed recommendations will optimize the bank's operating expenses without setting a fee for all business clients.

Keywords: marketing, marketing activity, marketing management, client, client behavior, segmentation, client retention, strategy, decision making support, modeling, forecasting, regression, economic and mathematical modeling, econometric models, machine learning, classification, clustering.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Праці, в яких опубліковані основні наукові результати дисертації

Статті у наукових фахових виданнях України та зарубіжних виданнях, проіндексованих у базах даних Scopus та Web of Science

1. Піскунова, О., & Ключко, Р. (2020). Прогнозування часу до наступного замовлення клієнта інтернет-магазину на підґрунті методів машинного навчання. *Економічний простір*, 161, 147–152. <https://doi.org/10.32782/2224-6282/161-27> (0.7 д.а. , з них 0.6 д.а. авторські: автором розроблено та реалізовано прогнозування часу до наступного замовлення клієнтів за допомогою методів машинного навчання).

2. Piskunova, O., Ligonenko, L., Klochko, R., & Frolova, T., Bilyk, T. (2021). Applying machine learning approach to start-up success prediction. *Scientific Horizons*, 24(11), 72–84. [https://doi.org/10.48077/scihor.24\(11\).2021.72-84](https://doi.org/10.48077/scihor.24(11).2021.72-84) (1.2 д.а., з них 0.9 д.а. авторські: автором реалізовано попередній процес обробки даних, зроблена побудова та аналіз моделі, що ґрунтується на використанні методів машинного навчання).

3. Klochko, R., & Piskunova, O. (2023). Marketing segmentation of Banks' corporate clients based on data mining technique. *Statistics in Transition New Series*, 24(5), 147–166. <https://doi.org/10.59170/stattrans-2023-068> (1.2 д.а., з них 1.1 авторські: автором реалізовано модель кластеризації клієнтів банку та розроблено цінові стратегії для кожного із виявлених сегментів).

4. Piskunova, O., Klochko, R., Bilyk, T., & Frolova, T. (2023). Modelling the bank customer activity duration based on the Cox Econometric Survival Model. *Financial and Credit Activity Problems of Theory and Practice*, 6(53), 21–31. <https://doi.org/10.55643/fcaptp.6.53.2023.4244> (0.9 д.а., з них 0.8 авторські: автором зроблена побудова та аналіз моделей активності роздрібних клієнтів банку, що ґрунтуються на використанні методу регресії виживання Кокса).

Розділ у колективній монографії

5. Піскунова, О., & Клочко, Р. (2023) Моделювання тривалості клієнтської активності в банку з використанням регресійної моделі виживання Вейбулла. Актуальні проблеми системного аналізу та моделювання процесів управління. За ред. В. Пономаренка, Л. Гур'янової, Я. Пеліової, Е. Ніжинського – Братислава-Харків, ВШЕМ – ХНЕУ ім. С. Кузнеця, 51–68. (0.85 д.а., з них 0.75 авторські: автором зроблена побудова та аналіз моделей активності роздрібних клієнтів банку, що ґрунтуються на використанні методу регресії виживання Вейбулла)

Праці, які засвідчують апробацію матеріалів дисертації

6. Piskunova, O., & Klochko, R. (2020). Classification of e-commerce customers based on Data Science techniques. *CEUR Workshop Proceedings*, 2649, 6 – 20 (0.8 д.а., з них 0.7 авторські: автором зроблена побудова та порівняльний аналіз моделей кластеризації та класифікації клієнтів онлайн-магазину).

7. Піскунова О., & Клочко, Р. (2020). Нова парадигма вітчизняного маркетингу на підґрунті Big Data статистики. Нові джерела та методи поширення даних у статистиці: матеріали XVIII Міжнародної науково-практичної конференції з нагоди Дня працівників статистики, Київ, «Інформаційно-аналітичне агентство», 37–41 (0.25 д.а, з них авторські 0.2: автором зроблено аналіз ключових напрямків розвитку маркетингу в епоху Big Data статистики).

8. Піскунова О., & Клочко, Р. (2020). Кластеризація клієнтів інтернет-магазину за рівнем їхньої купівельної активності. Матеріали XII міжнародної науково-практичної конференції «Сучасні проблеми моделювання соціально-економічних систем» (09-10 квітня 2020 р.). Братислава – Харків, ВШЕМ – ХНЕУ ім. С. Кузнеця (0.15 д.а., з них 0.1 авторські: автором запропоновано підхід до моделювання рівня купівельної активності клієнтів інтернет-магазину).

ЗМІСТ

ВСТУП.....	16
РОЗДІЛ 1 ТЕОРЕТИЧНІ ЗАСАДИ МОДЕЛЮВАННЯ ПОВЕДІНКИ КЛІЄНТІВ У СИСТЕМІ МАРКЕТИНГОВОГО МЕНЕДЖМЕНТУ	24
1.1 Сутність та особливості розвитку маркетингового менеджменту	24
1.2 Роль розуміння поведінки клієнтів у маркетинговому менеджменті.....	41
1.3 Аналітичний огляд існуючих підходів щодо моделювання поведінки клієнтів	63
1.4 Концептуальні положення щодо моделювання поведінки клієнтів у системі маркетингового менеджменту	88
Висновки до розділу 1	95
РОЗДІЛ 2 МОДЕЛЮВАННЯ ПОВЕДІНКИ КЛІЄНТІВ В УМОВАХ ОБМЕЖЕНОСТІ ДАНИХ	98
2.1 Особливості моделювання поведінки клієнтів в умовах обмеженості даних з використанням методів машинного навчання	98
2.2 Застосування інтелектуальних технологій до моделювання поведінки бізнес- клієнтів консалтингової агенції	122
2.3 Побудова системи моделей поведінки клієнтів в електронній комерції на підґрунті методів машинного навчання.....	133
Висновки до розділу 2	144
РОЗДІЛ 3 УДОСКОНАЛЕННЯ МАРКЕТИНГОВОГО МЕНЕДЖМЕНТУ В БАНКУ НА ПІДҐРУНТІ МОДЕЛЮВАННЯ ПОВЕДІНКИ КЛІЄНТІВ	147
3.1 Особливості моделювання поведінки клієнтів за наявності якісної інфраструктури даних на прикладі клієнтів банку.	147
3.2 Моделювання тривалості клієнтської активності з використанням економетричних моделей виживання	162
3.3 Удосконалення маркетингового менеджменту в банку шляхом побудови маркетингової сегментації бізнес-клієнтів	178
Висновки до розділу 3	187
ВИСНОВКИ.....	190
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	195
ДОДАТКИ.....	214

ВСТУП

Актуальність теми. Маркетингову діяльність зазвичай асоціюють із креативною складовою, в якій чи не єдиним критерієм прийняття рішень завжди була особиста думка топ-менеджменту комерційних організацій. Але в передових економіках маркетинг поступово перетворюється більше на технологічний процес, в якому рішення приймаються на основі алгоритмів моделювання. Бізнес більше не може покладатися лише на інтуїцію чи попередній досвід для визначення ефективних практик маркетингового менеджменту. Інноваційний розвиток комерційних організацій стає основним фактором їх конкурентоспроможності.

У нинішньому конкурентному бізнес-середовищі розуміння поведінки клієнтів має вирішальне значення для комерційних організацій, які прагнуть розробляти ефективні маркетингові стратегії. Сьогодні ефективність масового маркетингу постійно знижується, а у клієнтів спостерігається прагнення до індивідуалізованого досвіду, у зв'язку з розширенням спектру вибору та зниженням лояльності до конкретних брендів. Клієнти, які мають безліч альтернатив, все частіше демонструють примхливу і нестабільну поведінку.

Щоб відповідати вимогам сучасної економіки, комерційним організаціям необхідно при кожному прийнятті рішення зважати на думку своїх клієнтів. Можливо досягти справжнього успіху і встановити міцні зв'язки з клієнтами, якщо надавати пріоритет розумінню їх поведінки і персоналізації їхнього досвіду. Водночас, за щоденного контакту з тисячами клієнтів це важко здійснити. Значення економіко-математичного моделювання, з цієї точки зору, неможливо перебільшити, оскільки воно дозволяє перетворювати наявні первинні дані про клієнтів на знання про закономірності їх поведінки, що надає практичні переваги при розробці маркетингових стратегій. Алгоритми моделювання дають можливість сформувати детальне уявлення про фактори, що впливають на лояльність та утримання клієнтів, чого неможливо досягти за допомогою традиційного аналізу. Використовуючи методи аналізу виживання,

методи машинного навчання, включаючи алгоритми кластеризації та класифікації, можна побудувати системи обслуговування клієнтів з урахуванням їх уподобань та особливості поведінки. Це дозволить точно таргетувати маркетингові повідомлення, створювати і покращувати продукти та послуги, розробляти стратегії їх продажу. Поведінкове моделювання дозволяє комерційним організаціям швидко приймати рішення на основі даних, що є критично важливим у постійній боротьбі за лояльність клієнтів, підвищення ефективності маркетингової діяльності та збільшення продажів.

Тим не менш, підходи до моделювання також повинні враховувати практичні обмеження, що накладаються реальним бізнес-середовищем, враховуючи, що деякі комерційні організації мають обмеження в плані ресурсів та інфраструктури даних, що перешкоджає впровадженню підходів щодо моделювання поведінки клієнтів у системі маркетингового менеджменту.

Питання особливостей функціонування та розвитку маркетингового менеджменту та поведінки клієнтів досліджували такі вчені як: Шоу Е., Кук Д., Котлер Ф., Бейкер М., Лівінгстон Д., Фостер Д., Вотерворт Д., Шпонгольц У., Віренга Б., Ланкастер Д., Рейнольдс П., Бікгофф Н., Холленсен С., Опресник М., Каваллоне М., Дейтон Д., Мела К., Мурман К., Шварцкопф С., Тамілія Р., Ретчфорд Б., Бернстайн Д., Дюрінг Л., Меррітт Р., Камін Т., Ларіна Я., Савич О., Федорченко А., Архипова Т., Овсієнко Н., Павленко В., Пономаренко І., Моргулець О., Чорна О., Пилипенко В., Шкода Т., Олексюк О., Шафалюк О., Кофанов О., Зозульов О., Пилипчук В., Голіцин А., Гвоздицький В. та інші.

Суттєвий вплив на розвиток моделей та методів економіко-математичного моделювання в маркетинговому менеджменті зробили праці таких іноземних та вітчизняних вчених як: Віренга Б., Блаттберг Р., Кім Б., Неслін С., Бругген Г., Брей В., Ван дер Ланс Р., Міклошик А., Кухта М., Еванс Н., Райт Л., Стоун М., Моутінью Л., Багоцці Р., Го Ч., Гаррелл Ф., Вудуріс Х., Овусу Г., Дорн Р., Лесен Д., Джудічі П., Менвілль В., Чорноус Г., Негрей М., Агутін М., Обруч, Г., Піскунова О., Дербенцев В., Матвійчук А., Великоіваненко Г., Юнькова О., Даценко Н., Бабенко В., Хрустальов К., Ляшенко О., Кравець Т., Прокопенко М.,

Гур'янова Л., Яценко Р., Дубровіна Н., Чаговець Л., Чаговець В., Чернова Н. та інші.

Незважаючи на те, що дослідники широко вивчали методологічні, теоретичні та практичні елементи моделювання поведінки клієнтів у системах маркетингового менеджменту, все ще залишаються важливі сфери дослідження, які потребують вирішення для практичного застосування. Досі бракує досліджень, які зосереджені на комплексному вивченні особливостей моделювання поведінки клієнтів як за умови обмежень в обсягах і якості маркетингових даних, так і за умови якісної ІТ інфраструктури даних. Широка сфера застосування цього дослідження є вирішальним фактором, що визначає його значущість та актуальність. Відсутність комплексної точки зору, яка б забезпечила значні можливості для практиків при розробці маркетингових стратегій, незалежно від ступеня їхньої аналітичної зрілості, зумовили вибір теми, мети, завдання, об'єкта і предмета цього дисертаційного дослідження.

Зв'язок з науковими програмами, планами, темами. Дисертаційне дослідження виконано відповідно до плану науково-дослідних робіт Київського національного економічного університету імені Вадима Гетьмана та безпосередньо пов'язане з такими темами:

— науково-дослідна тема «Розвиток методології та інструментарію моделювання економічних систем у контексті підвищення економічної безпеки» (державний реєстраційний номер 0116U001428), в межах якої проведено дослідження сучасного стану економіко-математичного моделювання у сфері маркетингового менеджменту, спрямоване на покращення теоретичних і практичних засад. Здобувачем запропоновано підходи та моделі, необхідні для підвищення ефективності прийняття маркетингових рішень в умовах обмеженості даних.

— науково-дослідна тема «Статистичний аналіз та математичне моделювання економічних процесів в умовах глобальних зрушень в економіці» (державний реєстраційний номер 0123U101526). Участь автора полягає в удосконаленні маркетингового менеджменту банків шляхом авторської

розробки економетричних моделей виживання, що дозволяють покращити ефективність процесів утримання клієнтів.

Мета і завдання дослідження. Метою наукового дослідження є розробка концептуальних положень щодо моделювання поведінки клієнтів та побудова комплексу відповідних економіко-математичних моделей для вдосконалення системи маркетингового менеджменту.

Досягнення мети зумовило постановку та вирішення таких завдань:

- дослідити понятійний апарат, особливості та сучасні тенденції розвитку маркетингового менеджменту на основі аналізу праць вітчизняних та зарубіжних вчених;
- проаналізувати економіко-математичний інструментарій та існуючі підходи щодо моделювання поведінки клієнтів;
- сформулювати концептуальні положення щодо моделювання поведінки клієнтів у системі маркетингового менеджменту;
- розробити комплекс економіко-математичних моделей для вирішення задачі моделювання поведінки клієнтів за різної якості доступних даних;
- здійснити реалізацію побудованих моделей для прогнозування поведінки бізнес-клієнтів консалтингової агенції в умовах обмеженості даних;
- проаналізувати поведінку клієнтів в електронній комерції на підґрунті розробленої системи моделей;
- провести на основі методів економіко-математичного моделювання емпіричне дослідження факторів, що впливають на тривалість клієнтської активності в банку;
- сформулювати рекомендації щодо вдосконалення маркетингового менеджменту в банку шляхом застосування технологій моделювання поведінки клієнтів.

Об’єкт дослідження. Поведінка клієнтів з точки зору маркетингового менеджменту.

Предмет дослідження. Економетричні моделі та методи машинного навчання, що використовуються для моделювання поведінки клієнтів.

Методи дослідження. У дисертаційному дослідженні було використано статистичні методи, методи економіко-математичного моделювання, а саме: регресії виживання Вейбулла та Кокса, алгоритми машинного навчання, а саме: метод кластеризації k-середніх та методи класифікації (випадковий ліс, k-найближчих сусідів, лінійний дискримінантний аналіз, логістична регресія, метод опорних векторів, дерево класифікації та регресії). Також було застосовано загальнонаукові методи, а саме: методи синтезу та аналізу, графічні методи. Інформаційною базою дослідження є відкриті дані стосовно діяльності українських підприємств з ресурсу Dealroom, відкриті дані стосовно діяльності інтернет магазину з ресурсу UCI Machine Learning Repository, комерційні дані українського банку.

Наукова новизна отриманих результатів полягає в розробці концептуальних положень та відповідного комплексу економіко-математичних моделей для моделювання поведінки клієнтів за різної якості доступних даних з використанням алгоритмів машинного навчання та економетричних моделей виживання, що надає нові можливості щодо підвищення ефективності прийняття рішень в системі маркетингового менеджменту.

У процесі наукового дослідження отримані такі результати:
вперше:

- розроблено концептуальні положення щодо моделювання поведінки клієнтів у системі маркетингового менеджменту з використанням економетричних моделей та алгоритмів машинного навчання, що, на відміну від існуючих підходів, дозволяє виявляти закономірності клієнтської поведінки, як за умови обмежень в обсягах і якості маркетингових даних, так і за умови якісної ІТ інфраструктури даних;
- побудовано економетричні моделі виживання Кокса та Вейбулла для аналізу тривалості клієнтської активності, які, на відміну від існуючих підходів, дозволяють визначати фактори, що впливають на рівень утримання клієнтів, включаючи як соціально-демографічні характеристики клієнтів, так і активності з продажу додаткових продуктів клієнтам, за умови цензурованих спостережень.

удосконалено:

- модель прогнозування поведінки бізнес-клієнтів консалтингової агенції на підґрунті методів машинного навчання, для побудови якої було застосовано комплекс підходів з підвищення якості маркетингових даних, зокрема: усунення класового дисбалансу, бінаризація даних, заповнення пропущених даних тощо; а також комплекс підходів з підвищення якості побудованих моделей шляхом реалізації процесу автоматичного налаштування гіперпараметрів та проведення множинних перехресних перевірок метрик оцінки якості моделей, що дозволило отримати адекватні результати прогнозування за умови неякісних даних та обмеженої кількості спостережень;

- модель прогнозування поведінки клієнтів в електронній комерції за рахунок використання в якості фактору моделі результатів сегментації клієнтів, що були отримані шляхом попередньої кластеризації клієнтів на підґрунті методів машинного навчання, що дозволило більш точно спрогнозувати майбутню купівельну активність.

дістали подальшого розвитку:

- практичні аспекти побудови цінової стратегії для бізнес-клієнтів банку за допомогою розробки маркетингової сегментації, яка, на відміну від існуючих підходів, базується на алгоритмах кластеризації;

- рекомендації щодо застосування економетричних моделей та методів машинного навчання при аналізі особливостей поведінки клієнтів, які можуть бути використаними для розробки стратегії маркетингового менеджменту та підвищення ефективності прийняття маркетингових рішень.

Практичне значення одержаних результатів. Запропоновані автором концептуальні положення щодо моделювання поведінки клієнтів можуть бути впроваджені у бізнес-процеси комерційних організацій для удосконалення маркетингового менеджменту, зокрема для: підвищення ефективності процесу прийняття маркетингових рішень, формування дієвої маркетингової стратегії, оптимізації операційних витрат та маркетингових інвестицій, підвищення рівнів

утримання та лояльності клієнтів, реалізації результативної комунікаційної стратегії, розробки та удосконалення продуктів і послуг.

Підхід і набір моделей, методів та алгоритмів, застосований в роботі, в першу чергу може бути корисними для банків, інтернет-магазинів, маркетингових та консалтингових агенцій. Комплексний характер цього дисертаційного дослідження дозволяє використання розробленого концептуального підходу в процесі удосконалення системи маркетингового менеджменту українських чи закордонних комерційних організацій різних видів економічної діяльності, незалежно від ступеня їхньої аналітичної зрілості.

Отримані практичні результати дисертаційного дослідження підтверджені відповідними довідками про апробацію. Одержані наукові результати впроваджено в роботу:

1. ТОВ «АДАПТИВ МАРКЕТИНГ ІННОВЕЙШЕНЗ» (довідка № 1411-23 від 14 листопада 2023 року).

2. ТОВ «РАЗОМ КОМ'ЮНІКЕЙШНЗ» (довідка № 18-09/12 від 18 вересня 2023 року).

3. LEOPOLIS S.L.U (довідка № 3-6-11-23 від 6 листопада 2023 року).

4. Довідка від Київського національного економічного університету імені Вадима Гетьмана про використання результатів дисертаційної роботи при викладанні навчальних дисциплін «Кількісні методи в економіці (в R)» та «Економетричне моделювання в IT-середовищі» (довідка від 11 грудня 2023 року).

Особистий внесок здобувача. Дисертаційне дослідження є самостійно виконаною науковою роботою автора. Усі результати, висновки, розробки, рекомендації та інші результати дослідження одержані автором самостійно. Із наукових праць, які видані у співавторстві, використовувалися лише особисті напрацювання автора.

Апробація результатів дисертації. Основні теоретичні положення та практичні результати дисертації доповідались і обговорювались на 3 міжнародних та всеукраїнських науково-практичних конференціях:

- XII Міжнародній науково-практичній конференції «Сучасні проблеми моделювання соціально-економічних систем» (Харків 9-10 квітня 2020 р.);
- XVIII Міжнародній науково-практичній конференції з нагоди Дня працівників статистики. «Інформаційно-аналітичне агентство» (Київ, грудень 2020 р.);
- 27th Polish-Slovak-Ukrainian Scientific Seminar «Statistical Methods in Modern Economy – Theory and Applications» (Krakow, Poland, 19th October 2021).

Публікації за темою дисертації. Основні результати дисертаційного дослідження опубліковано у співавторстві у 8 наукових працях, загальним обсягом 6.05 д.а., (з яких особисто автору належать 5.15), з них: 4 – статті у наукових фахових виданнях України та зарубіжних виданнях, проіндексованих у базах даних Scopus та Web of Science; 1 – розділ у колективній монографії; 3 – тез доповідей на науково-практичних конференціях.

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, трьох розділів, висновків, списку використаних джерел, додатків. Обсяг основного тексту складає 177 сторінок друкованого тексту. Робота містить 35 таблиць на 7 сторінках та 42 рисунки на 15 сторінках. Список використаних джерел налічує 174 найменувань. Дисертація містить 7 додатків на 22 сторінках.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ЗАСАДИ МОДЕЛЮВАННЯ ПОВЕДІНКИ КЛІЄНТІВ У СИСТЕМІ МАРКЕТИНГОВОГО МЕНЕДЖМЕНТУ

1.1 Сутність та особливості розвитку маркетингового менеджменту

Розвиток цифрової ери, глобалізація ринку та нові моделі поведінки клієнтів суттєво трансформували сферу управління маркетингом. Маркетинговий менеджмент, що є невід'ємною складовою сучасних бізнес-стратегій, сьогодні перетворився на динамічну і складну галузь знань. Значення змісту та особливостей розвитку маркетингового менеджменту зростає з кожним днем. За своєю суттю, маркетинговий менеджмент пов'язаний з розробкою, формуванням та реалізацією підходів, спрямованих на створення, передачу та постачання цінності для клієнтів. Становлення маркетингового менеджменту, як наукової галузі відіграло ключову роль у стимулюванні ринкової конкуренції, просуванні інновацій та впливі на ринкову динаміку.

Маркетинг – це систематичний процес, що включає реалізацію різних елементів, таких як дизайн продукту, цінові стратегії, рекламні активності та дистрибуцію. Його основна мета сприяти укладанню взаємовигідних угод, які відповідають цілям, як окремих осіб, так і організацій. Маркетинг – це багатогранне соціальне явище, яке охоплює кілька вимірів, таких як економічні, соціальні, інтелектуальні, технічні та правові аспекти [136].

Термін «маркетинг» часто неправильно розуміють, що призводить до плутанини серед тих, хто помилково асоціює його з продажами, або звужують значення маркетингу лише до рекламної активності. Маркетинг охоплює стратегії і тактики, які організація використовує для взаємодії з клієнтами на широкому рівні, тоді як продажі стосуються конкретних взаємодій між комерційною організацією та окремими клієнтами [18].

Надалі будемо використовувати термін «комерційна організація» як основний, для характеристики юридичної особи, яка бере участь у комерційній

діяльності шляхом продажу товарів і послуг з метою отримання фінансового прибутку.

У своїй статті «Розширення концепції маркетингу» Філіп Котлер і Сідні Леві стверджують, що маркетинг – це всеохоплююча суспільна діяльність, яка виходить за рамки простого комерційного обміну товарами. Науковці стверджують, що суб'єкти поза межами бізнесу, такі як політики, освітні заклади чи благодійні організації, також займаються маркетинговою діяльністю, а принципи маркетингу, також можуть бути ефективно використані у просуванні послуг, осіб та концепцій [79].

Витоки теорії маркетингу можна простежити ще з античних часів, зокрема, з часів Аристотеля, Платона та Гомера. Вчені обговорювали значення продуктів та послуг з точки зору їх цінності, а також вирішальну роль, яку відіграють ефективна комунікація та дистрибуція у досягненні успіху в сфері торгівлі. Ці теми не втрачають своєї актуальності в сучасному маркетингу. В епоху Відродження відбулася формалізація деяких ринкових практик і методологій, що охоплюють поведінку споживачів, сегментацію ринку, канали розподілу та маркетингові дослідження [146]. Тим не менш, наукове вивчення маркетингу з'явилося невдовзі після початку 20-го століття [134].

Ерік Шоу, у своєму дослідженні «Роздуми про історію маркетингової думки», стверджує, що важливо розуміти внесок історичних дослідників маркетингу для того, щоб побудувати нові теоретичні рамки. Автор обговорює кількох осіб, які відіграли значну роль у розвитку маркетингової думки на початку 20-го століття. До таких першопрохідців належать Ральф Батлер, Арч Шоу, Пол Черінгтон, Фред Кларк, Пол Конверс та Ральф Брейєр [133]. Всебічний огляд еволюції системи маркетингу, протягом двадцятого століття, також провели Роджер Л. [87] та Шварцкопф, С. [129], висвітлюючи появу багатьох наукових шкіл у міру розвитку дисципліни.

Зародження промислової революції започаткувала собою важливий етап у розвитку ідеології маркетингу. Саме в цей період відбулося значне підвищення ефективності виробництва, що супроводжувалося зростанням потреби у

створенні та комерціалізації більшої кількості вироблених товарів. Як наслідок, з'явилися нові маркетингові методи та підходи, зокрема реклама та брендинг [146].

На початку свого існування маркетинг сприймався, як функція розподілу, з основним акцентом на фізичну дистрибуцію [20]. У цей період серед численних наукових шкіл у сфері маркетингу виділяють школу маркетингових функцій, школу маркетингових товарів та школу маркетингових інституцій [123].

Основна концепція **школи маркетингових функцій** полягає в тому, що маркетинг – це сукупність дій, які уможливають обмін товарами та послугами між клієнтами та постачальниками. Ця школа зосереджується на таких маркетингових функціях, як продаж, доставка, зберігання, стандартизація та оцінювання, фінансування та прийняття ризиків [134].

З 1920 по 1930 роки відбулося помітне розширення концепції маркетингу, яка охоплювала не лише просування товарів, але й різні аспекти дистрибуції та супутніх послуг, що сприяють переміщенню товарів від виробника до споживача. Термін «маркетинг» було сформульовано для того, щоб охопити складний комплекс послуг, пов'язаних з розподілом товарів, за винятком діяльності, яка передбачає зміну фізичних характеристик продукту [32].

У період з 1930 по 1940 роки під поняттям «маркетинг» розуміли, насамперед, процес просування товарів шляхом розповсюдження переконливої інформації, яка б формувала попит на них, і розвивала готовність платити встановлену ціну та докладати необхідних зусиль для отримання товару.

З 1940 по 1950 роки відбулося помітне розширення сфери застосування маркетингу, включаючи не лише фізичні аспекти дистрибуції, але й поширення його впливу на сферу виробництва продуктів [20].

Товарна **школа маркетингових думок** класифікувала товари та послуги, як об'єкти, що є предметом ринкового обміну. Основна концепцією було те, що маркетинг полягає у створенні вартості, шляхом виробництва та розповсюдження товарів і послуг [134].

Протягом 1940-1950-х років маркетинг пройшов процес інституціоналізації, що призвело до його сприйняття, як більш усталеної та практичної науки. Ідея маркетинг-міксу була створена Нілом Борденом у 1953 році [136].

Інституційна школа зосередилася на характеристиці та поведінці маркетингових посередників, які діють між виробниками та кінцевими клієнтами всередині каналів розподілу, включаючи оптових і роздрібних торговців. Основна концепція полягала в тому, що маркетинг є багатогранною соціальною та економічною структурою, що включає складну мережу інституцій та організацій [134].

Маркетинговий мікс – це набір стратегічних інструментів, що використовуються комерційними організаціями для ефективної реклами та просування своїх пропозицій у конкурентному ринковому середовищі. Система складається з декількох взаємопов'язаних компонентів, які гармонійно співпрацюють для досягнення запланованих результатів. Складові маркетингового міксу добре відомі як «чотири Р»: Product (продукт), Price (ціна), Promotion (просування) і Place (місце). Використання маркетингового міксу дозволяє комерційним організаціям створити особливу ідентичність за допомогою свого бренду та персоналізованих повідомлень, тим самим сприяючи диференціації від організацій-конкурентів [49].

Загалом, маркетинг-мікс є важливим інструментом у формуванні ефективних маркетингових стратегій, які відповідають вимогам їхньої цільової аудиторії. Завдяки ретельному управлінню його елементами, комерційні організації можуть ефективно залучати та утримувати клієнтську базу, успішно досягти своїх маркетингових цілей [7].

У період з 1950 по 1960 роки концепція маркетингу характеризувалася всеосяжністю, охоплюючи всі комерційні активності, пов'язані з рухом товарів і послуг від їх фізичного створення до кінцевого споживання [20]. В цей час відбувся помітний перехід до ринкової орієнтації у маркетингових стратегіях. Цей зсув спричинив за собою фундаментальну зміну в підходах маркетингового

планування комерційними організаціями, оскільки вони почали активно шукати способи зрозуміти запити клієнтів, а згодом адаптувати свої товари та послуги відповідно до цих виявлених потреб [22].

Поява комп'ютерів, на основі інтегральних схем, започаткувала трансформаційний етап у сфері маркетингових досліджень. Як наслідок, відбулося помітне зростання обчислювальних потужностей, хоча й дещо обмежене у світлі сучасних стандартів. У цей час основний акцент досліджень був спрямований на сферу маркетингового менеджменту та кількісних методів. За допомогою математичних моделей дослідники проводили аналіз поведінки клієнтів та розробляли ефективні маркетингові стратегії. У цей період з'явилася низка нових методів, таких як лінійне програмування, дерева рішень та теорія ігор. Вищезгадані методи використовувалися для розробки моделей поведінки клієнтів, оптимізації рішень, щодо маркетинг-міксу та аналізу впливу реклами на продажі [122].

До 1970 року відбулося широке розуміння фундаментальних аспектів філософії **маркетингового менеджменту**, що охоплювали елементи аналізу, планування, організації та регулювання, які розглядалися, як найважливіші складові функціонального менеджменту [10]. У цей час відбувся помітний перехід від маркетингового менеджменту до вивчення поведінки клієнтів. Науковці почали досліджувати психологічні та соціальні чинники, які впливають на поведінку клієнтів. З'явилася низка нових методологій, таких як оцінка ставлення, мапування переваг, конджойнт-аналіз, а також інші поведінкові та кількісні вимірювальні підходи. Вищезгадані методології використовувалися з метою сегментації ринку, визначення уподобань клієнтів та розробки стратегій позиціонування [122]. Початок цієї трансформації можна простежити в багатьох наукових працях, зокрема у впливових роботах Майкла Бейкера [7], Філіппа Котлера [79] та Джеймса М. Лівінгстона [93]. Зміна парадигми мала значний вплив на більшість сучасних наукових шкіл.

Школа маркетингового менеджменту приділяє значну увагу основним аспектам маркетингу, включаючи планування, реалізацію та контроль.

Центральна концепція стверджує, що маркетинг – це управлінська діяльність, яка охоплює систематичне вивчення, стратегічне формування, оперативне виконання та регуляторний нагляд за маркетинговими ініціативами. Науковці розглядають маркетинг, як систематичну процедуру створення клієнтської цінності та окреслює ключові управлінські обов'язки, включаючи розробку маркетингового міксу.

Школа маркетингових систем розглядає маркетинг, як комплексну структуру, що складається з декількох взаємопов'язаних елементів, таких як постачальники, виробники, дистриб'ютори, роздрібні торговці та клієнти. Центральна концепція обертається навколо складної природи маркетингу, яка вимагає гармонізації різноманітних видів діяльності та процесів. Крім того, ця школа приділяла значну увагу різноманітності каналів, що використовувалися для розповсюдження товарів і послуг, а також численним способам комунікації, що використовувалися для їх реклами та просування.

Школа поведінки споживачів робить акцент на вивченні поведінки споживачів, включаючи такі аспекти, як їхні бажання, вимоги, схильності та когнітивні процеси, що беруть участь у процесі прийняття рішень [134].

З розвитком маркетингу, науковці почали приділяти все більше уваги вивченню позиції маркетингу в суспільстві та його соціальних зобов'язань. У 1984 році Гордон Фоксалл опублікував роботу під назвою «Сфера маркетингу», в якій представив критичний погляд на зростаючу парадигму обміну в сфері маркетингу. У статті було критично проаналізовано розширення теорії маркетингу за межі її традиційного використання у сфері бізнесу, охоплюючи такі сфери, як «соціальний маркетинг» та «небізнесовий маркетинг» [50].

Соціальний маркетинг – це застосування маркетингових ідей і стратегій з метою впливу на поведінку людей та її модифікації, для сприяння суспільному добробуту. Ця практика передбачає використання тактик комерційного маркетингу для покращення здоров'я населення, збереження довкілля, соціальної справедливості та деяких інших питань. Основна мета соціального маркетингу –

сприяти поведінці, яка може підвищити добробут окремих людей і суспільства в цілому [99].

Іншим визначенням соціальний маркетингу є те що – це стратегічний підхід, який використовує маркетингові ідеї і тактики для полегшення і просування добровільних змін у поведінці, з кінцевою метою підвищення суспільного добробуту [25].

У період з 1980-х до 1990-х років відбувся значний перехід, який характеризувався переважним акцентом на стратегічному маркетингу та досягненні клієнтського щастя. Маркетологи помітили, що щастя клієнта може призвести до росту рівня утримання, залучення нових клієнтів та встановлення довготривалих відносин з ними, через створення цінності. Ці часи стали помітною епохою трансформаційних розробок у сфері маркетингових досліджень. Сфера кількісних досліджень, що спирається на економіку та статистику, стала унікальною галуззю вивчення [93].

У своїй праці "Маркетинг як інновація: Восьма парадигма" Кеннет Сіммондс стверджує, що концепція інновацій має важливе значення у сфері маркетингу, незважаючи на її помітну відсутність у традиційній теорії маркетингу. На думку Сіммондса, ступінь інновацій залежить від організаційного впливу маркетингу, і саме на маркетологів лягає тягар розвитку інновацій всередині комерційної організації [141].

З появою інтернету та широкого розповсюдження персональних комп'ютерів відбулася помітна трансформація у сфері маркетингових досліджень. Науковці почали досліджувати вплив інтернету на поведінку клієнтів і намагалися побудувати основи для розуміння їх поведінки в мережі. У цей період відбувся помітний прогрес у сфері імітаційних моделей прийняття рішень, що дозволило дослідникам моделювати та аналізувати поведінку клієнтів у контрольованому середовищі. Важливість системи управління взаємовідносинами з клієнтами (CRM) була визнана в академічних дослідженнях, оскільки бізнеси все частіше надавали пріоритет встановленню довготривалих зв'язків з клієнтами.

У цей час першочерговою метою стало покращення відносин з клієнтами шляхом використання цифрових платформ для налагодження тісніших зв'язків, оцінки їхньої поведінки та створення значущого досвіду. Комерційні організації перейшли від транзакційного підходу, до прийняття стратегії домінування послуг у сфері маркетингу. Ця нова стратегія робить акцент на використанні технологій для надання клієнтам індивідуального та цікавого досвіду в кожній точці контакту [22].

Багато науковців-маркетологів вивчали цей перехід у маркетинговій стратегії та поведінці клієнтів в епоху цифрових технологій, включаючи: Барон С. та Харріс К. [9], Віренга Б. та Бріджес Г. [158], Голдсміт Р. [56], Ель-Ансарі А. [46], Бейкер М. [8], Мендонса П. та Тейшейра С. [98].

Перше десятиліття 21-го століття стало свідком постійного розширення наукових досліджень, пов'язаних з інтернетом, коли вчені наполегливо вивчали вплив цього технологічного середовища на поведінку клієнтів. Концепція маркетингової прибутковості привернула значну увагу в академічних дослідженнях, оскільки організації змістили свій фокус на оптимізацію рентабельності інвестицій. Значний сплеск онлайн досліджень призвів до того, що інтернет став основним місцем для проведення досліджень у сфері маркетингу [122].

Загалом, діджитал трансформація економічних процесів залишила вагомий внесок у маркетингових наукових дослідженнях, зокрема варто відмітити праці Віейра В. [151], Мартінеса Л. [96], Вагнер К. [152], та Бругген Г. та Віренга Б. [149].

Наразі функція маркетингу полягає у виявленні та задоволенні потреб, бажань та очікувань клієнтів з метою створення цінності та, як наслідок, підвищення прибутковості комерційної організації. Тим не менш, існує невідповідність між твердженнями, викладеними в науковій літературі, та практичною реалізацією в організаціях, що призводить до розбіжності між описаними цілями маркетингу, і його реальною функцією в бізнесі. Тому вкрай важливо, щоб маркетинг відігравав помітну роль всередині організації, слугуючи

керівною силою, яка активно сприяє розвитку ринку, виявляючи розбіжності між теоретичними знаннями і практичною реалізацією [22].

Останніми роками спостерігається зростаюча тенденція до використання даних у маркетинговій науці. Стрімкий технологічний прогрес полегшив збір та аналіз величезної кількості даних, що стосуються поведінки клієнтів [123]. Ідея маркетингу зараз перебуває на стадії подальшого розвитку, що характеризується зростаючою увагою до персоналізованих маркетингових стратегій та використанням аналізу на основі даних для отримання уявлення про потреби та поведінку клієнтів.

Теоретико-методологічні та практичні аспекти впровадження інтелектуального аналізу даних в маркетинговому менеджменті висвітлено у працях таких учених, як Ларіна Я. [53], Павленко В., Пономаренко І., Моргулець О. та Федорченко А. [115], Савич О. та Шкода Т. [127], Олексюк О. та Шафалюк О. [168], Пилипчук В. та Голіцин А. [173], Ляшенко О., Кравець Т. та Прокопенко М. [91], Агутін М. [165].

Автори книги «Маркетинг Н2Н» Котлер Ф., Пфферш В. та Шпонгольц У. стверджують, що маркетинг слід розглядати, як механізм створення людської цінності, а не лише, як спосіб продажу товарів та заробляння грошей. Автори пропонують поняття маркетингу Н2Н (human-to-human) або людина-до-людини, як основу для створення нового маркетингового підходу, який надає пріоритет людському фактору, як у процесі планування, так і в процесі реалізації [80].

Виділяють такі фундаментальні елементи, які лежать в основі сучасної маркетингової концепції:

1. Прийняття клієнтоорієнтованого підходу є ключовою стратегією в сучасній бізнес-практиці.

2. Поняття інтегрованого маркетингу підкреслює необхідність прийняття цілісної стратегії маркетингу, в якій різні маркетингові операції, включаючи розробку продукту, рекламу, просування, продажі та PR, гармонійно координуються для досягнення запланованих результатів [44].

3. Постійні дослідження клієнтів є фундаментальним елементом, що лежить в основі філософії маркетингу. Для бізнесу важливо брати участь у постійних дослідженнях клієнтів, щоб мати всебічне розуміння їхніх потреб, смаків та ринкових тенденцій.

4. Встановлення відносин з клієнтами є важливим аспектом маркетингової ідеї, згідно з якою комерційні організації зобов'язані розвивати тривалі зв'язки з ними, шляхом надання високоякісних продуктів і послуг.

5. Маркетингова парадигма вимагає, щоб комерційні організації були в курсі новітніх технічних розробок, оскільки вони можуть покращити розуміння поведінки клієнтів, підвищити рівень їх залучення та задоволення [120].

Маркетинговий менеджмент охоплює більше, ніж просто продаж товарів чи послуг; він передбачає створення цінності для клієнтів та встановлення довготривалих зв'язків з ними. Маркетинг, як управлінська функція, намагається забезпечити керівництво та цілі для дій організації. Основною метою маркетингового менеджменту є розробка гнучкого бізнес-плану, який може ефективно досягти запланованих результатів і дозволить комерційним організаціям вистояти за дуже складних обставин.

Крім того, **маркетинговий менеджмент** – це дисциплінований підхід, який застосовує підприємницькі принципи для адаптації до мінливих обставин сьогодення. Він зосереджується на раціональному та логічному використанні людського капіталу, сировини, інфраструктури, фінансових активів та інших ресурсів для забезпечення прибутковості, стійкості та довгострокового розширення [7].

Роль маркетингу в організаціях зазнала трансформації, перетворившись з окремої функції на всеохоплюючу відповідальність, яка поширюється на всіх працівників. Цей рух зумовлений зміною поведінки клієнтів, які більше не сприймають маркетинг окремо від самого продукту, а радше розглядають його як невід'ємну частину продукту. Усі дії, що їх здійснює комерційна організація, підпадають під маркетинг і відіграють вирішальну роль у визначенні загального клієнтського досвіду.

Згідно цієї ідеї, досягнення ефективного залучення клієнтів вимагає організаційної трансформації, коли контакти з клієнтами більше не сприймаються, як виключна компетенція окремих відділів організації, а скоріше, як цілісна серія взаємопов'язаних зустрічей. Для того, щоб досягти ефективного залучення клієнтів, необхідно, щоб усі співробітники комерційної організації займали проактивну позицію у спілкуванні з клієнтами та визначенні важливих точок контакту з ними.

Загалом, маркетинговий менеджмент перебирає на себе роль двигуна залучення та утримання клієнтів. Це означає, що маркетинг бере на себе відповідальність за визначення пріоритетів і стимулювання обговорень в межах усієї організації. Фактично, бізнесу необхідно прийняти клієнтоорієнтований підхід і побудувати всю свою структуру на основі сприяння взаємодії з клієнтами. Для досягнення успіху необхідно, щоб маркетинг пронизував усю діяльність комерційної організації.

Маркетинговий менеджмент охоплює багато функцій. Ця діяльність включає планування, організацію, керівництво і контроль активностей, пов'язаних з виробництвом і розповсюдженням бажаних товарів і послуг, задовольняючи потреби і вподобання клієнтів [4]. Практика ефективного управління маркетингом охоплює розробку та реалізацію стратегій, спрямованих на залучення, утримання та розширення клієнтської бази. Це передбачає визначення маркетингових цілей, встановлення пріоритетів, побудову комплексної маркетингової стратегії та послідовне досягнення фінансових і організаційних цілей.

Маркетинговий менеджмент все більше покладається на методології, засновані на даних, і посилює свою аналітичну природу. Враховуючи стрімке зростання кількості даних про клієнтів та рівень аналізу, необхідний для їх ефективної обробки, організації повинні визначити чіткі ролі та сприяти міжфункціональній співпраці, щоб ефективно збирати, організовувати, аналізувати та поширювати дані, отримані в результаті різноманітних взаємодій з клієнтами. Загалом, сфера аналізу даних має потенціал для надання

маркетологам безлічі цінних інсайтів, які можуть ефективно сприяти покращенню взаємодії з клієнтами, підвищенню рівня їхньої задоволеності та збільшенню прибутку. Однак реалізація цього потенціалу вимагає всебічного розуміння того, як аналітика даних буде невпинно трансформувати сферу маркетингу [51].

Виходячи з емпіричних спостережень, очевидно, що більшість керівників вищої ланки усвідомлюють необхідність модернізації маркетингу. Однак, як видається, серед цих керівників існує певна невизначеність, щодо точної природи та складових цієї модернізації. Тому для бізнесу життєво важливо мати всебічне розуміння компонентів, з яких складається сучасна маркетингова модель (Таблиця 1.1). Сучасний маркетинг вимагає підвищення компетенцій, щоб ефективно адаптуватися до мінливої динаміки та вимог маркетингового ландшафту.

Таблиця 1.1

Рівні розвитку маркетингового менеджменту комерційної організації

Маркетинговий компонент	Низький рівень розвитку	Високий рівень розвитку
Стратегія та інсайти	Оновлення бренд-стратегії кожні два-три роки. Первинні дослідження та фокус-групи використовуються для отримання інсайтів.	Вдосконалення бренд-стратегії щонайменше раз на рік. Використання передової аналітики та джерел інсайтів у реальному часі.
Продукт і ціна	Орієнтація на продукт. Продукти та цінові рішення значною мірою ізольовані від ключових знань про клієнта.	Розвиток продуктового портфелю на основі клієнтських інсайтів та аналітики. Ціноутворення визначається поєднанням досліджень та аналітики.
Клієнтський досвід та персоналізація	Пропозиція кількох стандартизованих продуктів для широких клієнтських сегментів з обмеженою персоналізацією.	Персоналізований для всіх каналів та життєвого циклу, клієнтоорієнтований досвід, що ґрунтується на глибокому розумінні клієнтських потреб та больових точок.
Вимірювання та ROI	Відсутність комплексного та регулярного вимірювання ефективності маркетингових активностей. Кожна кампанія має чітко визначені KPI.	Використання омніканальної аналітики ROI на рівні клієнта і щотижнева оптимізація параметрів кампаній. Наявність автоматизованих дашбордів з різними рівнями деталізації для різних стейкхолдерів.
Дані та технології	Маркетингові технології знаходяться у власності та управлінні IT-відділу, а основні можливості рекламних технологій - у власності та управлінні партнерського агентства.	Маркетинговий відділ визначає сценарії використання клієнтських даних та тісно співпрацює з IT-відділом над управлінням технологічними ресурсами.

Джерело: [2]

Значний внесок в дослідження особливостей побудови системи маркетингового менеджменту внесли такі вчені, як Віренга Б. [157], Кім В. та Моборн Р. [74], Овсієнко Н. [167], Архипова Т. [166], Аскарза Е., Фейдер П. та Харді Б. [3], Вотерворт Д. [155].

Цифрові канали, а також розробки у сфері аналітики та дата сайнс зробили можливим для маркетологів нести відповідальність за створення цінності для клієнтів та організації в усіх каналах. Щоб мати позитивну рентабельність інвестицій (ROI), кожен учасник маркетингового менеджменту повинен діяти так, ніби інвестиції, які він робить, є його власними. Це передбачає регулярний моніторинг витрат, встановлення стандартів для виявлення тих активностей, які не приносять користі. Така фінансова дисципліна не лише допомагає маркетингу виконати своє призначення, як рушія зростання, вона також встановлює довіру з фінансовим департаментом, відкриває доступ до нового фінансування та демонструє цінність маркетингу для всієї організації [2]. Оцінка ефективності маркетингу є складним завданням, яке неможливо адекватно відобразити за допомогою одного показника. У сфері усталених маркетингових практик часто доводиться використовувати кілька і, часто, недосконалих методів, щоб кількісно оцінити успіх маркетингових активностей. Тим не менш, організації мають потенціал для прийняття стандартизованого підходу для оцінки результатів у різних каналах. Додаткові показники, які організації можуть використовувати для моніторингу ефективності своїх маркетингових активностей, включають коефіцієнт залучення, коефіцієнт кліків, коефіцієнт конверсії, коефіцієнт переглядів, дохідність клієнта, вартість залучення клієнтів, життєву цінність клієнта, тривалість життя клієнтів, згадки в соціальних мережах, аналіз настроїв і рівень рекомендацій (NPS). Кожен з цих показників стосується окремих аспектів маркетингової ефективності, а разом вони дають повне уявлення про рентабельність маркетингових інвестицій (ROI) [33].

Як було сказано раніше, управління витратами стосується систематичного моніторингу та проактивного контролю фінансових витрат комерційної організації, зокрема, пов'язаних з маркетинговою діяльністю. Управління

витратами включає в себе ретельний нагляд за всіма витратами, пов'язаними з маркетинговою діяльністю. Основною метою управління витратами є оптимізація поточного маркетингового бюджету, що дозволяє організаціям більш ефективно розподіляти ресурси.

Ефективне управління витратами в маркетинговому менеджменті передбачає:

1. Забезпечення справедливого ціноутворення. Одним з найважливіших аспектів здійснення економічних операцій є визначення справедливої ціни. Як для покупців, так і для продавців важливо встановити взаємоприйнятну ціну, яка точно відображає вартість товарів чи послуг. Маркетологи повинні мати всебічне розуміння послуг, які вони закупають, і переконатися, що витрати, які вони несуть, є обґрунтованими, незалежно від того, чи ці послуги надаються внутрішніми або зовнішніми партнерами.

2. Плата лише за необхідні послуги. Для того, щоб найбільш ефективно залучати своїх цільових клієнтів, маркетологи повинні надавати пріоритет послугам, які дозволяють цього досягти. Маркетологи повинні вміти визначати критерії успіху для окремих кампаній і встановлювати ключові показники ефективності (KPI). Маркетингові витрати мають бути сконцентровані на каналах, які, згідно з аналітикою, є найефективнішими для встановлення зв'язку з цільовими клієнтами.

3. Зміна підходів до виконання роботи. Найважливішим фактором, який визначає чи варто маркетологам переглядати свої витрати є спосіб виконання своєї роботи. Для цього необхідно стати більш економними та гнучкими, використовувати нові технології, дані та аналітику для отримання глибшого розуміння в режимі реального часу, а також будувати організаційні структури, які сприятимуть ефективності та гнучкості. Більшої рентабельності інвестицій та кращих результатів можна досягти, якщо краще орієнтуватися на правильних клієнтів [21].

Варто виділити шість перешкод, з якими можуть зіткнутися організації у сучасному маркетинговому менеджменті:

1. Однією з основних перешкод в управлінні маркетингом є завдання ефективної адаптації до мінливих потреб та уподобань клієнтів. У відповідь на мінливі очікування клієнтів організації повинні змінювати свою маркетингову стратегію і продуктову пропозицію, щоб зберегти свою актуальність і конкурентоспроможність.

2. Іншою проблемою, з якою стикаються в управлінні маркетингом, є ескалація конкуренції. Зі збільшенням кількості конкурентів, що виходять на ринок, завдання виокремлення продуктів і послуг, та розробки унікальної пропозиції по відношенню до інших гравців на ринку стає складним.

3. Наступним викликом, з яким стикається маркетинговий менеджмент, є швидкий темп розвитку технологій. З появою нових технологій організаціям необхідно змінювати свою маркетингову стратегію, щоб скористатися новими можливостями і зберегти конкурентну перевагу [64].

4. Глобалізація створює значні проблеми у сфері маркетингового менеджменту. Оскільки організації виходять на нові ринки, для них стає необхідним змінювати свою маркетингову стратегію, щоб задовольнити різноманітні потреби і схильності клієнтів з різних країн або культур.

5. Регуляторні та правові проблеми можуть створювати перешкоди у сфері управління маркетингом. Організації зобов'язані дотримуватися різноманітних правил і законів, що стосуються реклами, безпеки продуктів, особливостей надання послуг та інших маркетингових заходів.

6. Останньою проблемою, з якою стикаються в маркетинговому менеджменті, є обмеженість ресурсів. Для того, щоб досягти своїх маркетингових цілей, комерційним організаціям вкрай важливо розподіляти свої маркетингові кошти у спосіб, який буде одночасно результативним та ефективним [7].

Каваллоне М. виділяє п'ять сценаріїв поведінки комерційних організацій за яких довіра до принципів маркетингового менеджменту може бути підірвана.

Один із прикладів називається **«Переконання вигадкою»**, коли організації застосовують віртуальний маркетинговий підхід, включивши відділ

маркетингу в свою організаційну структуру лише теоретично, не здійснюючи реальних маркетингових заходів. Це стосується бізнесів, які обмежилися впровадженням поверхневої форми маркетингу, наприклад, внесенням косметичних змін до організаційної структури або лише представленням відділу на папері.

Другий сценарій, який називається **«Я теж»**, виникає тоді, коли комерційні організації починають маркетингові зусилля лише після того, як спостерігають за тим, як це роблять їхні конкуренти, тим самим не помічаючи потенціалу для отримання конкурентної переваги за допомогою інновацій.

Третій сценарій, відомий як **«Короткозорість у відносинах»** – це нерозуміння справжньої важливості встановлення та управління зв'язками між різними підрозділами, особливо з відділами, які безпосередньо не взаємодіють з маркетингом, такими як логістика, адміністрація та виробництво. Це призводить до відриву маркетингу від реального контексту, в якому він працює, а недостатня увага до внутрішніх клієнтів (тобто інших підрозділів) ще більше знижує потенціал маркетингу, як відділу, що є рушієм змін і забезпечує лідерство.

Четвертий сценарій стосується **«Відсутності (або страху) змішування»**, коли організації демонструють нездатність ефективно об'єднати функції маркетингу та продажів. Хоча вони мають доступ до цінного дослідницького інструменту у вигляді власного відділу продажів, який перебуває в постійному контакті з клієнтами, вони часто не можуть повною мірою використати величезну кількість ринкової інформації, яку надає цей відділ.

Кінцева дилема – **«Хто є клієнтом?»**. Питання ідентифікації клієнта може здатися тривіальним, але це питання, яке комерційні організації іноді нехтують ставити собі. Припущення про володіння знаннями про те, що вважається «корисним і вигідним», в деяких випадках перетворило маркетинг на процес, який більше не бере до уваги бажання клієнтів, а радше сам диктує ці бажання [22].

Шляхом інтеграції академічних і практичних знань можна досягти більш значних успіхів у галузі маркетингового менеджменту. Дейтон Д. А., Мела К. Ф.

та Мурман К. виділяють такі переваги співпраці науковців з маркетологами-практиками:

1. Генерація кращих ідей. Науковці потребують точки опори для того, щоб розвивати знання. Кілька років нестандартного мислення та спостереження за маркетингом і поведінкою клієнтів на ринку підвищують ймовірність того, що ідеї будуть інноваційними.

2. Покращення даних. Більш активне залучення до практичного маркетингу надає науковцям доступ до свіжих джерел даних і сприяє зміцненню довіри, необхідної для обміну такою інформацією. Додатковою перевагою є те, що часто нові дані дозволяють точніше ідентифікувати емпіричні явища і виключити альтернативні пояснення.

3. Створення нових підходів. Співпраця з практиками може призвести до створення нових інструментів і методів дослідження маркетингового менеджменту, які можна застосовувати як в академічному, так і в промисловому секторах.

Перевагами для практиків, щодо співпраці з науковцями є:

1. Оновлення точки зору. Співпраця з науковцями може забезпечити практиків новими ідеями та рішеннями, пропонуючи їм альтернативні перспективи та об'єктиви, через які можна розглядати певну проблему, запропонувати різні точки зору. Співпраця з науковцями може надати маркетологам нові перспективи, які спонукають їх глибше переглянути свої стратегії і тактики.

2. Набуття міжгалузевих знань. Маркетологи, приймаючи рішення, мають справу з конкретними змінними (наприклад, ціна, канал) і діють у певному галузевому контексті. Ці особливості можуть обмежувати їхню здатність бачити більш потужні узагальнення, пов'язані з тим, що вони вивчили, які можуть бути застосовані в інших частинах їхнього бізнесу, інших маркетингових інструментах або в інших комерційних організаціях, які вони можуть очолити з часом. Академічні маркетингові знання, навпаки, побудовані на основі міжгалузевого підходу і можуть задовольнити ці потреби.

4. Створення нових підходів. Численні методи, що мають відношення до індустрії, включаючи конджойнт-аналіз, моделювання маркетингового міксу і, нещодавно, моделювання атрибуції, виникли в академічних колах.

5. Холодний погляд на проблему. Практики можуть заплутатися в нових технологіях і тенденціях, займаючись своїм повсякденним бізнесом. Неупереджена позиція науковця може бути дуже важливою, оскільки вона надає практикам більш надійну інформацію, яка має вирішальне значення в процесі прийняття рішень [40].

1.2 Роль розуміння поведінки клієнтів у маркетинговому менеджменті

Сфера клієнтської поведінки виникла, як окремий науковий напрям, близько півстоліття тому. У міру того, як сфера маркетингу почала відмежовуватись від економіки, вивчення поведінки клієнтів також почало відмежовуватись від маркетингових досліджень. Поведінкова економіка стала визначною концепцією в економічній науці [137].

Термін «клієнт» може використовуватися як синонім терміну «споживач». Тим не менш, фахівці з маркетингу визнають, що використання терміну «клієнт» краще передає факт встановлення взаємозв'язків, заснованих на придбанні продукту або послуги. Термін «споживач» на противагу, означає когось, хто тільки має потенціал скористатися товаром чи послугою [35].

З настанням індустріальної епохи спостерігається безперервний процес впровадження інновацій. Очікується, що впровадження виробничих і комерційних інновацій сприятиме впровадженню нових маркетингових методів, спрямованих на те, щоб впливати на клієнтів, спонукати їх до здійснення покупок і сприяти формуванню довгострокової лояльності.

За останнє десятиліття відбулися помітні зміни в динаміці взаємодії, коли клієнти стали відігравати більш впливову роль у визначенні умов взаємодії. Клієнт суттєво порушив маркетинговий ландшафт, реалізувавши свою нову лідерську позицію. З економічної точки зору, запропоновано називати цей період

«Епохою клієнта» [35]. Дослідження, проведені Мелдрум М. та Макдональд М. [97], Штаусс Б. та Зайдель В. [143], Хомбург К., Шефер Х. та Шнайдер Ж. [66], Кукук С. [84], Осборн П. та Баллантайн Д. [108] поглибили розуміння багатовимірних відносин з різними клієнтами. Їхні роботи сприяли збагаченню, як теоретичних, так і практичних знань у цій сфері.

Історичні періоди були свідками появи нових маркетингових практик внаслідок розвитку виробництва, технологій та комунікацій. Вищезгадані практики здебільшого формувалися під впливом маркетологів, і не зазнавали впливу з боку клієнтів. Очікується, що поширення маркетингових стратегій, орієнтованих на клієнтів, призведе до непередбачуваних результатів. На відміну від традиційного маркетингу, орієнтованого на управління попитом, впровадження клієнтоорієнтованого маркетингу змістить акцент маркетингової функції в бік управління пропозицією. Також, впровадження клієнтоорієнтованого маркетингу сприятиме залученню клієнтів та комерційних організацій до спільного створення продуктів, стратегій ціноутворення та методів дистрибуції [138].

У наш час клієнти відіграють значну і впливову роль у взаємовідносинах. Спостерігається помітне зростання залучення клієнтів до стратегічного планування, розробки продуктів та маркетингових програм. Поява такої взаємодії між клієнтом і брендом свідчить про зміну ієрархії між клієнтами та організаціями (рис. 1.1).

Розвиток клієнтоорієнтованого маркетингу призвів до розширення залучення клієнтів до процесу виконання замовлень, що породило концепцію "маркетингу співтворчості". Маркетинг спільного створення передбачає спільне залучення, як маркетологів, так і клієнтів на різних етапах життєвого циклу продукту або послуги, включаючи розробку, виробництво і споживання [78].

Ступінь реалізації маркетингу співтворчості залежить від здатності комерційної організації збирати та використовувати знання про клієнтів. Впровадження маркетингових стратегій співтворчості може підвищити лояльність клієнтів і зменшити операційні витрати [138].



Рис. 1.1 Новий порядок створення продуктів

Джерело: [35]

Зосередимося на розкритті поняття клієнтоорієнтованості. Суть цієї концепції полягає в тому, щоб поставити клієнта в центр комерційної організації. У сучасному бізнес-середовищі отримання прибутку більше не залежить виключно від здатності генерувати дохід через окремі бізнес-транзакції. Натомість, фокус змістився у бік створення життєвої цінності, що отримується від кожного клієнта. Клієнти розглядаються, як найцінніший актив, що вимагає початкового залучення та подальшого розвитку і підтримки протягом усього періоду ділових відносин [113].

Концепція клієнтоорієнтованого мислення базується на кількох ключових передумовах:

- По-перше, рівень прибутковості клієнтів має тенденцію до зростання з часом. Хоча новопридбаний клієнт може не відразу зробити внесок у прибутковість бізнесу, очікується, що він стане прибутковим у майбутньому.
- По-друге, витрати, пов'язані з обслуговуванням і впровадженням нових пропозицій для поточних клієнтів, є нижчими порівняно з витратами, понесеними для новопридбаних клієнтів.
- По-третє, не всі клієнти однаково вигідні, але лояльні клієнти, як правило, менш чутливі до змін цін, більш схильні до реакції на маркетингові активності, і більш пробачливі до випадкових помилок, яких припускається бізнес. Отже, вони мають більшу цінність для комерційної організації [113].

Для бізнесу є важливим поглиблення розуміння купівельної цінності клієнта, поділ клієнтів на сегменти відповідно до купівельної цінності, розробка бізнес-моделі, яка гарантує необхідний рівень прибутковості. Для того, щоб генерувати цінність для своїх акціонерів, комерційна організація повинна не лише надавати цінність своїм клієнтам, але й отримувати цінність у відповідь.

У дослідженні "Теорія оцінки клієнтів: Концепції, метрики, стратегія та впровадження", В. Кумар, представлено теорію оцінки клієнтів. Ця теорія ґрунтується на економічних концепціях, і має на меті кількісно оцінити, контролювати та оптимізувати цінність, яку клієнти надають бізнесу. У роботі розглядається життєва цінність клієнта (CLV), як показник, який може запропонувати надійну та перспективну оцінку прямої цінності клієнтів [81]. Також, значну увагу питанню життєвої цінності клієнтів у своїх дослідженнях приділяли Мін С. [101] та Бонаккі М. [15].

Значним зрушенням в клієнтоорієнтованій ідеології є перехід від прибутковості продукту до прибутковості клієнта, як основного показника організаційного успіху. Часто товари та послуги мають синергетичний зв'язок, тобто, певні продукти мають потенціал слугувати каталізатором для започаткування взаємовідносин, а подальше отримання прибутку відбувається за рахунок інших продуктів. Хоча ці окремі товари можуть не мати фінансового обґрунтування для свого існування, вони часто слугують невід'ємними компонентами, що сприяють підвищенню загальної прибутковості організації. Не зважаючи на те, що комерційна організація може не отримати негайного прибутку від первинної покупки клієнта, подальші та повторні покупки можуть призвести до суттєво іншого результату [113]. Прикладом може слугувати надання банком кредиту під пільгові умови, що слугує важелем залучення, і подальшого продажу кредитної картки, що слугує важелем створення цінності для організації. Для бізнесу вкрай важливо мати всебічне розуміння механізмів, за допомогою яких генерується прибуток від клієнтів протягом усього періоду співпраці. Маркетингові стратегії повинні надавати пріоритет встановленню та підтримці відносин, як з існуючими, так і з потенційними клієнтами.

Використання систем управління взаємовідносинами з клієнтами (CRM) пережило сплеск популярності наприкінці 1990-х та на початку 2000-х років. Управління взаємовідносинами з клієнтами – це стратегічний підхід, який використовується бізнесом для підвищення прибутковості, доходів і задоволеності клієнтів [13].

Очевидно, що кожен клієнт має свої життєві цінності. Тому вкрай важливо впроваджувати комплексну програму управління клієнтами, спрямовану на залучення нових клієнтів, оптимізацію їхньої прибутковості, та перетворення збиткових клієнтів на прибуткових

Каваллоне М. визначає CRM, як стратегічний процес встановлення, розвитку, збереження та зміцнення довготривалих і взаємовигідних зв'язків між клієнтами та комерційною організацією. Ефективне управління взаємовідносинами з клієнтами ґрунтується на розумінні їх вимог і побажань [22].

Кумар В. та Рейнартц В, у своїй книзі, подали всебічний аналіз стратегічних та операційних елементів сучасного управління взаємовідносинами з клієнтами. У книзі постійно наголошується на необхідності розуміння економічної цінності клієнта, як основного керівного принципу для прийняття маркетингових рішень [82]. Управління взаємовідносинами з клієнтами може слугувати цінним інструментом для пошуку сегментів клієнтів, які приносять прибуток, шляхом аналізу їхньої купівельної поведінки [13].

В останні роки концепція CRM набула ширшого значення, охоплюючи комплексний процес встановлення та зміцнення вигідних відносин з клієнтами, шляхом надання їм цінності та підвищення рівня задоволеності. У цьому контексті CRM охоплює всі зусилля, спрямовані на залучення, утримання та розширення клієнтської бази [22].

Розвиток теорії CRM ще більше підкреслюється твердженням, що «CRM є центральною концепцією сучасного маркетингу». Це також підтверджується наявністю великої кількості сучасних досліджень даної концепції, зокрема працями Оберг К. [107] та Парович І. [114].

Твердження про те, що «всі клієнти різні», слугує фундаментальною передумовою у сфері маркетингу, підкреслюючи думку про те, що кожен клієнт має унікальні уподобання та вимоги, а отже, унеможливорює існування ідентичних клієнтів. Ця концепція має важливе значення для розробки маркетингових стратегій, оскільки дозволяє організаціям адаптувати свої пропозиції, послуги та рекламну активність до окремих сегментів клієнтів, а не застосовувати загальний підхід. Розуміючи та стратегічно орієнтуючись на різні групи клієнтів, відповідно до їхніх потреб, комерційні організації можуть підвищити ймовірність досягнення успіху та отримати більш значну конкурентну перевагу.

Управління гетерогенністю клієнтів є важливим стратегічним рішенням, яке організації повинні прийняти для того, щоб ефективно задовольнити різноманітний спектр клієнтських уподобань, присутніх на ринку [59].

Пальмат'є Р. В. та Крецеліус А. Т. виділяють п'ять ключових джерел гетерогенності клієнтів [112]:

- індивідуальні відмінності;
- життєвий досвід;
- функціональні потреби;
- самоідентичність;
- маркетингова діяльність.

Врахування індивідуальних особливостей, таких як стать та особистість, завжди було важливим для маркетологів. Ці відмінності роблять значний внесок у різноманітність моделей купівельних вподобань, особливо щодо таких виразних товарних категорій, як транспортні засоби або мода. Різноманітний життєвий досвід, що охоплює такі фактори, як культурне походження та фінансовий рівень, може суттєво впливати на особливості, які спостерігаються в процесі прийняття рішень про покупку. Функціональні потреби охоплюють матеріальні або практичні характеристики продукту, такі як його здатність протистояти зношуванню або енергозбереження, які мають різний ступінь важливості для клієнтів. Самоідентифікація є впливовою змінною, яка стимулює

розвиток унікальних самооцінок серед клієнтів і згодом впливає на їхні дії, в тому числі на формування схильності до конкретного бренду, який відповідає їхньому самовираженню. Зрештою, вплив маркетингових активностей на вподобання та сприйняття клієнтів має потенціал для зміни їхніх уподобань [112].

Сьогодні переважна більшість керівників визнають, що задоволення постійно зростаючого різноманіття вимог клієнтів є критично важливим. З одного боку, вони впроваджують нові програми та процеси, щоб задовольнити специфічні потреби кожного клієнта. З іншого боку, керівники визнають, що прийняття такої стратегії призведе до надмірних витрат і ускладнення їхнього функціонування. Ця скрутна ситуація являє собою управлінську перешкоду, яка потенційно може стати важливим чинником конкурентної переваги. Питання різноманітності створює значну перешкоду для створення цінності для організації через її прямий вплив на різницю між вартістю та готовністю платити.

Для багатьох комерційних організацій необхідність обслуговувати різні сегменти клієнтів є критично важливою для їхнього виживання. Наприклад, у комерційному банкінгу банк повинен обслуговувати мільйони роздрібних клієнтів, на додаток до тисяч великих корпоративних бізнес-клієнтів [83].

У сучасному конкурентному середовищі комерційні організації мають два варіанти реагування на різноманітні потреби клієнтів: задовольнити середньостатистичного клієнта універсальним продуктом або задовольнити різноманітні індивідуальні вимоги клієнтів.

Бізнеси з продуктово-орієнтованим підходом ставлять на перше місце задоволення ринкових потреб, визначаючи середні вимоги клієнтів на ринку. Масовий маркетинг використовується для задоволення типових потреб ринку. З іншого боку, клієнтоорієнтована організація визнає відмінності між клієнтськими сегментами та надає пріоритет пристосуванню своїх послуг до унікальних потреб кожного окремого клієнта або групи клієнтів. Коли до клієнтів ставляться однаково, можна зробити висновок, що комерційна організація вважає цінності, вимоги та пріоритети клієнтів неоднозначними або

несуттєвими. Таким чином, продукти та рівні обслуговування створюються за допомогою відтворюваної стандартизованої процедури, яка є простою та більш економічно ефективною.

Альтернативним підходом до пом'якшення вищезгаданих проблем може бути однакове ставлення до всіх клієнтів у межах певної категорії. Визначивши ці однорідні клієнтські сегменти, що характеризуються спільними цінностями, потребами та пріоритетами, стає можливим адаптувати різні продукти та послуги до їхніх конкретних вимог. Таким чином, основна проблема, пов'язана з відбором клієнтів, полягає в практиці сегментації клієнтів. Організації повинні одночасно задовольняти різноманітність потреб клієнтів і прагнути до скорочення витрат, або ж намагатися мінімізувати витрати без шкоди для відповідного рівня задоволеності клієнтів.

Клієнти в межах певного сегмента мають спільний набір цінностей, що призводить до однакового рівня задоволення певними особливостями продукту чи послуги. Існують різні методи сегментації клієнтів для маркетингових цілей, одним із найпоширеніших підходів є демографічна сегментація. Цей метод передбачає розподіл людей на категорії на основі їхніх демографічних характеристик, таких як вік, стать, рівень освіти та економічного достатку.

Розуміння купівельної цінності клієнтів може надати цінну інформацію про важливість конкретних переваг та особливостей продукту чи послуги для різних сегментів клієнтів. Одним із класичних підходів такої оцінки є RFM, де R (Recency) є мірою давності останнього замовлення, а F (Frequency) та M (Monetary) є кількістю та вартістю усіх замовлень клієнта відповідно. Таке розуміння може також виявити, які сегменти клієнтів схильні платити преміальну ціну або домовлятися про нижчу ціну за певні атрибути продукту чи послуги. Крім того, це може допомогти визначити сегмент клієнтів, який сприяє довгостроковій прибутковості. При розробці клієнтоорієнтованої стратегії використання сегментації клієнтів на основі цінностей дає цінну інформацію про купівельну поведінку, що дозволяє більш ефективно таргетувати та обслуговувати клієнтів [113].

Часто комерційні організації стикаються з проблемами у формуванні лояльності клієнтів через надмірні витрати, пов'язані з програмами лояльності та прямого продажу. З іншого боку, вони можуть ненавмисно віддалити клієнтів, намагаючись оптимізувати маркетингові витрати. У багатьох випадках не вдається досягти, як лояльності, так і низьких цін. Для маркетингу вкрай важливо прагнути до концепції оптимальної ефективності в усіх своїх програмах і процедурах [138].

Зосередимося на питанні типологізації взаємовідносин у бізнес-середовищі. Загалом, основними моделями комерційної взаємодії є:

- **B2B** (від бізнес до бізнесу) - стосується комерційного обміну товарами або послугами між бізнесами;
- **B2C** (від бізнес до споживача) - стосується комерційних операцій, коли комерційні організації безпосередньо пропонують товари або послуги індивідуальним споживачам;
- **B2G** (від бізнес до уряду) - це взаємодія, в яких організації беруть участь у наданні товарів або послуг державним установам або громадським організаціям;
- **C2C** (від споживача до споживача) - це практика фізичних осіб, які беруть участь у прямих транзакціях з продажу товарів чи послуг іншим особам;
- **C2B** (від споживача до бізнесу) – стосується практики, коли споживачі беруть участь у продажу товарів або послуг організаціям;

Вієйра В. А. та ін. бізнес-клієнтами визначають організації, які займаються закупівлею товарів чи послуг у інших організацій з метою полегшення та покращення власної операційної діяльності [151]. На відміну від маркетингу B2C, який орієнтований на індивідуальних клієнтів, маркетинг B2B спрямований на розвиток зв'язків з іншими організаціями та пропозиції індивідуальних рішень для задоволення їхніх потреб та усунення перешкод.

Клієнти бізнес-сегменту часто демонструють підвищену потребу у вичерпній інформації, що стосуються продуктів і послуг, перш ніж остаточно визначитися з вибором покупки. Крім того, їхні процеси прийняття рішень, як правило, є більш складними і передбачають участь кількох зацікавлених сторін.

Розуміння та вивчення поведінки бізнес-клієнтів має важливе значення в управлінні маркетингом, через різні переконливі причини.

1. Перш за все, важливо зазначити, що бізнес-клієнти, як правило, мають вищий рівень складності та вибагливості, порівняно з індивідуальними клієнтами. Відповідно, їхні рішення про покупку часто формуються під впливом різноманітних факторів, включаючи: корпоративну культуру, динаміку всередині закупівельного відділу та специфічні вимоги до вирішення проблем.

2. Крім того, варто зазначити, що бізнес-клієнти часто проходять через процеси прийняття рішень, які є більш тривалими та складними, ніж у індивідуальних клієнтів. Ці процедури, як правило, передбачають залучення кількох зацікавлених сторін і потребують більшого обсягу інформації та знань, щодо пропонованих продуктів і послуг.

3. Бізнес-клієнти часто роблять акцент на встановленні довготривалих партнерських відносин і надають великого значення довірі та авторитету при виборі постачальників. Отже, ефективне управління маркетингом вимагає концентрації зусиль на створенні та підтримці міцних відносин з клієнтами. Ці зусилля охоплюють персоналізоване спілкування, індивідуальні рішення та безперервне надання підтримки і сервісу.

4. Крім того, варто зазначити, що ринки B2B, як правило, мають порівняно обмежений масштаб, що характеризується меншою клієнтською базою. Однак важливо підкреслити, що ці клієнти часто беруть участь у більших за обсягом транзакціях порівняно з ринком B2C [151].

Одними з ключових гравців на ринку B2B є банки. Бізнес-клієнти для банків є основним джерелом доходу, але водночас і основним споживачем індивідуальних рішень, стосовно продуктів та послуг. Маркетингові команди в банках зосереджені на наданні найкращого персоналізованого клієнтського досвіду бізнес-сегменту, для збереження довготривалих взаємовідносин з клієнтами, що в свою чергу, забезпечує сталість високих фінансових результатів банку. Іншими ваговими гравцями на цьому ринку є консалтингові та маркетингові агентства, акселератори та інвестиційні фонди. Часто клієнтами

даних організацій є стартапи, які активно звертаються до них за фінансовою підтримкою, консалтингом та менторством, шукаючи шляхи для свого зростання.

Стартапи демонструють особливу клієнтську поведінку, на відміну від класичних підприємств, яка характеризується швидкою гнучкістю, готовністю йти на ризики, зосередженістю на інноваціях та сильний акцент на швидкому розширенні. Важливими елементами є обмежені фінансові ресурси, потреба в адаптивності, орієнтація на оригінальність, завоювання визнання та довіри. Вивчення їхньої діяльності дозволяє інвесторам робити обґрунтований вибір, маркетологам проводити точні та цілеспрямовані маркетингові заходи, вдосконалювати та покращувати продукт.

Також, свої дослідження, стосовно характеристики бізнес-клієнтів та стратегії ефективної роботи з ними, з точки зору маркетингу, присвятили Віртц. Б.В. [159], Віейра В.А. та ін. [151]. В тому числі, окремі дослідження стосовно впровадження маркетингового менеджменту з урахуванням особливостей поведінки стартапів займалися такі вітчизняні та закордонні вчені, як Шарма С. та Гоял Д. [52], Д'Авіно М., Де Сімоні В., Іаннуччі М. та Ширальді М. [45], Усман М. та Сунь Х. [148], Кофанов О. та Зозульов О. [77].

Однією з найважливіших передумов у сфері маркетингу є те, що всі клієнти з часом зазнають змін. На динамічний характер поведінки та вподобань клієнтів впливають нові тенденції, технологічний прогрес, інноваційні рішення та безліч інших факторів. Для того, щоб ефективно управляти змінами, комерційним організаціям необхідно мати повне уявлення про причини цих змін та вчасно виявляти відповідні тренди на випередження [112]. Детальному дослідженню особливості поведінки клієнтів присвятило свої роботи багато науковців, зокрема: Райан Е. та Плойхарт Р. [125], Канг М. та Гонг Т. [71].

Кундарі А. [35] характеризує поведінку клієнтів, як поведінкові моделі та процеси прийняття рішень, які демонструють люди та організації під час купівлі товарів чи послуг. Вона охоплює різні детермінанти, які впливають на прийняття рішень клієнтами, включаючи індивідуальні вподобання, суспільні впливи та

рекламні комунікації. Розуміючи поведінку клієнтів, комерційні організації можуть адаптувати свої маркетингові комунікації та товарні пропозиції для більш ефективного задоволення потреб та вподобань своїх клієнтів, що в кінцевому підсумку призведе до підвищення рівня їхньої задоволеності та лояльності.

Бернштейн Я. С. [11] стверджує, що процес вибору ґрунтується не на прямому порівнянні подій, а скоріше на порівнянні спогадів, пов'язаних з цими подіями. Розмірковуючи про майбутнє, люди, як правило, не концептуалізують свої майбутні події у традиційний спосіб. Концепцію майбутнього можна розуміти, як сукупність очікуваних спогадів.

Створення незабутніх вражень є ключовим елементом ефективного залучення клієнтів. Клієнти беруть участь у процесах прийняття рішень на основі своїх спогадів про минулі події, а не на основі самого досвіду. Маркетологи повинні застосовувати більш проактивний підхід, сприяючи формуванню незабутніх вражень у своїх клієнтів. Першочерговий акцент має бути зроблений на культивуванні цікавого досвіду, щоб клієнти зберегли його як сприятливу і варту уваги зустріч, а згодом поширювали добрі відгуки.

На рішення клієнтів скористатися послугою впливають різні фактори, зокрема внутрішні та зовнішні вигоди, соціальні фактори, такі як прагнення до соціальних зв'язків, вплив лідерів думок та референтних груп, а також когнітивні фактори, такі як знайомство та досвід. Крім того, спосіб життя та особистість людини, також, відіграють певну роль у формуванні процесу прийняття рішень. Клієнти демонструють різні вподобання та мотивації, коли йдеться про споживання. Отже, перед маркетологами стоїть завдання розпізнати ці елементи і націлити на них свої маркетингові ініціативи, щоб підвищити залученість:

1. Культурні норми – це колективні ідеї, цінності та умовності, які впливають на поведінку людей у межах певної культури чи соціальної групи. Маркетологи мають можливість використовувати культурні норми як стратегічний інструмент, розробляючи маркетингові кампанії, які ефективно резонують з культурними цінностями та уподобаннями цільової аудиторії.

2. **Соціальні впливи** охоплюють безліч елементів, які впливають на поведінкові моделі клієнтів у межах їхніх соціальних мереж, включаючи сімейні стосунки, дружбу, професійні контакти та інші громадські зв'язки. Маркетологи мають можливість використовувати соціальний вплив як стратегічний підхід, розробляючи маркетингові кампанії, які ефективно орієнтуються на соціальні потреби та прагнення певних сегментів аудиторії.

3. **Особисті цінності** – це суб'єктивні погляди та ставлення людей, які мають вплив на поведінку клієнтів.

4. **Психологічні елементи**, такі як мотивація, сприйняття та навчання, що відіграють значну роль у впливі на поведінку людини.

5. **Макросередовищні тенденції**, що охоплюють значні соціальні, політичні, економічні та технічні фактори, які суттєво впливають на поведінку клієнтів.

Процес прийняття рішення – це багатогранне явище, яке характеризується послідовністю етапів. Бернштейн Я. С. [11] серед основних етапів виділяє: усвідомлення потреби, пошук інформації, оцінку наявних варіантів, залучення людей до процесу прийняття рішення та подальшу поведінку після покупки.

1. **Визнання потреби.** Клієнти демонструють усвідомлення скрутного становища або потреби, що може бути спровоковано зовнішніми або внутрішніми стимулами. На цій фазі клієнти розпізнають потребу, яку вони мають намір задовольнити через споживання. Внутрішні стимули, такі як голод або спрага, або зовнішні стимули, такі як маркетингова комунікація, можуть викликати потребу.

2. **Пошук інформації.** Клієнти активно шукають інформацію про різні альтернативи, які є в їхньому розпорядженні і допомагають їм дійти висновку. Клієнти можуть отримувати інформацію з різних джерел, зокрема, з емпіричних джерел, що ґрунтуються на попередньому досвіді використання подібних продуктів, комерційних джерел, зокрема, реклами та електронних розсилок, публічних джерел, зокрема, засобів масової інформації та оглядів, або особистих

джерел, зокрема, від друзів та родини. На цьому етапі збираються дані, щоб визначити, які продукти та послуги задовольняють їхні потреби.

3. Оцінка альтернатив. Третій етап процесу прийняття рішень передбачає оцінку альтернатив шляхом порівняння їхніх цін, характеристик і переваг, щоб визначити, яка з них є найбільш прийнятною для конкретної потреби.

4. Процес прийняття рішення та поведінка після покупки є завершальною фазою в процесі прийняття рішення. Поведінка після покупки стосується вивчення дій клієнта після покупки, а також рівня його задоволеності придбаним продуктом або послугою. Значення цієї поведінки полягає в тому, що вона може дати маркетологам уявлення про задоволеність і лояльність, які відіграють важливу роль у стимулюванні повторних покупок, сприяють поширенню позитивних відгуків і, в кінцевому підсумку, підвищують прибутковість організації [11].

Поява смартфонів зробила революцію у сфері зв'язку, значно підвищивши рівень доступу клієнтів до інформації. Клієнти отримали можливість оперативного доступу до інтернету, де вони можуть отримувати інформацію про ціни та наявність необхідних товарів [35]. Інтернет став важливим технологічним інструментом, який дозволяє маркетологам доставляти персоналізований контент і здійснювати транзакції зі значно меншими витратами, порівняно з альтернативними медіа-платформами [138]. Проте, в епоху розвитку цифрових технологій офлайн взаємодія, однаково, залишається важливою в житті клієнта. Існує постійний зв'язок між цифровим та офлайн-світом, причому кожен з них впливає на інший і підсилює його. Кім Д. та ін. виявили, що загалом соціальна взаємодія в офлайні позитивно впливає на попит онлайн покупок, причому активна соціальна взаємодія в офлайні забезпечує позитивний інформаційний вплив [73]. Умайма Б. досліджував вплив соціальних медіа на купівельну поведінку. Результати свідчать про те, що соціальні медіа мають значний вплив на купівельний вибір клієнтів, особливо на етапах збору інформації та оцінки варіантів [109]. Схожі висновки були зроблені і у дослідженнях [130] та [63].

Процес засвоєння інформації та формування ставлення окремих клієнтів має важливе значення для маркетингу, оскільки він формує їхні потреби та впливає на стратегії їхнього задоволення. Основна мета маркетингу – розпізнати закономірні моделі поведінки в конкретних обставинах, таким чином підвищуючи здатність маркетолога задовольняти запити клієнтів. Цей фундаментальний процес формує основу маркетингу [85].

У зв'язку з еволюцією поведінки сучасних клієнтів, завоювання їхньої лояльності та переконання стати адвокатами може становити значний виклик. Зміна в поведінці клієнтів полягає в переході від пасивності до активного залучення, а поява адвокатів представляє вищий рівень залучення клієнтів.

Дослідженням особливостей поведінки клієнтів в процесі їх залучення займалися Маджид К. А. [95] та Свейн С.Д., Бергер П.Д., Вайнбер Б. Д. [145].

Створення оптимальних умов для залучення клієнтів та адвокації потребує значних витрат часу та системного підходу до взаємодії. Крім того, важливо зазначити, що досягнення цих результатів не залежить від традиційних стратегій інформування про продукт. Роль організаційних принципів та бачення стає все більш значущою у налагодженні відносин. Адвокати бренду, порівняно з іншими клієнтами, регулярно вивчають і тестують продукти чи послуги, щоб переконатися, що вони відповідають їхнім особистим поглядам, перш ніж зробити висновок. Потенціал людей стати адвокатом бренду залежить від ефективності маркетингового менеджменту на кожному з етапів життєвого циклу клієнта (рис. 1.2).

1. Початковим кроком у цьому процесі незмінно є підвищення рівня **обізнаності**. На цьому кроці маркетологи намагаються оцінити рівень поінформованості про бренд та сприйняття бренду серед клієнтів, щодо ідентичності та цінностей організації.

2. **Залучення клієнтів**. На додаток до простої обізнаності, проводиться оцінка того, чи клієнти активно взаємодіють з брендом, взаємодіючи з рекламою, беручи участь в маркетингових заходах, надаючи контактну інформацію, тощо.



Рис. 1.2 Шлях до адвокату бренду

Джерело: [35]

3. Прихильність клієнтів – визначається мірою, з якою клієнти посилили свої стосунки з брендом, завдяки таким діям, як публікації та коментування чи написання відгуків про бренд.

4. Адвокати бренду з'являються на останньому етапі, коли вони активно підтримують бренд, рекомендуючи його своїм знайомим, ділячись посиланнями, публікуючи пости в соціальних мережах або беручи участь у спільноті користувачів бренду [35].

Погляди та поведінка людей перебувають під безпосереднім впливом їхніх референтних груп. Вищезгадані групи охоплюють, як неформальні основні групи, включаючи сім'ю, друзів, сусідів та колег, так і вторинні групи, такі як релігійні та професійні об'єднання [11].

Каваллоне М. Запропонував вслане бачення траєкторії, яку проходять клієнти від початкової транзакції до досягнення стану лояльності до бренду. Вона складається з чотирьох окремих етапів, кожен з яких відіграє вирішальну роль у формуванні та встановленні лояльності [22].

1. Початковий етап включає в себе процес придбання, під час якого клієнт вирішує придбати певний товар або послугу. На прийняття цього рішення впливають різноманітні аспекти, включаючи потреби та бажання клієнтів, сприйняття цінності товару чи послуги, а також конкурентне середовище, в якому функціонує бізнес. Щоб бути ефективними, комерційні організації

повинні ретельно продумати, як вони можуть диференціювати себе від конкурентів, створюючи ціннісну пропозицію, яка буде дуже привабливою для клієнтів.

2. Наступним етапом розвитку є акт повторної покупки. Для того, щоб перевести клієнтів від разових транзакцій до постійних покупок, комерційні організації повинні надавати пріоритет продажу високоякісних продуктів та послуг, винятковому обслуговуванню клієнтів та розвитку міцних відносин з клієнтами. Зокрема, повинні бути розроблені стратегії, які б заохочували клієнтів здійснювати наступні транзакції, посилюючи сприйняття цінності їхньої першої покупки. Після того, як продавець успішно ідентифікував потенційного клієнта і сприяв здійсненню покупки, наступною метою є формування лояльності клієнта через реалізацію стратегій перехресних і повторних продажів.

Перехресні продажі – це практика пропонування додаткових товарів або послуг вже існуючому клієнту. На відміну від процесу придбання нового бізнесу, перехресні продажі пов'язані з потенційним ризиком порушення існуючих відносин з клієнтами. Тому важливо гарантувати, що додатковий продукт або послуга, які пропонуються клієнту, підвищують цінність, отриману від організації.

Допродаж – це стратегія продажу, за якої продавець намагається переконати клієнта придбати дорожчі товари, модернізацію або більше доповнень, з метою укладення вигіднішої, з фінансової точки зору, угоди. Практика допродажів, зазвичай, передбачає просування послуг або товарів, які приносять більший прибуток. Однак допродаж також може охоплювати ознайомлення клієнтів з альтернативними можливостями, про які вони раніше не здогадувалися [13].

Збільшення обсягу покупок клієнта означає два сприятливих наслідки:

- По-перше, маржинальність відносин підвищується відповідно до логічних принципів – обсяг продажів збільшується, незважаючи на те, що клієнтська база залишається незмінною.

– Другий позитивний аспект стосується конкуренції. Фактично, в міру того, як розширюється вибір продукції клієнтом, створюється віртуальний бар'єр, який ефективно перешкоджає входженню на ринок прямих конкурентів. Цей бар'єр перешкоджає входженню на ринок тих, чиї пропозиції стають менш привабливими через те, що додаткові продукти чи послуги клієнта покращують початковий продукт, придбаний на початку встановлення взаємозв'язків з організацією.

3. Наступний етап за методологією Каваллоне М. – це утримання клієнтів, на якому комерційні організації прагнуть підтримувати взаємодію з клієнтами та зберігати їхню зацікавленість у пропонованих продуктах і послугах. Ця мета може бути досягнута за допомогою різних підходів, включаючи впровадження програм лояльності або надання індивідуального досвіду для клієнтів. На цьому етапі важливо приділяти першочергову увагу створенню сприятливого емоційного зв'язку з клієнтами, тим самим сприяючи їх постійній підтримці та лояльності до бренду.

Термін «утримання клієнтів» можна тлумачити у двох різних значеннях, кожне з яких відображає різні підходи комерційних організацій до збереження взаємовідносин з клієнтами. Перше значення є позитивним і передбачає проактивні зусилля, спрямовані на встановлення та підтримку міцних відносин з клієнтами. Друга концепція передбачає реалізацію стратегій, спрямованих на створення бар'єрів, що перешкоджають клієнтам віддати перевагу конкурентам.

Особи, яким надається індивідуальна послуга або продукт, навряд чи перейдуть до іншого постачальника через тривалий процес донесення своїх потреб та відповідний час, необхідний альтернативному постачальнику для ознайомлення з клієнтом.

Визначивши рівень інвестицій для кожного кластера, організація повинна визначити ряд факторів, які мотивують і максимізують рівень лояльності клієнтів: зі зростанням лояльності клієнтів зростає і прибутковість бізнесу, оскільки лояльні клієнти, як правило, збільшують частоту і кількість своїх

покупок, вони більш схильні купувати додаткові продукти і легше піддаються обслуговуванню.

4. Завершальним етапом у просуванні є формування лояльності клієнта. На цьому етапі встановлюється міцна прив'язаність до бізнесу та його пропозицій. Вирішальним фактором у встановленні такого емоційного зв'язку з клієнтами є постійне забезпечення найкращого клієнтського досвіду, коли докладаються зусилля, щоб перевершити очікування клієнтів і задовольнити їхні запити. Це передбачає проактивне виявлення потреб клієнтів, надання високоякісних продуктів і послуг, а також культивування довіри та спільної атмосфери серед клієнтів [22].

Клієнти, які пов'язані з поширенням нових цифрових каналів, технологій і продуктів, суттєво змінили свої підходи до прийняття рішень. Вищезгадана нова поведінка спонукала до формулювання моделі «Клієнтська подорож прийняття рішення» (CDJ), яка складається з наступних п'яти етапів:

1. попередній розгляд;
2. активна оцінка;
3. момент придбання;
4. досвід після придбання продукту;
5. петля лояльності.

На відміну від попередніх моделей процесу прийняття рішень клієнтами, CDJ забезпечує три суттєві покращення:

1. **Розгляд бренду.** Клієнти вступають у фазу початкового розгляду із заздалегідь визначеним переліком брендів у своєму списку покупок, а отже, впізнаваність бренду має важливе значення для того, щоб забезпечити собі місце в списку. Однак, коли вони починають фазу оцінки та збирають більше даних, вони можуть вирішити збільшити кількість брендів у своєму списку, а не зменшити його, як це було у випадку з попередньо описаною моделлю.

2. **Розширення можливостей клієнтів.** Клієнти тепер мають владу над процесом і можуть отримувати інформацію відповідно до своїх конкретних вимог та етапу прийняття рішення. Зокрема, маючи багато джерел отримання

інформації, вони не залежать від маркетолога в наданні їм інформації про товар чи послугу.

3. Дві, а не одна форма лояльності. Надання післяпродажної підтримки для зміцнення або підтримки лояльності клієнтів не є чимось новим, однак, згідно з моделлю CDJ, не всі форми лояльності є однаковими. Згідно з результатами дослідження, існують активні лоялісти (адвокати), які залишаються відданими бренду та активно його просувають, а також пасивні лоялісти, які залишаються відданими доти, доки не з'являється краща альтернатива. Маркетологи повинні забезпечити, щоб пасивних лоялістів, яких у шість разів більше, ніж активних, не відволікалися на конкуруючі пропозиції. Незважаючи на те, що і ті, і інші є надзвичайно цінними клієнтами, вони вимагають різних підходів.

Ще однією моделлю процесу прийняття рішень є модель клієнтського шляху, запропонована маркетологами практиками, і яка відома як "Моменти істини". Згідно даного підходу стверджується, що успіх бренду визначається двома критичними моментами: першим моментом істини (FMOT) і другим моментом істини (SMOT).

FMOT відбувається, коли клієнт заходить до магазину, обирає і бере з полиці новий бренд. SMOT виникає, коли клієнт має позитивний або негативний досвід використання продукту вдома. За даними фахівців з маркетингу, клієнти з більшою ймовірністю здійнять повторну покупку, якщо бренд досягне успіху в обох цих сценаріях.

Точки дотику стали нагальною проблемою для маркетологів за останні два десятиліття. Однак до цього часу клієнт не був активно залучений до процесу дотику. Як наслідок, бренди повинні бути обізнаними та обережними, щодо своїх методів контакту та взаємодії зі ними. Це нове явище можна проілюструвати на прикладі моделі "Нульового моменту істини" (ZMOT) від Google [88].

ZMOT можна розглядати як попередник моделі FMOT, що доповнює FMOT і SMOT в той момент, коли клієнт вперше стикається зі стимулом. Таким

чином, відвідувачі отримують знання про продукт або послугу за допомогою своїх улюблених цифрових пристроїв. На той час, коли вони роблять замовлення онлайн, піднімають слухавку або відвідують магазин (FMOT), вони вже надзвичайно добре поінформовані. Вони також беруть до уваги відео, веб-сайти брендів, відгуки колег, сторонні рейтинги та огляди, традиційні ЗМІ та результати пошукових систем. Не дивно, що клієнти, ймовірно, додадуть додаткові бренди до свого списку вподобань і відмовляться від деяких зі своїх початкових виборів у світлі цього дослідження.

Сьогодні взаємодія з клієнтом – це більше, ніж просто закидати його повідомленням і сподіватися, що він до нього дослухається. Вкрай важливо усвідомити, що комерційні організації повинні бути повністю залучені в розмову в той момент, коли клієнти про це просять, а не тоді, коли бізнесу зручно.

Практика попередніх покупок суттєво вплинула на поведінку клієнтів. У минулому це було пов'язано насамперед з дорогими об'єктами, такими як автомобілі, електроніка або житло. Тепер люди віддають перевагу дослідженню перед покупкою навіть найдрібніших речей. Автор [35] стверджує, що інформаційних бар'єрів більше не існує. Покупці мають необмежений доступ до інформації в будь-який час і в будь-якому місці. Будь-який негативний досвід буде поширюватися через їхні соціальні медіа. Крім того, коментарі не припиняють своє існування в цифровому просторі, а зберігаються протягом невизначеного часу в безстроковому цифровому архіві, де вони і надалі будуть доступні на щоденній основі в майбутньому.

На основі усіх описаних моделей, Кундарі А. запропонував нову модель «Купівельна подорож клієнта» (рис 1.3), інтегрувавши компоненти FMOT+SMOT, ZMOT та CDJ, а також включивши власний контур впливу адвоката бренду. Ця модель, що включає численні точки дотику та петлі зворотного зв'язку, призначена для характеристики нелінійного процесу купівлі товарів та послуг, і являє собою всеохоплюючу модель, орієнтовану на клієнта.

Як зазначалося раніше, адвокати мають непропорційно великий вплив на репутацію та доходи бренду через поширення переконливих повідомлень та

контенту, що прискорює процес купівлі для інших. Адвокати бренду є активом, який варто захищати, оскільки завдяки активному просуванню бренду та його переваг, як для діючих, так і для потенційних клієнтів, вони мають більший вплив, ніж будь-які повідомлення, які бренд виробляє самостійно.

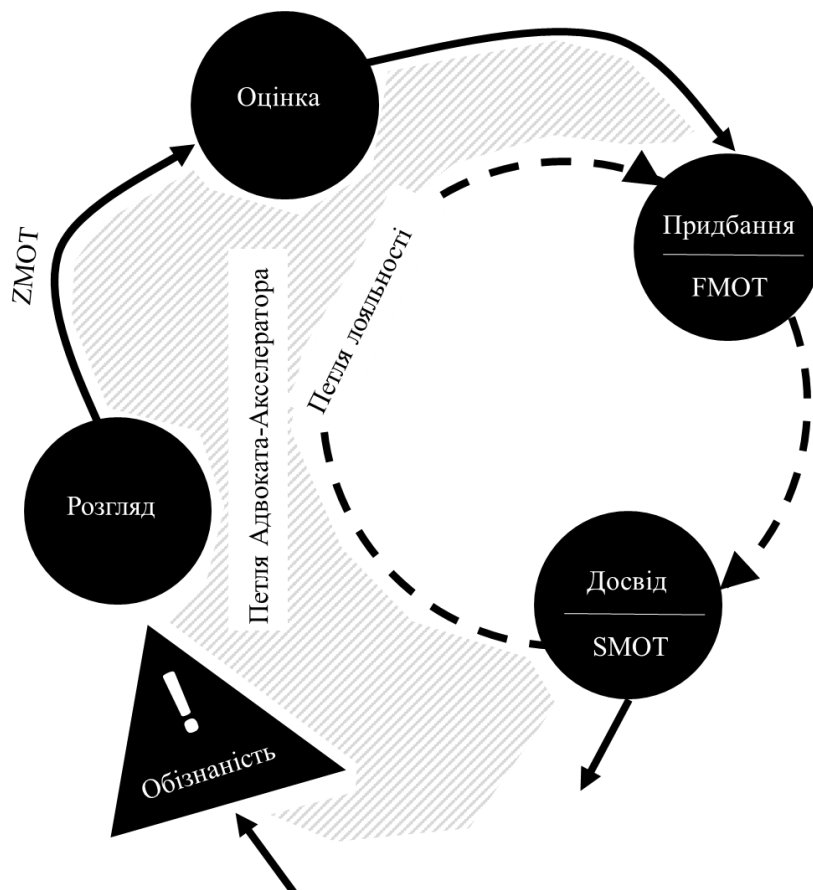


Рис. 1.3 Схema купівельного шляху клієнта та петля адвоката–акселератора

Джерело: [35]

Звідси випливає, що адвокати повинні бути включені в нову маркетингову модель. Асортимент сприйняття та думок, якими вони володіють, суттєво впливає, як на намір зробити покупку, так і на репутацію бренду чи продукту. Незалежно від того, чи це простий «коментар», чи розгорнутий відеопост, весь контент, пов'язаний з брендом, вкорінюється у подорож користувача і стає доступним для наступних клієнтів. Автор називає шар знань та думок, що формуються завдяки високоякісному клієнтському контенту, «петлею адвоката-акселератора» [35].

1.3 Аналітичний огляд існуючих підходів щодо моделювання поведінки клієнтів

Традиційні принципи маркетингових досліджень були повністю витіснені наслідками, що виникли внаслідок змін у соціальному, економічному, технологічному та маркетинговому ландшафтах [160]. Ми живемо в епоху, що характеризується повсюдним поширенням даних, коли всі аспекти нашого оточення нерозривно пов'язані з даними, а кожний аспект нашого існування ретельно задокументований в цифровій формі. Тому існує нагальна потреба в інструментах і процесах управління даними, які здатні своєчасно та ефективно витягувати з даних цінну інформацію або знання [126].

Використання даних для прийняття управлінських рішень стало центральним питанням у сфері бізнесу. Інтеграція та коректний аналіз даних про поведінку клієнтів, разом з подальшим використанням цих даних для прийняття маркетингових рішень, стали важливими стратегічними цілями в розвитку комерційних організацій [163]. Маркетинг існує для того, щоб допомогти організації збільшити клієнтську цінність. Досягнення цієї мети в нинішньому конкурентному бізнес-середовищі, яке орієнтоване на рентабельність інвестицій, перетворилося на завдання, яке добре підходить для використання інтелектуального маркетингу на основі моделювання. Частиною інтелектуального маркетингу є маркетинг баз даних.

Блаттберг Р. К. описує маркетинг баз даних, як управління комп'ютеризованою реляційною базою даних у режимі реального часу. Ця база даних містить вичерпні, актуальні та релевантні дані про клієнтів. Основною метою маркетингу баз даних є виявлення найбільш лояльних клієнтів з метою встановлення якісних, довгострокових відносин. Це досягається шляхом розробки математичних моделей, які полегшують цільову доставку необхідних повідомлень відповідним клієнтам в оптимальний час і в найефективнішому форматі [14]. Переваги використання маркетингу баз даних для моделювання поведінки клієнтів у електронній комерції, також, описує Чен Н. [29].

Фундаментальною концепцією маркетингу на основі баз даних є вивчення даних про клієнтів з метою підвищення їхньої цінності. Більш лаконічне і популярне визначення стверджує, що маркетинг баз даних передбачає використання клієнтських баз даних для підвищення ефективності маркетингу, шляхом ефективного залучення, утримання та розвитку клієнтів. Використовуючи ефективні методи бізнес-аналітики, організації можуть зменшити свої маркетингові витрати, завдяки кращому розумінню вподобань клієнтів та впровадженню стратегій персоналізації [161].

У сучасному бізнесі, який пріоритизує досягнення бажаних результатів, вище керівництво часто ставить фундаментальне питання стосовно окупності маркетингових інвестицій. Маркетинг баз даних має на меті виміряти ефективність маркетингових стратегій і підвищити їхню загальну результативність. Якщо розглядати маркетинг у ширшому контексті, то маркетинг баз даних має на меті виявити ту частину маркетингових зусиль комерційної організації, яка є неефективною або непродуктивною [14].

Науковці виділяють декілька передумов стрімкого росту значення інтелектуального маркетингу:

- **Зниження ефективності масового маркетингу.** Масова реклама та некастомізовані маркетингові активності дають незадовільні результати, супроводжуючись зростанням витрат і зменшенням прибутку.

- **Відповідальність за маркетинг.** Вище керівництво, в гонитві за результатами, вимагає, щоб кожна бізнес-функція, в тому числі і маркетинг, надавала обґрунтування свого існування. Вимога вищого керівництва до результатів, що ґрунтуються на фактах, безпосередньо впливає на фокус маркетингу баз даних, який полягає в аналізі даних і вимірюванні результатів.

- **Посилення уваги до відносин з клієнтами.** Зараз організації приділяють більше уваги відносинам з клієнтами, ніж будь-коли раніше. Вони спостерігають комерціалізацію своїх продуктів і зниження лояльності клієнтів [14].

З огляду на описані передумови, Брінкер С. та Геллер, Дж. запропонували виділяти окрему роль, на яку покладається відповідальність за концептуалізацію

та впровадження технологічних рішень, які відповідають потребам маркетингової діяльності. **Маркетингові технологи** – це люди, які володіють комплексним розумінням, як маркетингу, так і інформаційних технологій. Маркетингові технологи, які досягають успіху, докладають цілеспрямованих зусиль, щоб зрозуміти контекст, в якому працюють технології та їх різноманітне використання. Вони допомагають маркетологам, яким бракує технічних знань, розробляти більш ефективні кампанії, програми та клієнтський досвід [17].

В даний час багато галузей переживають реалізацію своїх прагнень до управління даними. Рітейлери та банки мають явну перевагу в управлінні життєвим циклом клієнта, завдяки великому обсягу цінних даних, які вони мають у своєму розпорядженні. Впровадження технологій моделювання поведінки клієнтів, в даних сферах, може значно підвищити ефективність таких маркетингових ініціатив:

1. Залучення нових клієнтів. Методи моделювання дозволяють кількісно оцінити вплив конкретних рекламних акцій, що дозволяє бізнесам сконцентрувати свої маркетингові витрати на тих інструментах, які приносять найбільшу віддачу [5].

2. Розвиток та утримання існуючих клієнтів. Завдяки моделюванню та аналізу інтересів, поведінкових особливостей та факторів продовження «життя» окремих клієнтів, створюються можливості для додаткових і перехресних продажів, завдяки персоналізованим пропозиціям [106].

3. Збереження лояльності поточних клієнтів і повернення довіри втрачених. Проведення аналізу управління життєвим циклом клієнта може виявити сфери, де існує ймовірність передчасного відтоку клієнтів, що дасть змогу розробити відповідні превентивні заходи. Маркетологи мають можливість розпізнавати та визначати споживчі пріоритети клієнтів, а потім надавати їм окремі рівні лояльності, ексклюзивні заохочення або індивідуальні умови обслуговування [34].

Наразі ринок зазнає значних трансформаційних змін. Серед керівників комерційних організацій спостерігається помітна тенденція, коли вони все

частіше не віддають перевагу частим, масштабним і дорогим дослідницьким ініціативам. Вони все частіше застосовують внутрішній підхід, щоб розвивати аналітичну функцію, як фундаментальну здатність, проводячи свої дослідження безперервно, миттєво, адаптивно та ітеративно.

Інформація в режимі реального часу наразі використовується для отримання більш точного уявлення про споживчі настрої. Це може призвести до прискореного впровадження модифікацій у послуги та продукти, завдяки глибшому розумінню емоційних реакцій на бренди, на відміну від звичайних методів опитування. Для того, щоб досягти оптимальних результатів у відстеженні в режимі реального часу, важливо володіти відповідними технологіями роботи з даними. Маркетологи повинні встановити чіткі межі, щоб забезпечити збір точних даних, які відповідають конкретним маркетинговим цілям кожного бренду [160].

Основна мета даних – сприяти поглибленому розумінню особливостей бізнес-активності комерційної організації та вдосконалення процесів прийняття рішень. Генерування інсайтів на основі даних відіграє вирішальну роль у сприянні процесам прийняття рішень. Дуже важливо визначити найбільш підходящі джерела даних, які можуть ефективно підтримувати ці процеси. Проте, досі велика кількість комерційних організацій не користується значним потенціалом для збільшення прибутку, нехтуючи можливостями, які пропонують великі дані. [34]

Часто, клієнтська аналітика зосереджується на звітах про те, що сталося. Однак для того, щоб діяти на випередження і, по-справжньому, формувати майбутнє бізнесу, важливо передбачити, чого хочуть клієнти, і як вони відреагують на це. Однак стає все більш важливим, щоб не тільки аналіз ефективності бізнесу проводився на основі даних в реальному часі, але й дії у відповідь на результати аналізу могли виконуватися в режимі реального часу і миттєво змінювати параметри бізнес-процесів [105].

Детальне висвітлення значення даних у маркетинговому менеджменті було зроблено у дослідженнях Чжан Ч. та ін. [162], Дам Н., Дінь Т. та Менвілл В. [37].

Загалом, Браун Б., у своєму дослідженні [19] стверджує, що інформація про клієнтів повинна розглядатися як стратегічна. Проте, більшість організацій використовують лише невелику частину даних, що є в їхньому розпорядженні. Вони не усвідомлюють всі можливості застосування даних про активність клієнтів, а тому часто збирають лише ту інформацію, яку вважають цінною на свій розсуд. Зазвичай це, досить, примітивна групова статистика дзвінків, або операцій клієнтів у розрізі регіонів, продуктів чи сервісів. Цю інформацію комерційні організації застосовують для побудови «ручної», часто за доходом або регіоном, сегментації клієнтів; створення нових продуктів та сервісів, які б зацікавили лише певну групу клієнтів; відкриття нових точок продажу або відділень обслуговування у регіонах із найбільшою концентрацією [170].

Частими проблемами є необхідність аналізувати застарілі системи та ізольовані бази даних. Цілком можливо, що моделі та інформаційні звіти ґрунтуватимуться на застарілих даних, а основні процеси потребуватимуть значного ручного втручання. Для того, щоб оптимізувати цінність даних про клієнтів, організації повинні виявити та виокремити критичні закономірності. Детальна інформація про покупки клієнтів, відвідування веб-сайтів, запити до служби підтримки та тривалість перебування на сайті може в сукупності сприяти створенню повної картини споживчих уподобань і моделей покупок. Інтегруючи відповідні дані про клієнтів і забезпечуючи їхню доступність у межах всієї організації, можна зменшити надлишковий збір інформації та ручне введення даних. Крім того, цей процес може призвести до покращення якості обслуговування клієнтів, зниження цін і підвищення якості надання послуг.

Використання застарілих даних для розробки нових моделей також може призвести до помилкових результатів моделювання. Щоб поглибити своє розуміння, бізнес-лідери повинні прийняти комплексну стратегію збору даних. Ця стратегія передбачає не лише збір поведінкових комерційних даних, але й інтеграцію аналітичних даних третіх сторін, в тому числі відкритих даних, що стосуються їхньої організації, клієнтів та конкурентів. Ця інтеграція слугуватиме доповненням до наявних внутрішніх даних про клієнтів. Комерційні організації,

які починають цей процес, відкривають для себе значні переваги, інтегруючи дані про мобільність клієнтів і продажі, отримані від сторонніх постачальників, а також відкриті дані про діяльність конкурентів у свою аналітичну базу [12].

Маркетологи виділяють багато факторів, що зумовлюють обмежений успіх ініціатив з аналітики даних у створенні суттєвих маркетингових переваг:

- відсутність негайного фінансового ефекту є недостатнім обґрунтуванням для виділення додаткових маркетингових витрат [170];
- керівники вищої ланки демонструють дефіцит розуміння та впевненості в аналітиці даних та методах моделювання;
- існуюча IT інфраструктура є недостатньою та обмеженою для ефективного впровадження розробок у сфері аналітики, моделювання та автоматизації;
- процеси маркетингового менеджменту, які вимагають залучення декількох рівнів затвердження, є тривалими і вимагають багато часу;
- може спостерігатися неузгодженість між управлінськими стимулами та пріоритетами бізнесу;
- брак відповідних навичок, необхідних для побудови моделей та аналізу даних.

Складність полягає в тому, що численні перешкоди, які гальмують прогрес ініціатив у сфері великих даних та моделювання поведінки клієнтів, не ґрунтуються на логічних міркуваннях. Вони демонструють емоційні реакції. Основна перешкода в ефективному маркетинговому менеджменті полягає в людській поведінці, незважаючи на складні технічні та адміністративні тонкощі, пов'язані з цим [34]. Традиційні процеси прийняття рішень, які застосовуються в управлінні комерційною організацією, в основному залежать від особистого досвіду, що може бути обмежуючим фактором у нинішній період, який характеризується надлишком інформації. У світлі досягнень сучасних інформаційних технологій існує необхідність прийняти більш інтелектуальний та ефективний підхід до прийняття рішень, наприклад впровадити технології моделювання в процес маркетингового менеджменту [67].

Використання передових аналітичних рекомендацій має потенціал для забезпечення значного впливу на ефективність маркетингового менеджменту, за умови, що вони будуть прийняті та впроваджені представниками відділу продажів та спеціалістами по роботі з клієнтами. Нижче представлено аналіз трьох поширених поведінкових перешкод впровадження підходів інтелектуального маркетингу:

Виклик 1: Уявна складність завдання та відсутність цінності у витрачених зусиллях. Значна кількість представників відділу продажів вважає, що природа сучасних аналітичних інструментів робить їх надмірно складними, що призводить до недостатніх переваг порівняно з часом, витраченим на освоєння їхнього використання. Це означає, що, незважаючи на наявність дуже зручного у використанні інструменту, керівники організацій повинні докладати значних зусиль, щоб переконати менеджерів у тому, що використання аналітики не є складним і є справді вигідним для підвищення ефективності їх роботи.

Гарним прикладом даної проблеми є процес роботи з клієнтами бізнес-сегменту. Найдосвідченіші комерційні організації, що працюють з роздрібними клієнтами, часто, здатні ефективно використовувати великі масиви даних, що є в їхньому розпорядженні. І навпаки, організації, що працюють у форматі «B2B», схильні переважно контролювати дані, а не використовувати їхній потенціал для обґрунтування стратегічного вибору. Ефективна аналітика може допомогти організаціям розпізнати основні фактори, що визначають ціни для різних клієнтських сегментів і продуктів, включаючи аспекти, які часто не беруться до уваги, такі як ширший економічний контекст, товарні переваги та переговори, що проводяться менеджерами з продажу. Впровадження нових стратегій ціноутворення є як комунікаційним, так і операційним викликом. Щоб полегшити розуміння і прийняття нових цінових стратегій, комерційні організації виділяють великі ресурси на реалізацію цілеспрямованих ініціатив змін у своїх відділах продажів. Для комерційних організацій вкрай важливо налагодити тісні робочі стосунки з продавцями, щоб ефективно доносити до них

обґрунтування цінових рекомендацій та забезпечувати всебічне розуміння системи ціноутворення.

Виклик 2: Схильність до самооцінки власної вищості. Численні менеджери вважають, що їхня інтуїція та досвід можуть дати більш точні рішення порівняно з аналітичними методами. Очевидно, що професійно виконане аналітичне рішення здатне запропонувати кращі та доречніші відповіді, порівняно з усіма інтуїтивними пропозиціями. Щоб переконати осіб, що приймають рішення, необхідно продемонструвати, як аналітика може підвищити ефективність їхньої роботи і збільшити їхні фінансові результати. Безумовно, слід визнати, що здійснення особистих оцінок залишається важливим. Використання обґрунтованих суджень і великого досвіду залишається необхідним для підвищення корисності та ефективності рекомендацій, отриманих на основі аналізу даних.

Виклик 3: Існує брак довіри у відносинах. Психологічні побоювання, пов'язані з витісненням людей машинами, є, мабуть, однією з найскладніших проблем для вирішення. Хоча автоматизація успішно усунула певні малоцінні та повторювані завдання, а технології довели свою високу ефективність у наданні рекомендацій клієнтам, залишається очевидним, що люди все ще віддають перевагу спілкуванню з іншими людьми. У сучасному бізнес-середовищі зросла роль менеджерів з продажу та агентів з обслуговування клієнтів, завдяки їхній здатності розуміти потреби клієнтів та орієнтуватися в складних транзакціях, особливо в тих, що стосуються продажу комплексних продуктів [34].

Важливою проблемою, пов'язаною з використанням складних економіко-математичних моделей, чи то в контексті прогнозування, чи то маркетингових досліджень у більш широкому сенсі, є обмежене розуміння керівниками процедур, що лежать в їх основі, та механізмів, за допомогою яких вони синтезують результати. Наприклад, використання електронних таблиць для проведення базового аналізу клієнтської бази, може дати меншу точність порівняно зі складними статистичними методами. Однак перевага полягає у зрозумілості та доступності цих висновків для інших менеджерів в організації.

Необхідність знайти баланс між складністю обраних підходів до аналізу та зрозумілістю презентації своєї роботи широкій аудиторії є постійною проблемою для маркетингових дослідників [160].

Якщо фокусуватися на конкретних кроках забезпечення якісного математичного моделювання поведінки клієнтів, то слід врахувати такі рекомендації:

- Своєчасність стала вирішальною, оскільки внутрішні клієнти (наприклад менеджери з продажу) більше не мають терпіння чекати на докладні звіти, які надаються через тижні або навіть місяці після проведення опитувань. Наразі вони надають перевагу отриманню інформації в короткі терміни, а саме за кілька тижнів або навіть днів, оскільки це збігається з тривалістю їхніх циклів ухвалення рішень.

- Ключовими факторами, які слід враховувати, є зручність, доступність та універсальність. Важливо, щоб інформація була зрозумілою та доступною для пошуку, дозволяючи швидко ідентифікувати та візуалізувати важливі процеси. Крім того, інформація повинна бути легко доступною для різних цілей.

- Креативність відіграє вирішальну роль у сфері маркетингових досліджень, оскільки маркетингові спеціалісти відповідають не лише за поширення знань, а й за сприяння процесам прийняття рішень.

- Інтеграція означає вимогу представляти інформацію у форматі, який можна легко пов'язати з іншими наборами даних, що дозволяє візуалізувати її колективно.

- Забезпечення надійності має вирішальне значення в маркетингових дослідженнях, оскільки передбачає отримання даних, які є дійсно репрезентативними і вільними від проблеми упередженості.

- Дотримання стандартів комплаєнсу має вирішальне значення, зосереджуючись не лише на традиційних для маркетингових досліджень вимогах анонімності, але й на зобов'язаннях виконувати обіцянки, дані клієнтам [160].

Загалом питання аналітики в маркетинговому менеджменті, особливо стосовно економіко-математичного моделювання, отримало всебічний огляд в науковій літературі. Наостанок варто відмітити праці Дам Н., Дінг Т., Менвілл В. [72] та Озава Т., Секігучі Т., Цуда К. [110], в яких надано вичерпну класифікацію кожного компонента маркетингової аналітики, а також пов'язані з ним моделі та підходи до інтелектуального аналізу.

Тепер зосередимося на визначенні сутності процесу моделювання поведінки клієнтів. Каррі Б. Та Моутінью Л. [36] висловлюють думку, що традиційні методи використання комп'ютерів для допомоги у прийнятті рішень, в основному, покладаються на числові розрахунки. Основною функцією комп'ютера в цих вправах є виконання обчислень. Використання інтелекту не є обов'язковою умовою для функціонування комп'ютера. Система призначена для виконання арифметичних операцій, а інтелект міститься або в математичній моделі, що використовується, або в програмах, необхідних для її реалізації. Інтелект, задіяний у цій діяльності, в першу чергу знаходиться в людському компоненті. Точне визначення поняття «інтелект» важко сформулювати. Комп'ютерна модель або програма, яка претендує на те, щоб бути інтелектуальною, повинна володіти всіма або деякими з наступних властивостей: розуміння; міркування та висновки; навчання та адаптація.

Економіко-математичне моделювання передбачає аналіз великих масивів даних у певних галузях, що є дуже трудомістким завданням при ручному виконанні. Для того, щоб сприяти ефективному прийняттю рішень, необхідно, щоб технологія мала здатність передбачати потенційні результати, пов'язані з кожним доступним варіантом, а також визначати оптимальний вибір для конкретного контексту. В цьому контексті корисними будуть алгоритми машинного навчання.

Розробка моделі прийняття рішень на основі машинного навчання спрямована на створення передових і універсальних підходів для вирішення складних проблем навчання і прийняття рішень в ситуаціях, що характеризуються невизначеністю [124].

Машинне навчання охоплює адаптивні методи, які дозволяють комп'ютерам або машинам здобувати знання в процесі навчання на досвіді та прикладах. Використання цих підходів не лише забезпечує значну економію коштів протягом тривалого періоду, але й підвищує зручність для клієнтів. У галузі маркетингу машинне навчання має значний потенціал для застосування в декількох сферах, таких як процеси прийняття рішень, взаємодія з клієнтами та розробка стратегії [100].

Пайл Д. та Хосе К. С. [119] вважають, що машинне навчання ґрунтується на низці попередніх фундаментальних компонентів, починаючи з класичної статистики. Використання статистичних висновків відіграє вирішальну роль у розробці та застосуванні сучасних систем машинного навчання. Однак важливо визнати, що класичні статистичні методи були сформульовані в період з 18-го до початку 20-го століть, коли доступні набори даних були значно меншими за обсягом, порівняно з величезними обсягами даних, які доступні зараз.

На відміну від машинного навчання, яке зосереджується на генеруванні інтригуючих гіпотез, статистика історично надавала пріоритет перевірці гіпотез. Тим не менш, методи машинного навчання зазнали глибокого впливу статистичних принципів. Іноді розмежування між методами машинного навчання і статистичними методами може бути складним. Дослідники машинного навчання надають перевагу терміну «узагальнення» для позначення різних створених інструментів аналізу даних. Ці інструменти застосовуються для таких цілей, як прогнозування, класифікація, кластеризація та виявлення асоціацій. Початківці в галузі машинного навчання часто стикаються з розгубленістю через широкий спектр алгоритмів і методів, які викладаються в таких дисциплінах, як статистичне навчання, інтелектуальний аналіз даних та дата сайнс.

Основна мета **інтелектуального аналізу даних** полягає у використанні зібраних даних, перетворенні їх на змістовну інформацію і, зрештою, в перетворенні її на практичні результати. Дам Н. [37] стверджує, що інтелектуальний потенціал походить від включення експертних знань у процес

аналізу, використання методів аналізу, заснованих на знаннях, а також генерування і поширення нової інформації в процесі аналізу. Всебічний аналіз методів інтелектуального прийняття рішень, їх практичне застосування у сфері виробництва та роздрібної торгівлі подано у дослідженні Гуо З. [61].

Як вже було озвучено, фундаментальна основа цих дисциплін здебільшого походить з теорії ймовірності та статистики. Тим не менш, цілком ймовірно, що деякі статистики заперечуватимуть думку про те, що машинне навчання вдихає нове життя в статистику, а отже, породжує вічні дебати, що нагадують проблему курки та яйця. Розумно стверджувати, що використання статистики в різних галузях потребувало каналу для її безперешкодного поширення, а машинне навчання ефективно забезпечило швидкий і ефективний засіб для виконання цієї вимоги.

Статистика включає в себе наукове дослідження вилучення знань з даних, а також діяльність, пов'язану з вимірюванням, управлінням і передачею невизначеності. Ця сфера досліджень слугує життєво важливим елементом у різних академічних галузях, включаючи соціальні науки, природничі науки, біомедичні науки, фізичні науки, інженерію та бізнес-аналітику. Використання статистичних підходів має першорядне значення, оскільки вони не тільки сприяють науковим відкриттям, але й дозволяють кількісно оцінити притаманну їм надійність, повторюваність і загальну невизначеність, пов'язану з цими висновками.

Дата сайнс – це нова дисципліна, яка є перспективною для вирішення сучасних проблем, пов'язаних з управлінням і аналізом великих масивів даних, які іноді називають «великими даними» [28].

Рамасубраманіан К. та Сінгх А. [121] описують дата сайнс, як комплексну структуру, що охоплює різні дисципліни, які володіють здатністю витягувати цінну інформацію з даних і розробляти інтелектуальні системи. Дата сайнс робить більший акцент на значущості даних на противагу алгоритмам навчання.

Великі дані та наука про дані відкривають широкі можливості для науковців, маркетологів, інженерів та ІТ-підприємств. Крім того, вони

відкривають значні перспективи для математиків і комп'ютерних наук у вивченні нових математичних методів і алгоритмів [28].

Нинішня зміна парадигми призвела до зростання та розробки численних продуктів на основі даних. Дата сайнс створила всеосяжну структуру, яка прагне інтегрувати багато наборів навичок, інструментів і технологій. Дрю Конвей, відомий американський науковець з даних, запропонував відображати сутність дата сайнс за допомогою діаграми Венна (рис. 1.4). На цьому малюнку Конвей точно розмістив машинне навчання на перетині програмних навичок, математичних і статистичних знань, та розуміння предметної області.



Рис. 1.4 Визначення дата сайнс

Джерело: [31]

Однак, кожна дисципліна прийняла унікальні перспективи і підходи, виправдані в її конкретному контексті, що призвело до появи різноманітних систем і сфер дослідження [121].

Відносно нещодавня інтеграція машинного навчання в бізнес процеси, як основного інструменту управління, викликала численні запитання. Нинішній момент є сприятливим для розгляду цих питань, оскільки конкурентна важливість бізнес-моделей, вдосконалених завдяки машинному навчанню, знаходиться на межі значного зростання. Можна стверджувати, що організації,

яким бракує знань з математики або які не здатні розвинути ці знання в найближчому майбутньому, вже вважаються застарілими.

Роль людини переосмислена таким чином, щоб вона перебувала на передньому краї процесу, з особливим акцентом на таких видах діяльності, як аналіз, оцінка, діагностика, планування, прогнозування та моделювання сценаріїв. Люди володіють високим рівнем майстерності у встановленні соціальних зв'язків і виконанні складних завдань, а також демонструють здібності до критичного мислення і розвитку креативності. Машинне навчання має виняткову майстерність у зберіганні та запам'ятовуванні величезних обсягів даних, а також у виконанні дуже складних обчислень на основі цих наборів даних [160].

Постійний розвиток досліджень математичного моделювання, статистичного аналізу та машинного навчання в маркетинговому менеджменті, зокрема і для моделювання поведінки клієнтів, можна підтвердити працями таких іноземних науковців: Ізенман А. Я. [69], Вебб А. Р. та Копсі К. Д. [156], Брей В. А. [16], Інджадат М., Мубайєд А., Насіф А. Б., Шамі А. [68], Шах Н., Інженер С., Бхагат Н., Чаухан Х., Шах М. [131], Москато П., Вріс Н. Дж. [103], Шлерет К., Скіра Б., Шульц Ф. [128], Най В., Лю Л., Ванг С., Донг Д. [104], Грілло Х., Алемани М. М. Е., Ортіс А., Фуєртес-Мікель В. С. [160], Хенкс Н., Сусіло Ю. О., Уллах М. [144], Шин Д.-Х., Кім В.-Й. [139], Гопагоні Д. [58], Лі С. [89], Ванг Н. [94], Гаддам С. [52], Де Мауро А., Сестіно А. та Бакконі А. [38].

Теоретичними дослідженнями та практичним впровадженням технологій машинного навчання у вітчизняній науковій літературі займалися такі вчені, як Дербенцев В. та Матвійчук А. [42], Великоіваненко Г. [150], Негрей М. [70], Юнькова О. [174], Даценко Н. [43], Бабенко В., Хрустальов К. та Обруч Г. [41], Чорноус Г. [30], Гур'янова Л., Яценко Р. Та Дубровіна Н. [62], Чаговець Л., Чаговець В. та Чернова Н. [23].

Машинне навчання – це технологія, яка дозволяє системам самостійно навчатися і вдосконалюватися на основі досвіду, без необхідності явного програмування. Алгоритми машинного навчання можна розділити на чотири

основні категорії, а саме: контрольоване, неконтрольоване, напівконтрольоване та навчання з підкріпленням.

Завдання вибору відповідного алгоритму навчання, який добре підходить для передбачуваного застосування в певній галузі, становить значний виклик. Причина різноманітності алгоритмів навчання полягає в їхніх різних цілях, а також у потенційній розбіжності в результатах роботи алгоритмів в межах однієї категорії, на яку можуть впливати особливості якості даних. Тому дуже важливо розуміти фундаментальні принципи різних алгоритмів машинного навчання та їхню придатність для застосування в різних практичних сферах [126].

Групування за типом проблеми означає класифікацію моделей на основі конкретних цілей, які вони мають на меті досягти. Кожна група демонструє порівняльний набір алгоритмічних методів та принципів. На рис. 1.5 подано схему категоризації машинного навчання та виділено декілька широко використовуваних методів у кожній категорії.



Рис. 1.5 Типологія методів машинного навчання

Джерело: доповнено автором на основі [121]

1. Навчання під контролем (з учителем) – це підхід до машинного навчання, який застосовується при роботі з набором даних, що містить змінну відгуку. Залежна змінна може бути неперервною або категорійною. Алгоритм отримує знання про змінну відгуку, аналізуючи її зв'язок із заданим набором предикторних змінних. Крім того, проблему, що розглядається, можна

класифікувати як задачу класифікації, коли залежна змінна є категорійною, і як задачу регресії, коли вона є неперервною.

2. При роботі з наборами даних, в яких відсутня змінна відгуку, відсутність підказок для змінних-предикторів вимагає застосування методів **неконтрольованого навчання (без учителя)**. Процес навчання залежить від метрики подібності або відстані, яка обчислюється між кожним рядком всередині набору даних. Кластеризація є основною стратегією, що використовується в неконтрольованому навчанні. Неконтрольоване навчання охоплює широкий спектр застосувань, включаючи сегментацію клієнтів у маркетингу, аналіз соціальних мереж або ж сегментацію зображень.

3. **Напівкерowane навчання.** У вищезгаданих двох категоріях набір даних або не має міток для всіх спостережень, або має мітки для всіх спостережень. Напівкерowane навчання стосується аналізу наборів даних, які містять лише частково заповнені змінні відгуку. У багатьох реальних сценаріях витрати, пов'язані з маркуванням, часто є значними через необхідність залучення кваліфікованих фахівців для виконання цього завдання. У ситуаціях, коли більшість спостережень не мають міток, але деякі з них мають їх, найбільш придатними для побудови моделей стають напівконтрольовані методи. Ці підходи використовують концепцію, згідно з якою, незважаючи на невідому групову приналежність немаркованих даних, ці дані містять важливу інформацію про параметри груп.

4. **Навчання з підкріпленням.** Навчання з підкріпленням є підходом, коли вхідні дані складаються виключно з початкового стану даних і немає детермінованої реакції, а є лише безліч потенційних результатів.

Алгоритм навчання з підкріпленням демонструє безперервний процес навчання шляхом ітеративного отримання знань з навколишнього середовища. Під час своєї роботи система набуває знань через взаємодію з навколишнім середовищем, поступово охоплюючи всі потенційні стани в межах свого дослідження.

Алгоритми економіко-математичного моделювання можна класифікувати за схожістю підходу та результатів, які вони дають. Такий підхід полегшує створення сценаріїв використання всередині однієї групи, а отже, дає змогу вирішувати ширший спектр проблемних ситуацій.

Регресійні методи. В академічних та дослідницьких колах методи, засновані на регресії, є найбільш поширеними та широко впровадженими. Їх легко описувати та інтегрувати в реальне бізнес-середовище. Зв'язок між залежною змінною та набором незалежних змінних оцінюється за допомогою ймовірнісного методу або мінімізації функції помилки [121].

Найбільш популярними регресійними алгоритмами є:

- лінійна регресія;
- логістична регресія;
- поліноміальна регресія;
- моделі виживання Кокса та Вейбулла;

Регресійний аналіз – це метод статистичного моделювання, який використовується для дослідження зв'язку між залежною змінною і набором незалежних змінних [161].

Клейнбаум Д. Г. та Клейн М.У [75] підтверджують, що у сфері економіко-математичного моделювання регресійний аналіз використовується для оцінки умовного очікування залежної змінної, коли відомі значення незалежних змінних. Ця оцінка по суті являє собою середнє значення залежної змінної для заданого набору значень незалежних змінних.

Параметричні підходи ґрунтуються на припущенні, що спостережувані дані є похідними від розподілу імовірностей, який є відомим і характеризується заздалегідь визначеним набором характеристик. Наприклад, лінійна регресія ґрунтується на припущенні про нормальний розподіл, тоді як логістична регресія, серед інших припущень, базується на припущенні про біноміальний розподіл. Це припущення дозволяє використовувати ці підходи і на наборах даних меншого розміру.

Непараметричні підходи характеризуються відсутністю припущень щодо базового розподілу імовірностей. Натомість ці методи покладаються на побудову емпіричних розподілів на основі наявних даних.

Аналіз виживання – це набір математичних методів, які використовуються для вивчення даних, де основна увага приділяється тривалості періоду до моменту настання певної події. Термін «час» у цьому контексті означає тривалість у роках, місяцях, тижнях або днях від початку періоду спостереження за об'єктом дослідження до настання певної події. Термін «подія» стосується різних явищ, які можуть вплинути на людину або клієнта, таких як смертність, деактивація, закриття або будь-який інший специфічний досвід, що становить інтерес.

У контексті аналізу виживання прийнято позначати часову змінну як час виживання, оскільки вона відображає тривалість, протягом якої індивід продовжує існувати або «виживає» протягом певного періоду спостереження.

Аналіз виживання часто вимагає розгляду фундаментальної аналітичної проблеми, відомої як цензурування. Цензурування – це явище, коли існує часткове знання про тривалість виживання об'єкта спостереження, хоча точний час виживання залишається невідомим [75].

Модель Кокса є непараметричною регресією, тому що не робить жодних припущень щодо конкретної форми функції небезпек. Модель Вейбулла, на противагу, припускає, що функція небезпеки слідує розподілу Вейбулла. Це встановлює певний зв'язок між змінними і ризиком виникнення події протягом певного періоду часу.

Дослідження та апробація алгоритмів виживання широко висвітлена такими науковцями: Фармейр, Л. та Тутц, Г. [48], Еландт-Джонсон, Р. К. та Джонсон, Н. Л. [47], Харрелл, Ф. Е. [65], Лю, Х. [92].

Недоліком регресійних алгоритмів є те, що застосовувані підходи не встановлюють причинно-наслідкового зв'язку, оскільки загальновизнано, що сильний зв'язок не обов'язково означає причинно-наслідковий зв'язок. Однак ці

методи дають досліднику даних уявлення про взаємозв'язок між рядами даних на основі певних припущень щодо параметрів.

Встановлення причинно-наслідкового зв'язку залежить від обачності та комерційної кмітливості, які застосовуються під час процесу моделювання. Вкрай важливо враховувати концепцію причинно-наслідкового зв'язку, оскільки реальні бізнес-процеси часто розходяться з необхідною інтерпретацією кореляцій, визначених моделлю. Тому дуже важливо володіти всебічним розумінням обговорюваної процедури та експертизою в предметній області. Для того, щоб витягти справжню цінність з даних, необхідно критично проаналізувати і поставити під сумнів основні припущення, що лежать в їх основі [75].

Класифікаційні алгоритми на основі дерев. Ці алгоритми працюють шляхом застосування послідовних умовних принципів до фактичних даних. Як правило, правила застосовуються послідовно, а класифікація визначається після виконання кожної умови.

Найбільш популярними алгоритмами на основі дерева рішень є:

- дерева класифікації та регресії (CART);
- випадковий ліс (Random forest) [121].

На відміну від інших методів машинного навчання, які покладаються на математичні методи, модель дерева рішень вважається непараметричною, оскільки вона не робить жодних базових припущень щодо структури моделі. Необхідно проявляти обережність при визначенні придатності дерева рішень для вирішення певних проблем. Простота інтерпретації та розуміння дерев рішень призвела до їх широкого використання в різних сферах [154].

Дерево класифікації та регресії - це підхід на основі дерева регресії, який використовує суму квадратичного відхилення від середнього значення (залишкову суму квадратів), як міру домішок у вузлах.

Випадковий ліс є однією з найпопулярніших ансамблевих моделей на основі дерева рішень. Точність цих моделей, як правило, вища, ніж у більшості інших дерев рішень. Базова структура моделі дерева рішень полегшує

інтерпретацію, і вона надзвичайно інтуїтивно зрозуміла для застосування до реальних сценаріїв [121]. Приклад процесу роботи алгоритму дерева рішень подано на рис. 1.6.

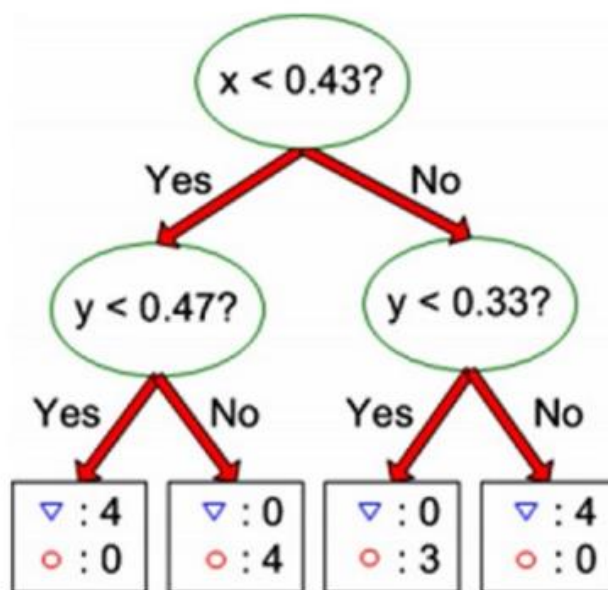


Рис. 1.6 Дерево рішень з двома змінними

Джерело: [28]

Універсальні моделі машинного навчання. Дані алгоритми можна використовувати як для класифікації, так і для вирішення регресійних задач, залежно від набору даних і формулювання проблеми.

Найбільш популярними алгоритмами є:

- k-найближчих сусідів (K-Nearest Neighbors);
- метод опорних векторів (Support Vector Machine);
- лінійний дискримінантний аналіз (Linear Discriminant Analysis).

k-найближчих сусідів працює, порівнюючи відстань екземпляра запиту з іншими навчальними вибірками і вибираючи k найближчих сусідів. Потім більшість з цих k-сусідніх класів використовуються для прогнозування екземпляра запиту [90]. На рис. 1.7 відображено принципи роботи методу k-найближчих сусідів.

Метод опорних векторів (SVM) має на меті визначити найбільш ідеальну гіперплощину, яка ефективно розділяє два різні класи. У контексті двох вимірів гіперплощину можна розуміти як лінію.

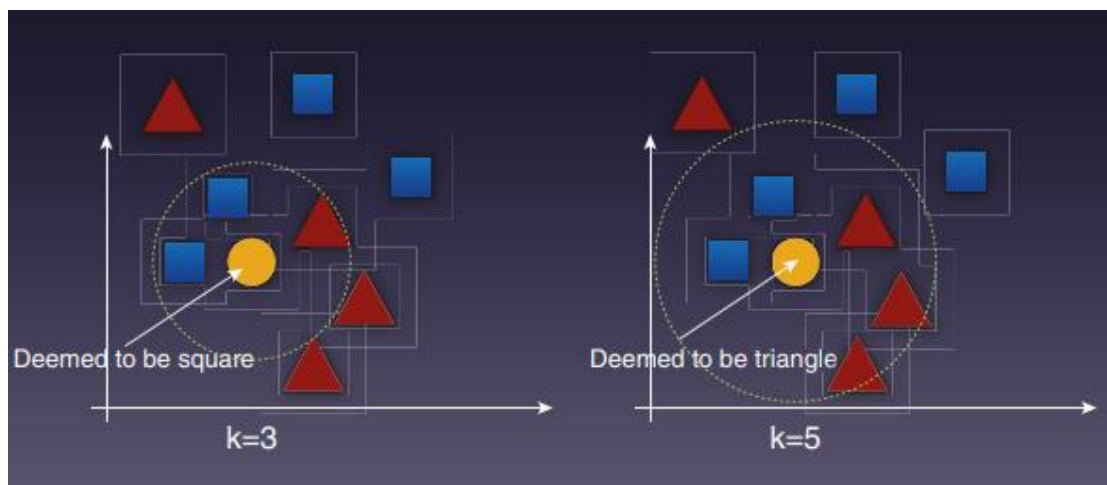


Рис. 1.7 Схематичне зображення алгоритму класифікації KNN

Джерело: [90]

Цей метод полягає в тому, щоб знайти найкращу лінію поділу між двома класами (середню лінію для обох класів). Основна ідея цього методу полягає в тому, щоб отримати лінію поділу на основі деяких граничних точок двох класів, причому існує лише два класи, які потрібно розділити. Однак, для точки довільної форми, яка належить до двох категорій, наявність центральної точки для класу або категорії не має сенсу. У багатьох випадках найбільше сенсу дає використання межі ребер [28]. У контексті двовимірного простору, як показано на рис. 1.8, точки, розташовані на границях, називаються опорними векторами, а пряма, що проходить через середину цих границь, позначається як ідеальна гіперплощина.

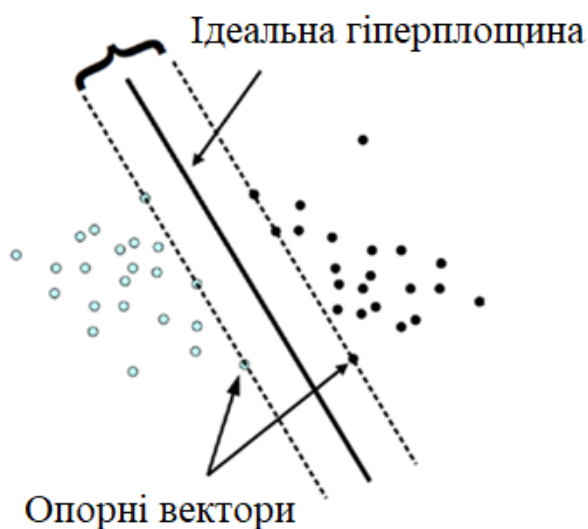


Рис. 1.8 Класифікація методом опорних векторів

Джерело: [28]

Лінійний дискримінантний аналіз (LDA) широко вивчається в статистиці, машинному навчанні та розпізнаванні образів. LDA призначений для пошуку оптимального перетворення для вилучення дискримінантних ознак, які характеризують два або більше класів об'єктів. LDA – це керований алгоритм вилучення ознак, в якому потрібні, як навчальні вибірки, так і відповідні їм мітки класів. Якщо навчальні вибірки кожного класу розподілені нормально, і всі класи мають спільну коваріаційну матрицю, LDA може знайти оптимальне перетворення для вилучення ознак. Оскільки інформація про клас або мітку використовується в процесі навчання, LDA чітко намагається розділити дані між різних класів [164].

Алгоритми кластеризації працюють на основі передумови максимізації внутрішньокластерної схожості при мінімізації міжкластерної схожості. Визначення підходів до формування кластера залежить від міри подібності. Кластерний аналіз – це процес, що полягає в класифікації заданої множини об'єктів на практичні та значущі кластери, в яких об'єкти однорідні між собою в середині сформованого кластеру, і суттєво відрізняються один від одного за своїми характеристиками, якщо порівнювати їх з іншими кластерами.

Для кластеризації існують різноманітні алгоритми. Методи кластеризації можна умовно поділити на дві категорії, як показано на рис. 1.9: ієрархічні та неієрархічні.



Рис. 1.9 Типологія методів кластеризації

Джерело: [14]

Ієрархічна кластеризація шляхом послідовного поділу кластерів або послідовного об'єднання кластерів генерує деревоподібну структуру.

Агломеративні методи – кластеризація починається з n -кластерного рішення для n клієнтів, в якому кожен клієнт представляє свій власний кластер. Об'єднання двох найбільш схожих клієнтів дає $(n - 1)$ кластерне рішення, за яким слідує $(n - 2)$ кластерне рішення і так далі. Цей процес об'єднання триває до тих пір, поки кожен клієнт не буде об'єднаний в єдиний кластер.

Роз'єднуючі методи – кластеризація діє у зворотному напрямку. Процес починається з одного кластерного рішення, що включає всіх n клієнтів. Згодом n клієнтів розбиваються на два кластери таким чином, щоб відрізнити їх один від одного. Цей процес сегментації продовжується до тих пір, поки кожен клієнт не стане кластером.

Останнім часом набули популярності неієрархічні методи кластеризації для знаходження оптимального кластерного рішення, на основі заданої користувачем кількості кластерів [57].

Імовірнісна кластеризація використовує ймовірнісні моделі замість того, щоб покладатися лише на вимірювання відстані для групування. Алгоритм припускає, що дані є результатом комбінації ймовірнісних розподілів, і має на меті визначити оптимальні параметри цих розподілів для точного узгодження кластерів. Присвоєння кластерів є імовірнісним, вказуючи на ймовірність приналежності до них, а не на строгі, остаточні позначення. Дана методологія може розглядатися, як імовірнісна ітерація алгоритму k -середніх, яка долає його численні недоліки в обмін на ряд базових припущень.

Карта самоорганізації є підкатегорією нейромережових моделей, які можуть бути використані для виявлення кластерів [14].

Метод **k -середніх** є найпоширенішим методом неієрархічної кластеризації, де k – це кількість груп, задана апіорі. Метод k -середніх – це метод неконтрольованої класифікації. Маючи набір векторів, якщо метою є кластеризація набору даних на дві категорії, алгоритм дозволяє знайти два

центри так, щоб загальна сума відстаней від кожної точки даних до центру була мінімальною [28].

Підхід k-середніх вважається більш підходящим вибором, ніж агломеративна кластеризація, для задач з великими обсягами даних завдяки його вищій ефективності обробки. Обчислювальна складність алгоритму k-середніх лінійно залежить від кількості спостережень, тоді як агломеративний алгоритм часто демонструє кубічну складність, яка залежить від використовуваної міри схожості. Крім того, k-середніх не потребує значного обсягу пам'яті комп'ютера, оскільки він не передбачає збереження матриць відстаней. Результат роботи алгоритму кластеризації k-середніх залежить від його початкової конфігурації [14].

Одним із потенційних недоліків використання алгоритму k-середніх є можливість отримання викривлених результатів за наявності викидів у наборі даних. Неаномальні одиниці мають тенденцію бути віднесеними до обмеженої кількості груп, в той час як викиди мають тенденцію бути віднесеними до ізольованих і невеликих груп. Це явище може призвести до того, що зазвичай називають «кластерами слонів», тобто кластерами, які є надмірно великими і охоплюють більшість спостережень [55].

Фундаментальним принципом, який об'єднує різні алгоритми кластерного аналізу, є максимізація міжгрупової варіативності при мінімізації внутрішньогрупової варіативності за певною метрикою. Замість того, щоб посилатися на конкретні алгоритми, кластеризація – це процес формування груп відповідно до міри схожості.

Це дозволить отримати уявлення про розподіл даних, які можуть бути використані для подальших алгоритмів або для генерування більш обґрунтованих рішень на основі даних. У машинному навчанні мета не завжди полягає в тому, щоб знайти заздалегідь визначену цільову змінну, скоріше, дослідницький аналіз дає багату інформацію про самі дані [121].

Існує поширена помилкова думка щодо різниці між кластерним аналізом і класифікацією. Обидві стратегії передбачають сегментацію об'єктів на кілька

груп. Однак, Джудічі П. та Фігіні С. [55] наголошують, що вони переслідують різні цілі. У сфері класифікації передбачається, що кількість категорій і приналежність кожного об'єкта до цих груп відома. Мета полягає в тому, щоб спрогнозувати сегментну приналежність нового об'єкта після оцінки класифікаційної моделі. І навпаки, кластеризація не передбачає використання попередньо встановлених сегментів. Основна мета цього завдання – розподілити учасників на окремі кластери на основі їхньої схожості та спільних характеристик. У контексті економіко-математичного моделювання класифікація є звичайною діяльністю, пов'язаною з цілеспрямованим виявленням інформації, тоді як кластеризація є прикладом нецілеспрямованого виявлення знань.

В контексті проблеми класифікації існує залежна змінна. І навпаки, кластеризація не передбачає наявності заздалегідь визначених груп. Метою дослідження є розподіл учасників на змінну кількість однорідних сегментів на основі їхніх характеристик або факторів.

Через притаманну суб'єктивність, пов'язану з визначенням кількості кластерів та вибором методу кластеризації, що супроводжується мірою подібності, кластеризацію іноді розглядають як дослідницьку техніку в рамках аналізу даних. Притаманний кластерному аналізу дослідницький характер часто ставить питання про точність і обґрунтованість отриманої схеми сегментації.

Варто зазначити, що у більшості підходів до кластеризації відсутня статистична модель або теоретична база, яка б стверджувала існування базового групування, що потребує відкриття. Відповідно, це значно ускладнює питання визначення точності отриманого рішення. У більшості підходів до кластеризації відсутній формальний розгляд причин похибки та дисперсії. Тому на результати кластеризації впливатиме наявність викидів або шумових точок [55].

Рамасубраманіан К., Сінгх А. у своїй роботі [121] підкреслюють, що бізнес-інсайти, як правило, випливають з простих і зрозумілих моделей. Рівень складності алгоритму прямо корелює з рівнем складності його пояснення. Якщо побудована модель має характеристики, які не можуть бути ефективно пояснені

бізнесу, комерційна активність втратить свою цінність, оскільки результат моделі не матиме практичного значення. Отже, основна мета економіко-математичного моделювання зводиться нанівець.

Для того, щоб отримати практичне застосування будь-якої розробленої моделі, необхідно розгорнути її в реальному середовищі для використання кінцевими користувачами. У динамічному операційному середовищі включення більшої кількості функцій у модель відповідає збільшенню припливу даних у реальну систему, що потенційно може вимагати доступу до абсолютно нової бази даних. Вищезгадане явище призводить до численних змін і взаємозалежностей в ІТ-системі. Оцінка витрат на впровадження та обслуговування повинна бути ретельно продумана.

Якщо певну поведінку можна пояснити за допомогою меншої кількості функцій та факторів моделі, доцільно реалізувати її за допомогою меншої кількості. Пріоритетність обчислювальної гнучкості та здатності швидко надавати результати моделювання часто переважає над перевагами, які дає більш багата на ознаки модель.

1.4 Концептуальні положення щодо моделювання поведінки клієнтів у системі маркетингового менеджменту

Розуміння та передбачення поведінки клієнтів має важливе значення для існування та подальшого розвитку комерційних організацій у сучасному конкурентному середовищі. Моделювання поведінки клієнтів є важливим компонентом у розумінні та прогнозуванні клієнтської активності. Це дозволяє комерційним організаціям приймати обґрунтовані рішення щодо маркетингової стратегії, розробки продуктів та обслуговування клієнтів.

Моделювання клієнтської поведінки має вирішальне значення для усіх комерційних організацій, які взаємодіють з великою кількістю як індивідуальних, так і бізнес-клієнтів. Метою цього дослідження є створення

універсального концептуального підходу, який може бути використаний для будь-якої галузі, типу клієнта або якості даних.

Хоча створення єдиного підходу до моделювання поведінки клієнтів є надзвичайно цінним, все ж таки в одному науковому дослідженні важко охопити весь спектр бізнес-сценаріїв. Тому у даному науковому дослідженні сфокусовано увагу на трьох ключових сферах економічної діяльності: консалтингових послугах, електронної комерції та банках. Комерційні організації, які працюють в цих сферах, взаємодіють з великою кількістю різнотипних клієнтів, що полегшує розробку універсальних підходів до моделювання їх поведінки. На рис. 1.10 представлена характеристика досліджених у науковій роботі комерційних організацій.

Дослідження клієнтів консалтингових агенцій дає можливість проаналізувати поведінку організацій з точки зору ринкового позиціонування, тобто зрозуміти закономірності їх поведінки при взаємодії з конкурентами. Поведінкою клієнтів в цьому випадку слід вважати: рекламну активність клієнта; впровадження нових або зміна існуючих продуктів та послуг; рост ключових показників ефективності; відкриття нових торгових точок чи представництв; активність з отримання додаткового фінансування. Це в свою чергу дає можливість розробити стратегії маркетингового менеджменту для досягнення фінансового успіху клієнта. Ще однією передумовою вибору консалтингових агенцій, є те що клієнтами цих організацій часто є різнотипні малі підприємства та стартапи, що мають важливе значення для економічного зростання, оскільки вони відіграють вирішальну роль в економічному розвитку, сприяючи інноваціям, конкурентоспроможності, створенню робочих місць та економічній стабільності [142].

Сучасні цифрові трансформації в економіці та суспільстві сприяють зростанню електронної комерції. Після впливу пандемії коронавірусу та повномасштабної війни в Україні електронна комерція стала важливим фактором економічного розвитку, що має потенціал для подальшого стимулювання економіки та сприяння її зростанню. Значення вивчення

закономірностей поведінки клієнтів електронної комерції продовжуватиме зростати [140].

	Консалтингові агенції	Електронна комерція	Банки
Типологія клієнтів	Бізнес клієнти: Будь які організації, включаючи: - Банки - Підприємства електронної комерції - Мікропідприємства / Стартапи	Бізнес клієнти: - Організації, які купують товари для потреб бізнесу - Підприємства, які продають товари на маркетплейсі Індивідуальні клієнти: - Покупці товарів для персонального використання	Бізнес клієнти: - Юридичні особи - Фізичні особи підприємці Індивідуальні клієнти: - Фізичні особи
Що є поведінкою клієнтів	- рекламна активність та комунікації - впровадження нових або зміна існуючих продуктів та послуг - темпи росту ключових показників ефективності - відкриття нових торгових точок чи представництв - активність з отримання додаткового фінансування	- Купівельні особливості клієнтів (кількість, сума, частота покупок) - Споживчі переваги (які категорії товарів купують) - Ідентифікаційні вподобання клієнтів, інтереси, лайфстайл. - Особливості процесу прийняття рішень клієнтом	- Будь яка взаємодія клієнтів з банківськими продуктами та послугами. - Будь яка зовнішня незалежна активність клієнтів, що може впливати на особливості взаємодії з банком та його комерційні результати (фінансовий стан клієнта, співпраця з іншими банками, лайфстайл) - Особливості процесу прийняття рішень клієнтом
Які завдання вирішують за допомогою моделювання поведінки клієнтів	- Підвищення долі ринку - Покращення показників знання бренду - Пошук шляхів росту ключових показників ефективності методом бенчмаркінгу - Пошук нових бізнес-можливостей для досягнення фінансового успіху клієнта	- Підвищення рівня утримання та лояльності клієнтів - Збільшення життєвої цінності клієнта - Оптимізація маркетингових витрат - Нарощення об'ємів продажів - Покращення клієнтського досвіду	- Покращення продуктів та послуг - Підвищення рівня утримання та лояльності клієнтів - Збільшення життєвої цінності клієнта - Оптимізація операційних витрат - Нарощення об'ємів додаткових продажів
Характеристика даних про поведінку клієнта	- Обмеженість даних стосовно комерційної активності клієнтів - Неякісні дані про комерційну активність конкурентів (часто отримані з відкритих джерел) - Недостатній обсяг спостережень стосовно поведінки клієнтів	- Повнота даних стосовно купівельних особливостей та споживчих переваг клієнтів у представників великого бізнесу та обмеженість даних у малого бізнесу - Обмеженість даних стосовно ідентифікаційних вподобань клієнтів (недостатній обсяг спостережень та історичність, погана якість)	- Повнота даних стосовно взаємодії клієнтів з продуктами та послугами банку - Обмеженість даних стосовно активності клієнтів за межами банку (недостатній обсяг спостережень та історичність, погана якість)

Рис. 1.10 Особливості характеристик комерційних організацій розглянутих у дослідженні у контексті моделювання поведінки їх клієнтів

Джерело: розроблено автором

Різноманітна клієнтська база банків, яка включає як роздрібних, так і бізнес-клієнтів, пропонує широкий спектр сценаріїв для моделювання. Враховуючи обидва типи клієнтів, стає можливим побудувати підходи до моделювання, які є дійсно універсальними. Крім того, розуміння поведінки

банківських клієнтів дає суттєві переваги через ключову роль банківського сектору у стимулюванні економічного зростання [147].

У цьому дисертаційному дослідженні розглядаються концептуальні положення моделювання поведінки клієнтів у двох різних контекстах. Перша частина (Розділ 2) – це дослідження процесу моделювання в ситуаціях, які характеризуються дефіцитом та обмеженістю даних, що зумовлює необхідність розробки моделей на основі часткової, розрізненої або низькоякісної інформації.

Друга частина (Розділ 3) заглиблюється в дослідження моделювання поведінки клієнтів в контексті доступу до більш вичерпних даних для навчання моделей. На відміну від підходу з обмеженим обсягом даних, ця частина оперує припущенням про широкі можливості інформаційних технологій, про наявність доступу до великих масивів даних, що уможливорює більш цілеспрямоване та глибоке дослідження поведінки клієнтів.

Концептуальний підхід щодо моделювання поведінки клієнтів в системі маркетингового менеджменту складається з чотирьох етапів: постановка бізнес-задачі маркетингового менеджменту; збір та розвідувальний аналіз наявних даних; формулювання завдань та вибір інструментарію моделювання; комп'ютерна реалізація моделей та їх застосування у маркетингу. Нижче представлено детальний опис кожного з етапів, враховуючи особливості прийняття рішень, які відбуваються на кожному з них (рис. 1.11).

У сфері моделювання клієнтської поведінки, яка постійно розвивається, дуже важливо розуміти стратегії ефективного подолання перешкод, пов'язаних з обмеженою інформацією та неякісними даними. Розділ 2 створює основу для розуміння цих труднощів та пошуку підходів для їх вирішення. Ця конкретна ситуація зазвичай передбачає використання відкритих даних або вже існуючих лімітованих ресурсів, що робить її особливо актуальною для невеликих комерційних організацій, які мають обмежену ІТ-інфраструктуру даних, стартапів, для активностей із запуску нових продуктів та тестування нових маркетингових механік, або для вирішення задач, які не передбачають накопичення великих об'ємів даних.

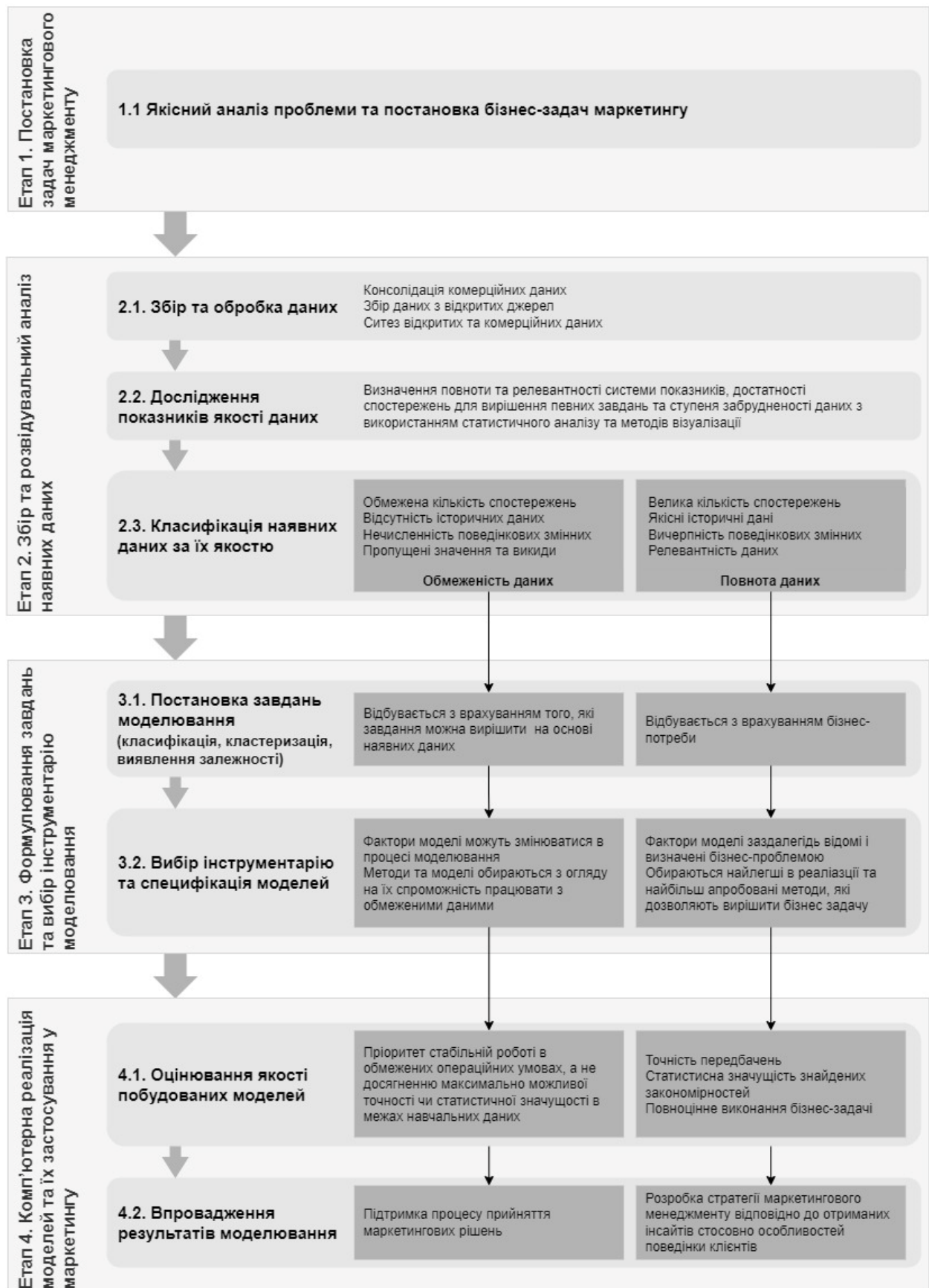


Рис. 1.11 Концептуальна схема запропонованого підходу щодо моделювання поведінки клієнтів у системі маркетингового менеджменту

Джерело: розроблено автором

Першою проблемою, експериментальне дослідження якої описане у Розділі 2, є моделювання поведінки клієнтів у контексті прогнозування поведінки бізнес-клієнтів консалтингової агенції. Прогнозування потенційного успіху нових бізнес-проектів та аналіз факторів, що впливають на цей успіх, має важливе значення для формування маркетингових стратегій та створення конкурентних переваг для підприємств, що зароджуються. Проведення аналізу конкурентів дозволяє виявити ті особливості поведінки, які можуть вплинути на відносний успіх або невдачу бізнесу.

Консалтингові агентства мають можливість адаптувати стратегії маркетингового позиціонування клієнтів у відповідь, підкреслюючи сильні сторони клієнта і водночас усуваючи слабкі сторони конкурентів. Враховуючи ці знання, отримані в результаті моделювання, консалтингові агентства можуть вдосконалити маркетингові стратегії для: підвищення долі ринку; покращення показників знання бренду; росту ключових показників ефективності; досягнення фінансового успіху клієнта.

Тим не менш, моделі у сфері консалтингу здебільшого залежать від відкритих даних стосовно діяльності підприємств, надійність яких може бути сумнівною. Тому в цьому дослідженні використовується підхід щодо моделювання, спеціально розроблена для ситуацій, що характеризуються високим рівнем шуму.

Іншим прикладом слугуватиме експериментальне дослідження щодо моделювання поведінки клієнтів у сфері електронної комерції. Багато малих гравців у сфері електронної комерції, як правило, мають дефіцит комплексної аналітичної інфраструктури, яка підтримує прогнозування покупок. Тому це дослідження присвячене розробці практичних моделей з елементарними даними про транзакції, які можна отримати навіть без значних витрат на інформаційні технології. Завдяки ретельному вивченню фундаментальних поведінкових особливостей і ретельному вибору відповідних рівнів складності моделі можливо розробити цінні системи прогнозування навіть за умови обмеженості наявних даних.

Загалом, Розділ 2 має на меті надати практичні знання, які проливають світло на тонкощі роботи в умовах обмежених параметрів, а також представити комплексний погляд на стратегії оптимізації ефективності моделювання.

Крім того, в цих концептуальних положеннях пропонується застосовувати підхід до оцінки якості моделей в умовах обмеженої кількості даних. Багато загальнозживаних процедур тестування мають тенденцію давати занадто позитивні результати при оцінці ефективності моделей. З огляду на це, запропоновані моделі використовують надійні методи регуляризації, щоб зменшити ризик надмірного навчання, оскільки здатність узагальнювати результати є надзвичайно важливою при роботі з обмеженими наборами даних. Моделі розроблені таким чином, щоб надавати пріоритет стабільній роботі в обмежених операційних умовах, а не досягненню максимально можливої точності в межах навчальних даних.

У Розділ 3 запропоновано експериментальну реалізацію концептуальних положень щодо поведінки клієнтів за наявності вичерпних даних. Наявність якісних даних, що охарактеризовані на етапі 2.3 концептуальних положень за умови повноти даних, зазвичай є результатом значних інвестицій, переважно, великих комерційних організацій з потужною ІТ-інфраструктурою даних в таких сферах як: інформаційні технології, банки, телекомунікації та роздрібна торгівля.

У цьому дослідженні запропоновано розглядати, як банки, що володіють великими об'ємами даних і ресурсів, можуть використовувати технології економетричного моделювання та машинного навчання для підвищення ефективності управління маркетингом. Мета полягає в удосконаленні маркетингових стратегій шляхом розробки моделей, які охоплюють складні фактори, що впливають на лояльність клієнтів і прибутковість у багатогранному середовищі банку.

У цьому дослідженні зосереджено увагу на тому, як перетворити ідеї, отримані на основі широко-апробованих моделей та методів на ефективні маркетингові стратегії. Показником якості математичного моделювання є

статистична значущість знайдених закономірностей та релевантність отриманих результатів бізнес особливостям досліджуваної сфери.

У Розділі 3 реалізовано експериментальне дослідження особливостей моделювання поведінки для двох конкретних категорій клієнтів.

1. Клієнти, які взяли кредит на придбання автомобіля. Незважаючи на високий рівень добробуту, клієнти автокредитування демонструють тривожно низький рівень утримання.

2. Сегмент корпоративних бізнес-клієнтів. Корпоративні клієнти становлять значну комерційну цінність завдяки великим об'ємам депозитів і потенціалу для перехресних продажів. Тим не менш, складна природа предмета дослідження, також, створює проблеми при спробі використання універсальних маркетингових стратегій. У сучасному конкурентному середовищі бізнес-клієнти потребують індивідуального та персоналізованого обслуговування.

Комплексний характер цього дисертаційного дослідження проявляється в тому, що воно демонструє аналіз двох сценаріїв моделювання поведінки клієнтів, а саме практики моделювання в умовах обмеженості даних і моделювання в умовах достатньої кількості та якості даних. Виявлені підходи, тести та найкращі практики пропонують цінну інформацію для практиків при розробці стратегій, незалежно від ступеня їхньої аналітичної зрілості.

Висновки до розділу 1

У першому розділі дисертаційного дослідження було досліджено еволюцію становлення поняття маркетингового менеджменту. У контексті дослідження теорії маркетингового менеджменту визначено, що роль маркетингу в організаціях зазнала трансформації, перетворившись з окремої функції на всеохоплюючу відповідальність, яка поширюється на всіх працівників. Обґрунтовано, що усі дії, що їх здійснює комерційна організація,

підпадають під маркетинг і відіграють вирішальну роль у визначенні загального клієнтського досвіду.

Обґрунтовано використання терміну «клієнт», як більш прийняттого для використання в дослідженні, ніж «споживач», через його кращу асоціацією із встановленням взаємозв'язків, заснованих на придбанні продукту або послуги. Виокремлення основних засад маркетингового менеджменту дозволило виявити, що встановлення відносин з клієнтами є найважливішим аспектом маркетингової ідеї.

Відповідно до проведеного аналізу, було обґрунтовано, що у сучасній бізнес-динаміці отримання прибутку більше не залежить виключно від здатності генерувати дохід через окремі бізнес-транзакції. Натомість, фокус змістився у бік створення життєвої цінності, що отримується від кожного клієнта. Для того, щоб генерувати цінність для своїх акціонерів, комерційна організація повинна не лише надавати цінність своїм клієнтам, але й отримувати цінність у відповідь.

На основі аналізу особливостей класифікації клієнтів доведено значення дослідження поведінки бізнес-клієнтів. Виявлено, що одними з ключових гравців на ринку B2B є банки. Іншими ваговими гравцями на цьому ринку є консалтингові та маркетингові агентства, акселератори та інвестиційні фонди. Визначено, що часто клієнтами даних організацій є стартапи, які активно звертаються до них за фінансовою підтримкою та консалтингом, шукаючи шляхи для свого зростання. З'ясовано, що вивчення їхньої діяльності дозволяє маркетингологам проводити точні та цілеспрямовані маркетингові заходи, вдосконалювати та покращувати продукт.

Узагальнення положень різних підходів до класифікації клієнтів довело, що з розвитком гетерогенності клієнтів і появою більш досконалих тактик таргетування, поведінкова сегментація стала найбільш доцільною для комерційних організацій, які реалізують маркетингові стратегії, орієнтовані на лояльного клієнта. Досліджено, що на процес формування лояльності клієнта можна впливати шляхом розробки поведінкової сегментації клієнтів, вивчення

особливостей поведінки клієнтів в кожному із сегментів та реалізації стратегій перехресних і повторних продажів для кожної із наявних груп.

Розгляд сутності понять «маркетинг баз даних» та «інтелектуальний маркетинг» надав можливість дійти висновків, що роль технологій моделювання в маркетингу полягає у виявленні маркетингових зусиль, які є неефективними або непродуктивними, удосконалити процеси прийняття рішень. Додатково проаналізовано бар'єри, що перешкоджають потенціалу впровадження процесів моделювання в маркетинговий менеджмент, включаючи суб'єктивні судження осіб, що приймають управлінські рішення, суперечливі короткострокові та довгострокові результати впровадження, а також недостатній рівень володіння методологіями моделювання. Проаналізовано тенденції з підвищення ролі ефективного управління витратами в маркетинговому менеджменті.

Обґрунтовано важливість прийняття комплексної стратегії збору даних. Ця стратегія передбачає не лише збір поведінкових комерційних даних, але й інтеграцію аналітичних даних третіх сторін, в тому числі відкритих даних, що стосуються їхньої організації, клієнтів та конкурентів. Доведено важливість якості даних для прогнозування поведінки клієнтів.

Розглянуто найбільш використовувані моделі та методи дослідження поведінки клієнтів. Наголошено на тому, що рівень складності алгоритму прямо корелює з рівнем складності його пояснення. Пріоритетність обчислювальної гнучкості та здатності швидко надавати результати часто переважає над перевагами, які дає більш багата на ознаки модель.

Розроблено концептуальні положення щодо моделювання поведінки клієнтів як в умовах обмеженості даних, так і в умовах якісних даних. Концептуальний підхід складається з чотирьох етапів: постановка бізнес-задачі маркетингового менеджменту; збір та розвідувальний аналіз наявних даних; формулювання завдань та вибір інструментарію моделювання; комп'ютерна реалізація моделей та їх застосування у маркетингу.

Основні результати Розділу 1 опубліковано в наукових працях автора [117, 169, 170, 172].

РОЗДІЛ 2

МОДЕЛЮВАННЯ ПОВЕДІНКИ КЛІЄНТІВ В УМОВАХ ОБМЕЖЕНОСТІ ДАНИХ

2.1 Особливості моделювання поведінки клієнтів в умовах обмеженості даних з використанням методів машинного навчання

У сучасну епоху, яка характеризується прийняттям рішень на основі даних, моделювання поведінки клієнтів набуло великого значення для комерційних організацій, які прагнуть зрозуміти свою цільову аудиторію та вдосконалити методи маркетингу. Тим не менш, можуть виникнути певні обставини, коли даних недостатньо або вони мають низьку якість, що знижує ефективність зусиль з моделювання.

Одним з можливих обмежень дослідження є невеликий розмір вибірки, доступних для аналізу. В окремих випадках, особливо для нещодавно впроваджених продуктів або спеціалізованих ринків, може існувати обмежена кількість спостережень, які можна використати для аналізу. Відсутність достатньої кількості спостережень створює труднощі у виявленні закономірностей, формуванні точних прогнозів та отриманні глибокого розуміння поведінки клієнтів.

Ключовим фактором, який значно впливає на точність і надійність моделей поведінки клієнтів, є якість даних. Різні фактори, такі як наявність пропущених даних, невідповідність форматів даних, або якщо наявні дані містять значну кількість шуму, упереджень та неточностей, можуть мати негативний вплив на процес моделювання.

Наявність даних низької якості не тільки створює перешкоди для початкового аналізу, але й може призвести до помилкових інтерпретацій та неефективного вибору маркетингової стратегії. Тим не менш, організації не завжди можуть відкласти розробку аналітичних рішень до того часу, поки не

будуть доступні значні обсяги якісних даних. У таких випадках необхідно ретельно підходити до вибору методів моделювання, які здатні ефективно працювати з обмеженими даними.

На процес моделювання клієнтської поведінки значною мірою впливають особливості доступних джерел даних. Існує дві основні класифікації джерел даних, а саме: відкриті дані та комерційні дані, які мають свої унікальні характеристики, переваги та недоліки.

Відкриті дані – це термін, що використовується для опису інформації, яка є загальнодоступною і може бути отримана без будь-яких обмежень або витрат. Як правило, джерелами інформації є державні та громадські організації, науково-дослідні інститути або волонтерські ініціативи бізнесу. Розглядаючи використання відкритих даних з метою моделювання поведінки клієнтів, важливо враховувати деякі важливі елементи.

1. Доступність. Відкриті дані характеризуються високим ступенем доступності, оскільки їх може легко отримати будь-хто, не стикаючись з жодними обмеженнями чи перешкодами. Отже, отримання даних не потребує складних процедур або дорогих ліцензійних угод. Така доступність дає змогу організаціям і дослідникам використовувати різноманітні джерела даних, щоб поглибити своє розуміння поведінки клієнтів.

2. Економічна ефективність. Відкриті дані часто доступні без фінансового навантаження або з незначним фінансовим навантаженням, що робить їх привабливим вибором для комерційних організацій, які працюють в умовах обмежених фінансових ресурсів.

3. Різноманітність даних. Джерела відкритих даних часто складаються з різноманітного масиву інформації, що охоплює кілька категорій, таких як соціально-демографічні дані, економічні показники, статистика діджитал активності тощо. Наявність різноманітної інформації може мати значні переваги з точки зору отримання повного уявлення про поведінку клієнтів.

Тим не менш, важливо визнати, що відкриті дані мають певні обмеження та недоліки, коли йдеться про моделювання поведінки клієнтів.

1. Якість даних. Незважаючи на старанні спроби підтримувати відкритість, дані все ще можуть містити неповну або некоректну інформацію. Ретельна перевірка даних і всебічне розуміння процесу збору даних є важливими факторами для усунення можливих упереджень або помилок.

2. Масштаби кастомізації обмежені. Відкриті джерела даних можуть не повністю відповідати специфічним вимогам бізнесу або не відображати особливості поведінки клієнтів. Пристосування моделей до конкретних бізнес-вимог може включати збір додаткових даних або здійснення інтеграційних заходів.

Комерційні дані – це дані, якими володіє та розпоряджається виключно певна організація. Це інформація, яка є ексклюзивною для певного бізнесу, а отже, відкриває особливі перспективи, щодо аналізу демографічних характеристик клієнтів та операційних процесів. Дані часто охоплюють комплексні профілі клієнтів, історію транзакцій та дані про взаємодію з продуктами, опитування, що дає глибоке уявлення про поведінку окремих клієнтів. Розглядаючи можливість використання комерційних даних з метою моделювання клієнтської поведінки, необхідно взяти до уваги багато важливих елементів. Використання комерційних даних для моделювання поведінки клієнтів має кілька переваг:

1. Значна релевантність. Організації мають повну владу над даними, що гарантує їхню цілісність і точність.

2. Створення конкурентних переваг. Використання ексклюзивних даних може сприяти досягненню конкурентних переваг, що, в свою чергу, дає змогу надавати персоналізований клієнтський досвід та здійснювати цілеспрямовані маркетингові активності.

3. Гнучкість. Комерційні організації мають можливість налаштовувати свої методи збору даних, щоб отримати необхідну інформацію про клієнтів, а отже, полегшити розробку точних моделей, які відповідають їхнім власним бізнес-вимогам.

Обмеження використання комерційних даних для моделювання поведінки клієнтів:

1. Фінанси та ресурси. Процес отримання, зберігання та управління даними може накладати значні фінансові обмеження і вимагати значного виділення ресурсів.

2. Конфіденційність та етичність. Дані проблеми є важливими питаннями в контексті моделювання поведінки клієнтів. Використання комерційних даних вимагає суворого дотримання стандартів конфіденційності, а також, зобов'язує сумлінно ставитися до етичних проблем, пов'язаних з використанням інформації про клієнтів.

Всебічне розуміння відмінних рис відкритих і комерційних даних є важливим для того, щоб кваліфіковано моделювати поведінку клієнтів. Вибір джерел даних залежить від конкретних цілей моделювання та наявних ресурсів. Використання відкритих даних може забезпечити широке розуміння та фінансову ефективність, тоді як комерційні дані надають більш точну та детальну інформацію. Інтегруючи обидва джерела, якщо це доцільно, можна отримати більш глибоке розуміння поведінки клієнтів, що дозволить приймати обґрунтовані рішення на основі даних і підвищити рівень задоволеності клієнтів. Наприклад, моделюючи тенденції поведінки конкурентів на основі відкритих даних, можна здобути глибокі знання, які допоможуть скорегувати власні маркетингові стратегії розвитку. В даному дослідженні було зосереджено увагу на моделюванні поведінки бізнес-клієнтів консалтингової агенції – українських стартап підприємств.

Нижче наведено приклади маркетингових сценаріїв, які характеризуються недостатніми або неякісними даними, а отже, створюють труднощі у прогнозуванні поведінки клієнтів.

1. Маркетингові та консалтингові агенції часто обслуговують клієнтів, які не бажають розкривати комерційні дані через занепокоєння щодо потенційного поширення інформації поза їхнім контролем. Неможливість агентств адекватно оцінити та змоделювати поведінку клієнтів ускладнюється обмеженістю власних

даних або даних з відкритих джерел, на які вони змушені покладатися через відсутність якісної інформації від клієнтів.

2. Комерційні організації, які прагнуть вийти на ринки нових регіонів, стикаються зі значними проблемами, пов'язаними з управлінням даними. Обмеженість спостереження за поведінкою клієнтів пояснюється обмеженою доступністю публічних даних, а також притаманними їм особливостям споживчої поведінки.

3. Орієнтування на нову групу клієнтів. Коли маркетингові спеціалісти прагнуть залучити нову групу клієнтів, їхнє розуміння поведінки цієї групи все ще перебуває на ранніх стадіях.

4. Обмеженість бюджетів на тестування є суттєвим ускладненням, оскільки багато стратегій, таких як цінове тестування, вимагають значних витрат при впровадженні. Проведення пілотних проектів з обмеженим обсягом вибірки дає недостатньо даних для точного прогнозування ефектів маркетингових кампаній, які можуть виникнути в результаті повноцінного впровадження.

5. Аналіз мікросегментів, який передбачає моделювання конкретних підмножин клієнтської бази, є проблематичним, зважаючи на кількість спостережень. Хоча кластер, можливо, є візуально привабливим, використання вузьких вибірок може призвести до непередбачуваних і ненадійних прогнозів.

6. Якщо роздрібна мережа відкриває новий магазин, відсутність попередніх транзакцій, пов'язаних з цим конкретним місцем, створює проблему для точного прогнозування місцевого попиту. Завдання прогнозування продажів, визначення потреб у персоналі, управління запасами та вирішення інших аспектів стає більш складним за відсутності даних про клієнтів даного магазину.

7. Організації, які мають слабку ІТ інфраструктуру, стикаються зі значними труднощами в агрегуванні вичерпних даних про клієнтів, що є важливим елементом для ефективного застосування алгоритмів моделювання.

Незалежно від будь-яких обмежень у доступності та якості даних, комерційним організаціям необхідно виділяти ресурси на розробку індивідуальних аналітичних моделей, які відповідають їхнім конкретним бізнес-

потребам. Навіть елементарні моделі, навчені на обмеженому наборі даних, можуть запропонувати цінну інформацію, слугувати орієнтирами та покращити альтернативні гіпотези, що не базуються на моделях. Використовуючи методи роботи з малими обсягами даних, комерційні організації мають можливість отримати значну цінність від процесу моделювання поведінки клієнтів.

Вирішальний аспект полягає в інноваційному використанні аналітичних методів, які найбільше підходять для наявних даних, а також у врахуванні потенціалу для вдосконалення моделювання в майбутньому. Замість того, щоб просто чекати на отримання оптимальних наборів даних, комерційним організаціям необхідно проактивно ініціювати цей процес, формуючи відповідні бізнес-запити, виявляючи цінні закономірності у вже доступних даних, і беручи участь в розробці моделей, щоб сприяти їх постійному вдосконаленню.

Вибір підходів залежить від цілей і допустимих компромісів, пов'язаних з бізнес-задачею. На точність і надійність даних, що використовуються для моделювання поведінки клієнтів, значною мірою впливають методи, які застосовуються при зборі та обробці даних. Труднощі з даними, такі як шум, упередженість та неточності, можуть виникати через недоліки підходів, що використовуються для збору необроблених даних з багатьох джерел, включаючи взаємодію клієнтів, транзакції, опитування та пристрої, що використовуються клієнтами. Систематичні похибки можуть виникати, коли застосовуються нерівномірні процеси збору даних по декількох каналах, наприклад, перегляди інформації в інтернеті та транзакції в магазині. Погіршення якості також спричинене неналежною перевіркою та фільтрацією неточних або аномальних даних.

У процесі обробки даних може виникнути багато проблем через упередженість маркування даних, неналежну інтеграцію даних з різних джерел, неефективну обробку відсутньої інформації або застосування помилкових перетворень даних. Створення та підтримка стандартизованих і добре задокументованих процедур збору та обробки даних про клієнтів є надзвичайно важливими для того, щоб зменшити поширення проблем з якістю, які потенційно

можуть суттєво знизити ефективність і надійність прогнозних моделей. Стратегічні інвестиції на ранніх стадіях розвитку для створення надійних каналів передачі даних і систем управління приносять значні вигоди в майбутньому, особливо при впровадженні складних моделей для аналізу клієнтів.

Непроста задача моделювання в умовах обмеженості даних потребує використання великої кількості математичних та статистичних методів для її вирішення. Для того, щоб зробити процес якомога зрозумілим і прозорим, було розглянуто його у вигляді дев'яти кроків. Виконання цих кроків дозволить створити та реалізувати комплекс моделей, необхідний для досягнення мети даного дослідження. Деталізована схема процесу подана на рис. 2.1.

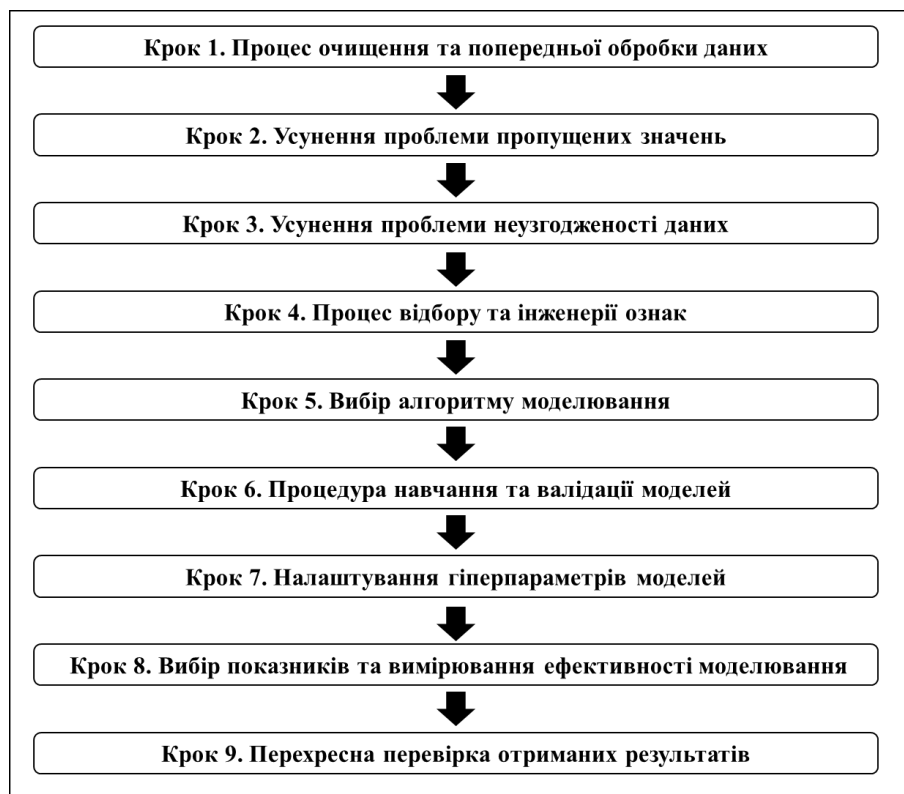


Рис. 2.1 Процес моделювання в умовах обмежених даних

Джерело: розроблено автором

У випадку з неякісними даними необхідні процедури включають очищення даних, фільтрацію, дозаповнення відсутніх значень та інженерію ознак для вилучення цінних закономірностей із неякісних або неповноцінних даних. Інші важливі заходи з підготовки даних включають виявлення та

усунення пропусків, виправлення помилок у даних, нормалізацію ознак та усунення дисбалансу класів.

Далі зосередимося на деталізації підходів щодо моделювання поведінки клієнтів в умовах обмеженості даних.

1. Процес очищення та попередньої обробки даних передбачає виявлення та видалення помилок, невідповідностей та нерелевантної інформації з набору даних, а також перетворення даних у формат, придатний для аналізу. Тобто з даних прибираються усі змінні, які не можуть мати впливу на процес вирішення поставленої проблеми. Також із даних фільтруються спостереження, які мають пропущені значення, що неможливо доповнити.

Наявність викидів може спотворити розподіл даних, а отже, негативно вплинути на точність моделей. Видаляються спостереження викиди, та ті, які не є релевантними для дослідження. Наприклад, якщо є ціль передбачити час до наступного замовлення на основі статистики останніх трьох покупок, то нерелевантними є клієнти, які зробили всього лише 2 покупки. Не підходящими для моделювання поведінки клієнтів можуть бути спостереження, які є застарілими, і які, з експертної точки зору, не відповідають поточним особливостям поведінки клієнтів.

2. Проблема відсутності даних є серйозною проблемою в дослідженнях та аналізі даних. Відсутність даних може виникнути через декілька факторів, таких як відсутність відповідей в опитуваннях, незавершені транзакції або технічні труднощі. Управління відсутніми даними є складним завданням через те, що за відсутності належного підходу вони можуть призвести до упередженості або неточностей в отриманих результатах. Оптимальним підходом до вирішення проблеми відсутніх значень є або усунення, або ігнорування їх при роботі з великим набором даних. У випадку, коли доступний набір даних обмежений за розміром, виникає необхідність розглянути імплікацію ознак, як засіб усунення відсутніх значень шляхом їх заповнення.

Доповнення даних – це процес заповнення відсутніх значень у наборі даних за допомогою різних статистичних методів. В нашому дослідженні було

прийнято рішення про використання одного із основних методів заповнення пропущених значень шляхом усереднення непропущених спостережень [86].

Пропущені категорійні значення пропонується доповнювати шляхом пошуку відповідної інформації з інших джерел, або видаленням спостереження. Також пропущене значення може бути незалежною ознакою, яка характеризує спостереження (пропущені значення заповнюються новою категорією «Відсутнє значення»). В такому випадку алгоритми машинного навчання будуть краще розуміти і оцінювати вплив ознаки на залежну змінну.

3. Ще однією з проблем в аналізі даних є **наявність неузгодженості даних**. Наявність непослідовних форматів даних або одиниць виміру може призвести до помилок у процесі моделювання. Належна **нормалізація характеристик даних** є надзвичайно важливою перед використанням методів кластеризації в галузі машинного навчання. Коли змінні демонструють значні відмінності в одиницях виміру і розподілі, конкретні одиниці виміру можуть мати домінуючий вплив на обчислення відстані, що використовуються для кластеризації порівнянних точок даних. Процедура нормалізації змінних до спільного числового діапазону має на меті запобігти надмірному впливу будь-якої однієї ознаки на процес кластеризації, таким чином забезпечуючи рівний внесок усіх вхідних змінних.

4. Процес відбору та інженерії ознак. Термін інженерія ознак (feature engineering) був введений для того, щоб підкреслити важливість експертизи предметної області, необхідної для вибору наборів ознак для алгоритмів моделювання. Для встановлення значущих зв'язків між змінною відгуку та змінними-предикторами необхідно володіти знаннями про специфічні властивості ознак, а також про їхні взаємозв'язки [121].

Перший етап інженерії ознак для алгоритмів машинного навчання полягає у **створенні цільової змінної (змінної відгуку)** моделі. Ця змінна слугує фундаментальною основою для поточного завдання класифікації. Часто відсутність чіткої та однозначної характеристики зумовлює необхідність

розробки комбінованої змінної шляхом трансформації та консолідації вже існуючих ознак або необхідність отримання її з інших джерел.

Усунення класового дисбалансу є надзвичайно важливим при моделюванні поведінки клієнтів, оскільки це гарантує отримання точних і надійних прогнозів. Коли категорійна змінна має велику кількість класів, які розподілені нерівномірно, модель має проблеми з точною оцінкою її параметрів через обмежену кількість даних, доступних для кожного класу. Як наслідок, це призводить до неоптимальних оцінок моделі та зниження загальної продуктивності. Дану проблему було запропоновано вирішувати за рахунок об'єднання декількох класів схожості в один клас, прибираючи таким чином диспропорцію в кількості спостережень на один клас.

Одним із підходів у створенні змінних є **сегментація клієнтів, і подальше використання даних сегментів у моделюванні поведінки**. Методи кластеризації здатні виявити закономірності та зв'язки, які важко розпізнати з початкових змінних, шляхом виявлення кластерів клієнтів на основі їхніх спільних якостей. Згодом ці спостереження можуть бути використані для створення нових змінних, які охоплюють ці категорії, пропонуючи більш точну характеристику клієнтів. У нашому дослідженні сегментація клієнтських баз на групи схожості була досягнута шляхом кластеризації клієнтів на основі спостережуваної частоти, давності і вартості покупок.

Модель RFM (Recency, Frequency, Monetary) – це широко використовувана стратегія, що застосовується для аналізу цінності клієнтів. На основі емпіричних даних науковці визначили три важливі змінні, пов'язані з купівлею, які впливають на ймовірність майбутніх покупок клієнтів. Трьома основними змінними для узагальнення історії покупок клієнтів є давність (Recency), частота (Frequency) і вартість (Monetary). Використовуючи вимірювання RFM для окремих клієнтів, дозволяє оцінити їхню ймовірність відгуку на маркетингові активності.

Початкова змінна, давність (Recency), означає останній випадок, коли клієнт здійснював транзакцію. Метою цього критерію є вивчення зв'язку між

часовою близькістю події та ймовірністю виникнення реакції. В нашому дослідженні даний критерій визначається, як кількість днів з моменту останнього замовлення.

Друга змінна, позначена як частота (Frequency), кількісно вимірює частоту покупок протягом певного періоду часу. В нашому дослідженні, зважаючи на доступні дані, було зосереджено увагу на покупках за останні 18 місяців. Емпірично виявлено, що клієнти, які частіше купують товари, мають більше шансів здійснити додаткове замовлення.

Вартість (Monetary) для кожного окремого клієнта визначається, як сумарний дохід від усіх транзакцій (вартість замовлень) клієнта протягом певного періоду часу. У нашому випадку показник вимірюється в доларах. Подібно до частоти, існує позитивна кореляція між вартістю та ймовірністю відгуку на маркетингові активності [14].

Неперервні змінні було вирішено **перетворювати в бінарні форми**, щоб полегшити їх розбиття за допомогою класифікаторів і дерев рішень. Перетворення даних у бінарні значення полегшує процес інтерпретації та оцінки впливу на модель.

Наявність мультиколінеарності, що означає високу кореляцію між вхідними змінними, може ускладнити процес оцінки значення окремих змінних. **Консолідація потенційно дублюючих показників в одну змінну** підвищує ясність їхньої значущості. Коли ми стикаємося з двома змінними, кожна з яких створює труднощі у впливі на модель, стає вкрай важливо задіяти креативність для інженерії змінних, щоб об'єднати їхні ефекти. Для досягнення мети покращення здатності до навчання було вирішено здійснити **модифікацію та комбінацію наявних характеристик** для створення нових змінних моделей.

5. Вибір алгоритму моделювання. Методи машинного навчання дуже добре підходять для ефективної роботи з наборами даних, що характеризуються обмеженим розміром вибірки, шумом та багатьма іншими недоліками. На відміну від традиційних економіко-математичних підходів, алгоритми

машинного навчання здатні виявляти складні закономірності та кореляції навіть у випадках, коли наявні дані обмежені за якістю або кількістю.

Цей феномен можна пояснити їхньою здатністю здобувати знання з даних і надавати прогнози, не покладаючись на явне програмування. По суті, алгоритми машинного навчання мають здатність самостійно розпізнавати закономірності в даних, навіть якщо ці закономірності не є очевидними.

Однією з основних переваг машинного навчання є його здатність до екстраполяції з обмеженого набору даних. Це означає, що алгоритми машинного навчання здатні робити точні прогнози на нових, невідомих даних, навіть якщо вони навчаються на дуже обмеженому наборі даних. Цей підхід виявляється дуже корисним для роботи з проблемними даними, враховуючи труднощі та витрати, пов'язані зі збором значних обсягів якісних даних.

Алгоритми машинного навчання також демонструють стійкість до шуму. Це означає, що вони здатні надавати точні прогнози навіть за наявності зашумлених або помилкових даних. Це явище можна пояснити здатністю алгоритмів машинного навчання виводити висновки з колективних закономірностей, присутніх у даних, а отже, зменшувати залежність від окремих точок даних.

Крім того, методи машинного навчання можуть бути використані для вирішення проблеми відсутності даних. Це означає, що комерційні організації мають можливість генерувати прогнози навіть у випадках, коли деякі дані відсутні. Алгоритми машинного навчання здатні здобувати знання з наявних закономірностей в даних і згодом синтезувати відсутні значення, використовуючи ці закономірності.

Під час дослідження буде застосовано 6 моделей класифікації:

- випадковий ліс (RF);
- логістична регресія (LR);
- лінійний дискримінантний аналіз (LDA);
- метод опорних векторів (SVM);
- k – найближчих сусідів (kNN);

- дерево класифікації та регресії (CART).

Також, для сегментації клієнтів, було обрано метод кластеризації *k*-середніх (*k*-means).

Дерево класифікації та регресії є однією з реалізацій дерева рішень. Періодичні вузли дерев класифікації та регресії є коренем, а внутрішні вузли – гілками. Кінцеві вузли є листовими вузлами. Кожен періодичний вузол представляє одну вхідну змінну (*x*) і точку розбиття цієї змінної; листові вузли представляють вихідну змінну (*y*).

CART використовує показник Gini, як критерій для розділення вузлів. Нижче наведено покрокову візуалізацію алгоритму CART:

1. Підрахунок показника Gini. Для заданого набору даних *S* з класами *c* показник Gini обчислюється за формулою:

$$Gini(S) = 1 - \sum_{i=1}^c P_i^2, \quad (2.1)$$

де P_i – частка прикладів класу *i* у множині *S*,

c – кількість класів.

2. Обчислення показника Gini для кожного атрибута, для всіх можливих поділів. Вибір атрибута і розбиття, які дають найменшу зважену суму домішок Gini для періодичних вузлів.

$$Gini_{split}(S, A) = \sum_{v \in Values(A)} \frac{|S_v|}{|S|} * Gini(S_v), \quad (2.2)$$

де *A* – це атрибут,

Values(A) – множина всіх можливих значень атрибута *A*,

S_v – підмножина *S*, для якої атрибут *A* має значення *v*,

$Gini(S_v)$ – це показник Gini для S_v .

3. Створення вузлу дерева рішень для вибраного атрибута і точки розбиття.
4. Поділ даних на підмножини на основі значень обраного атрибута.
5. Повторення процесу рекурсивно для кожної підмножини, поки не буде виконано умову зупинки [54].

Випадковий ліс – це ансамблева модель, яка будує кілька дерев і класифікує об'єкти на основі «голосування». Об'єкт належить до класу, який має

більшість голосів від усіх дерев. Алгоритм навчає кілька дерев рішень на різних наборах даних і використовує середнє значення для підвищення точності прогнозування моделі.

Далі наведено короткий опис роботи алгоритму випадкового лісу:

1. Нехай N = кількість спостережень, n = кількість дерев рішень (задаються користувачем), а M - кількість змінних у наборі даних.
2. Необхідно вибрати підмножину m змінних з M , де $m \ll M$, і побудувати n дерев рішень, використовуючи випадковий набір з m змінних.
3. Далі потрібно «виростити» кожне дерево якомога більшим.
4. На останньому етапі використовується мажоритарне голосування для визначення класу спостереження.

Для побудови кожного дерева рішень використовується випадково вибрана підмножина з N спостережень без заміни [121].

Логістична регресія – це метод моделювання ймовірності події. Як і лінійна регресія, вона допомагає зрозуміти зв'язок між однією або кількома змінними та цільовою змінною, за винятком того, що в цьому випадку цільова змінна є бінарною: її значення дорівнює 0 або 1.

Для реалізації алгоритму першочергово необхідно обчислити лінійну регресію x :

$$x = b + \sum_{i=1}^n w_i * x_i, \quad (2.3)$$

де x – лінійна регресія,

b - похибка,

w_i – вага для ознаки x_i ,

x_i - значення спостереження i .

Далі необхідно застосувати сигмоїдну функцію до лінійної регресії, щоб розбити результати між 0 та 1:

$$\sigma(x) = \frac{1}{1+e^{-x}}, \quad (2.4)$$

де $\sigma(x)$ – сигмоїдна функція.

Вихід сигмоїдної функції представляє ймовірність належності до позитивного класу.

Останнім кроком є визначення функції втрат для вимірювання різниці між прогнозованими ймовірностями та фактичними значеннями (*Loss*):

$$Loss = -\frac{1}{m} \sum_{i=1}^m [y_i * \log(P_i) + (1 - y_i) * \log(1 - P_i)], \quad (2.5)$$

де m – кількість спостережень,

y_i – фактична значення (0 або 1) для спостереження i ,

P_i – прогнозована ймовірність належності до позитивного класу для спостереження i [111].

Лінійний дискримінантний аналіз – це різновид багатовимірної аналізу даних, призначений для вирішення проблем розпізнавання випадкових образів. Він використовується для того, щоб вирішити, які фактори розділяють («дискримінують») певні набори даних (так звані «групи») [164].

Метод опорних векторів – це набір схожих алгоритмів керованого навчання, що використовуються для задач класифікації та регресійного аналізу. Особливістю методу опорних векторів є постійне зменшення емпіричної помилки класифікації та прагнення до збільшення відстані, тому цей метод також відомий як метод класифікації на максимальну відстань [153].

k-найближчих сусідів – передбачає, що об'єкти поділяються на різні класи, щоб їх можна було класифікувати на основі їхньої схожості. Мірою подібності може бути відстань між об'єктами. Кластеризація за методом k-середніх має на меті розділити об'єкти на k кластерів так, щоб точки даних в одному кластері були схожими, а точки даних в різних кластерах були якомога більш віддаленими. Відстань між ними визначає схожість двох точок.

Етапами реалізації алгоритму є:

1. визначення необхідної кількості кластерів - k ;
2. визначення початкового значення центроїда;
3. розрахунок відстані від існуючих об'єктів до центроїда;

4. віднесення кожного об'єкта до відповідного кластеру в залежності від розрахованих відстаней.

Цей метод дуже поширений в економічних дослідженнях, але його практичне застосування для кластеризації має певні труднощі.

По-перше, кінцеві результати чутливі до початкового випадкового вибору центрів груп. Для вирішення цієї проблеми було запропоновано процедуру, що передбачає багаторазове виконання алгоритму з різним випадковим розподілом початкових центроїдів. Ітерація з мінімальним значенням W_{total} була обрана як остаточний варіант кластеризації. Сума квадратів всередині кластера (W_{total}) вимірює середню квадратичну відстань всіх точок всередині кластера до центроїда кластера [117].

$$W_{total} = \sum_{l=1}^k \frac{\rho(C^l)}{n^l}, \quad (2.6)$$

де $\rho(C^l)$ сума евклідових відстаней між точками всередині кластера l ;

n^l — кількість точок у кластері l ;

k — кількість кластерів.

Сума евклідових відстаней між точками всередині кластера l обчислюється за формулою:

$$\rho(C^l) = \sum_{i=1}^n \rho(X^i, C^l), \quad (2.7)$$

де n — кількість точок у кластері l ;

C^l — відсоток ваги кластера l [14].

6. Процедура навчання та валідації моделі вимагає поділу даних на окремі набори для цілей навчання та тестування. Якщо розподіл даних не є репрезентативним, цілком ймовірно, що моделі не зможуть добре узагальнити нові, невідомі дані, що призведе до низької ефективності моделі.

Алгоритми навчання здатні ефективно фіксувати та враховувати певні закономірності, присутні в даних. Однак важливо зазначити, що цей процес іноді може призвести до перенавчання, коли алгоритм надмірно підлаштовується під навчальні дані, не узагальнюючи належним чином нові дані.

Явище перенавчання відноситься до ситуації, в якій статистична модель надмірно адаптована. Алгоритми машинного навчання можуть проявляти надмірну специфічність при обробці великих обсягів поведінкових даних, що, можливо, призводить до труднощів з точною категоризацією окремих точок даних в рамках більш широких категорій. Це означає, що машинному навчанню не вистачає ефективності для розпізнавання певних типів поведінки, що робить процес неефективним [160].

Оцінка моделей на тестових наборах, які не використовувалися в процесі навчання, дозволяє об'єктивно оцінити їхню здатність до узагальнення, оскільки забезпечує неупереджене обчислення помилки моделі.

Зазвичай, науковцями пропонується здійснювати розподіл у пропорції 75% - навчальна, 25% - тестова. Що було і реалізовано в даному дослідженні. Перша множина використовувалася для налаштування параметрів моделі. Друга – для тестування алгоритму на раніше не бачених значеннях.

7. Особливістю алгоритмів машинного навчання є те, що вони базуються на великій кількості **гіперпараметрів**. Ці гіперпараметри мають значення за замовчуванням, які застосовуються, якщо попередньо нічого не змінюється в налаштуваннях алгоритму. Очевидно, що вони не завжди дають найточніші прогнози.

Наприклад, одним із гіперпараметрів в кластеризації k-means є вибір оптимального значення k - фіксованої кількості кластерів для розбиття.

Існує кілька варіантів вирішення цього завдання:

- кількість визначається потребами бізнесу. Цей підхід зазвичай використовується, якщо існує перевірена система класифікації клієнтів у бізнесі. Прикладом може бути розподіл клієнтів за рівнем купівельної активності (низький, нижче середнього, середній, високий);

- кількість підбирається за допомогою алгоритмів машинного навчання. Цей підхід використовується, коли особа, яка приймає рішення, не має уявлення про типологію своїх клієнтів. Алгоритми машинного навчання допомагають виділити класи клієнтів на основі рівня схожості їхньої поведінки;

– поєднання першого та другого підходів. Найпоширеніший підхід - коли рішення приймається на основі, як розуміння бізнесу, так і результатів моделювання [117].

Найбільшим популярним методом вибору кількості кластерів є «Метод середнього силуету» (average silhouette width). Значення силуету показує, наскільки об'єкт схожий на свій кластер в порівнянні з іншими кластерами.

Припустимо, що дані були кластеризовані на k кластерів. Для точки $X^i \in C^l$ (точка X^i знаходиться у кластері C^l) нехай:

$$a(X^i) = \frac{1}{|C^l|-1} \sum_{X^j \in C^l, X^i \neq X^j} \rho(X^i, X^j), \quad (2.8)$$

де $a(X^i)$ – середня відстань від $X^i \in C^l$ до інших об'єктів з кластера C^l ,

$|C^l|$ – кількість об'єктів у кластерів C^l .

Ми можемо інтерпретувати $a(X^i)$ як міру того, наскільки добре X^i присвоєно його кластеру (чим менше значення, тим краще призначення).

Потім визначаємо середню несхожість точки X^i до деякого кластера C^k як середню відстань від X^i до всіх точок C^k (де $C^k \neq C^l$). Для кожної точки даних $X^i \in C^l$, тепер визначимо:

$$b(X^i) = \min_{k \neq l} \frac{1}{|C^k|} \sum_{X^j \in C^k} \rho(X^i, X^j), \quad (2.9)$$

де $b(X^i)$ – найменша середня відстань X^i до всіх точок будь-якого іншого кластеру, де X^i не є членом.

Кластер з цією найменшою середньою різницею вважається "сусіднім кластером" до C^l , оскільки це наступний кластер, що найкраще підходить для точки X^i . Тепер визначимо силует однієї точки даних X^i :

$$s(X^i) = \frac{b(X^i) - a(X^i)}{\max(b(X^i), a(X^i))}, \quad (2.10)$$

де $a(X^i)$ – середня відстань від $X^i \in C^l$ до інших об'єктів з кластера C^l ,

$b(X^i)$ – найменша середня відстань X^i до всіх точок будь-якого іншого кластеру.

Силует варіюється від -1 до +1, де високе значення вказує, що об'єкт добре зіставлений з власним кластером. Якщо більшість об'єктів мають високе значення, тоді конфігурація кластеризації є підходящою

Окрім критерію силуету, існує багато критеріїв, за допомогою яких можна попередньо наближено визначити статистично необхідну кількість груп, перелік яких подано у таблиці 2.1.

Таблиця 2.1

Критерії визначення кількості кластерів в алгоритмі k -середніх

Назва критерію	Позначення	Критерій вибору
Індекс Х'юберта	Hubert	Графічний метод
Індекс Лебарт	Dindex	Графічний метод
Індекс Кшановські та Лай	KL	Максимальне значення індексу
Індекс СН. Калінські та Харабаш	CH	Максимальне значення індексу
Індекс Хартігана.	Hartigan	Максимальна різниця між рівнями ієрархії індексу
Кубічний критерій Сарле	CCC	Максимальне значення індексу
Індекс Скотта	Scott	Максимальна різниця між рівнями ієрархії індексу
Індекс Маріюта	Marriot	Максимальне значення другої різниці між рівнями індексу
Індекс Міллігана та Купера 1	TrCovW	Максимальна різниця між рівнями ієрархії індексу
Індекс Міллігана та Купера 2	TraceW	Максимальне значення абсолютної другої різниці між рівнями індексу
Індекс Фрідмана	Friedman	Максимальна різниця між рівнями ієрархії індексу
Індекс силуету Кауфмана	Silhouette	Максимальне значення індексу
Індекс Ратковського	Ratkowsky	Максимальне значення індексу
Індекс Болла	Ball	Максимальна різниця між рівнями ієрархії індексу
Індекс Міллігана	Ptbiserial	Максимальне значення індексу
Індекс Данна	Dunn	Максимальне значення індексу
Індекс Рубіна	Rubin	Мінімальне значення другої різниці між рівнями індексу
С-індекс Х'юберта	Cindex	Мінімальне значення індексу
DB-індекс Девіса та Боулдіна	DB	Мінімальне значення індексу
Індекс Дуди	Duda	Найменша кількість кластерів, для яких $index > criticalValue$
Індекс Pseudot2 Дуди та Харта	Pseudot2	Найменша кількість кластерів, для яких $index < criticalValue$
Індекс Біла	Beale	кількість кластерів, для яких критичне значення індексу $\geq \alpha$
Індекс Фрея	Frey	рівень кластера до того, як значення індексу < 1.00
Індекс МакКлейна	McClain	Мінімальне значення індексу
Індекс SD Халкіді 1	SDindex	Мінімальне значення індексу
Індекс SDbw Халкіді 2	SDbw	Мінімальне значення індексу

Джерело: [24]

Гіперпараметри для класифікаційних алгоритмів, зазвичай, задаються у певному програмному забезпеченні, яке використовується для калькуляцій. У

нашому дослідженні всі обчислення здійснюються у ПЗ RStudio з використанням мови програмування R.

Гіперпараметрами моделі дерева рішень CART є:

- `cost_complexity` - складність моделі;
- `tree_depth` - максимальна глибина дерева;
- `min_n` - мінімальна кількість точок у вузлі, яка знадобиться вузлу для подальшого розділення.

Гіперпараметри моделі «випадковий ліс»:

- `mtry` - кількість змінних (предикторів), які будуть випадковим чином вибиратися при кожному поділі при створенні моделей;
- `trees` - кількість дерев, що містяться в ансамблі;
- `min_n` - мінімальна кількість точок у вузлі, яка буде необхідна для подальшого поділу.

Гіперпараметрами моделі логістичної регресії є:

- `penalty` - загальна кількість регуляризації в моделі.
- `Mixture` - суміш різних типів регуляризації.

Гіперпараметрами моделі опорних векторів є:

- `Cost` – визначає можливі помилкові класифікації. По суті, це штраф, який накладає модель за помилку: чим вище значення `C`, тим менша ймовірність того, що алгоритм SVM неправильно класифікує точку.

Гіперпараметрами моделі k -найближчих сусідів є:

- `k` – мінімальна кількість спостережень на якій базується рішення про присвоєння класу [118].

Для вибору гіперпараметрів буде використано метод випадкового перебору параметрів. Оскільки існує нескінченна кількість варіантів сполучень гіперпараметрів, потрібен простіший метод їх визначення. Одним з варіантів є випадковий пошук. Наперед визначається певна кількість випадково вибраних значень сполук гіперпараметрів. Далі досліджується модель для кожної сполуки і проводиться перехресна перевірка.

8. У машинному навчанні важливим завданням є визначення **ефективності моделювання**. Тому, коли йдеться про класифікацію, можна покладатися на криву AUC (Area Under the Curve) - ROC (Receiver Operating Characteristics). Це одна з найважливіших метрик для перевірки ефективності будь-якої моделі класифікації. Крива AUC-ROC є мірою ефективності для задач класифікації при різних порогових значеннях. ROC - це крива ймовірності, а AUC – ступінь або міра розподілу. Вона описує, як модель може розрізняти класи. Чим вище AUC, тим краще модель прогнозує 0 як 0 і 1 як 1.

ROC-крива будується як відношення частоти істинних передбачень (TPR) до частоти хибних передбачень (FPR), де TPR відкладається на осі Y, а FPR - на осі X [118]. Візуальний приклад кривої ROC можна побачити на рис. 2.2.

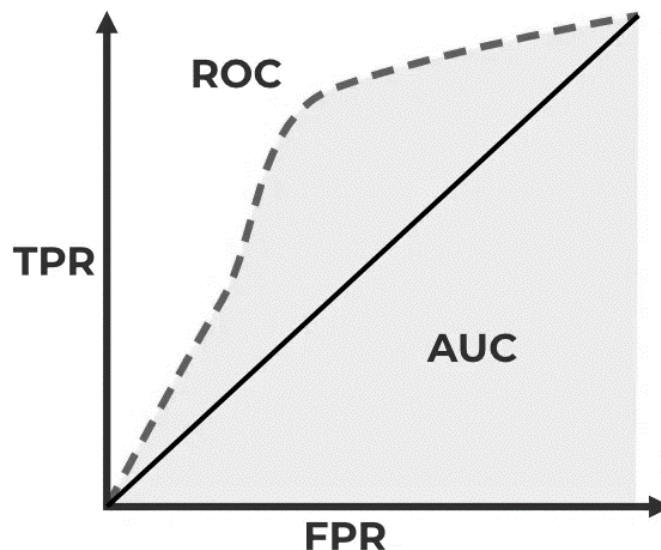


Рис. 2.2 Візуалізація кривої ROC та критерії AUC

Джерело: [121]

Для визначення цільових показників якості TPR/FPR необхідно побудувати «Матрицю невідповідності», яка будується на основі чотирьох показників:

- TP - правильно розподілений позитивний клас;
- TN - правильно розподілений негативний клас;
- FP - неправильно розподілений позитивний клас;
- FN - неправильно розподілений негативний клас.

Матриця невідповідності для моделей, в яких алгоритми намагаються розділити спостереження на два класи, подана в таблиці 2.2.

Більш складною виглядає побудова матриці невідповідності для трьох і більше класів розподілу, що є не типовим завданням для класифікації.

Таблиця 2.2

Матриця невідповідності при двокласовій класифікації

Двокласова класифікація		Реальні значення	
		1	0
Прогнозні значення	1	TP	FP
	0	FN	TN

Джерело: [55]

Позитивним класом (1), в такій ситуації буде вважатися той, перевірка якого відбувається. Негативним – всі інші класи (таблиця 2.3).

Таблиця 2.3

Матриця невідповідності для розподілу на чотири класи

Період наступної покупки		Реальні значення			
		1	0	0	0
Прогнозні значення	1	TP	FP	FP	FP
	0	FN	TN	TN	TN
	0	FN	TN	TN	TN
	0	FN	TN	TN	TN

Джерело: розроблено автором

Показник TPR (True Positive Rate), який має ще декілька назв (Recall /Sensitivity) – це частка правильно розподілених позитивних значень серед усіх позитивних значень і визначається наступним чином:

$$TPR = \frac{TP}{TP+FN}, \quad (2.11)$$

де TP - істинно позитивна класифікація,

FN - хибно негативна класифікація.

FPR (False Positive Rate) – частка неправильно розподілених від'ємних значень до всіх від'ємних значень обчислюється за формулою:

$$FPR = 1 - Spec = \frac{FP}{TN+FP}, \quad (2.12)$$

де TN - істинно негативна класифікація,

FP - хибно позитивна класифікація.

Specificity – відношення правильно розподілених від'ємних значень до всіх від'ємних значень обчислюється за формулою:

$$Spec = \frac{TN}{TN+FP}, \quad (1.13)$$

де TN - істинно негативна класифікація,

FP - хибно позитивна класифікація.

Відмінна модель має AUC близьке до 1, що означає, що вона добре розділяє класи. Погана модель має AUC близьке до 0, що означає найгірший результат розподілу. Це фактично означає, що алгоритм зіставляє зворотні значення. Він класифікує 0 як 1 і 1 як 0. За умови коли AUC дорівнює 0.5, це означає, що модель взагалі не має здатності розділяти класи [118].

Іншим класичним методом оцінки якості класифікаційної моделі, які було обрано для виконання дослідження, є точність (Accuracy), яка розраховується за формулою:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (2.14)$$

де TN - істинно негативна класифікація,

FP - хибно позитивна класифікація,

TP - істинно позитивна класифікація,

FN - хибно негативна класифікація.

Precision - ця метрика стосується кількості правильних позитивних прогнозів. Фактично, показник відповідає на питання: «скільки з тих прогнозів, які мали бути позитивними, насправді виявилися правильними?». Ця метрика визначається як:

$$Precision = \frac{TP}{TP+FP}, \quad (2.15)$$

де TP - істинно позитивна класифікація,

FP - хибно позитивна класифікація.

Останньою метрикою в нашому дослідженні буде зважена оцінка точності (F), яка розраховується за формулою:

$$F = \frac{TP}{TP + \frac{1}{2}(FP + FN)}, \quad (2.16)$$

де FP – хибно позитивна класифікація,

TP - істинно позитивна класифікація,

FN - хибнонегативна класифікація [90].

9. Останнім важливим елементом побудови моделей поведінки клієнтів в умовах обмежених даних є **перехресна перевірка** всіх метрик точності прогнозування. Перехресна перевірка – це статистичний метод, який використовується для оцінки прогностичних можливостей моделей машинного навчання. Процес перехресної перевірки вимагає використання вже існуючого набору даних для виконання окремої операції з метою екстраполяції результатів на непов'язаний набір даних. Перехресна перевірка – це постійна процедура в машинному навчанні, яка передбачає включення кожної нової точки даних у початковий набір даних і подальшу переоцінку інтерпретації результатів на основі нової інформації [160].

Цей підхід ефективно пом'якшує недоліки, пов'язані з використанням єдиного поділу навчання-тестування, шляхом методичного поділу наявних наборів даних на окремі навчальні та валідаційні підмножини. Цей процес розбиття виконується систематично, що дозволяє повторювати моделювання та аналіз. Наприклад, процес 10-кратної перехресної перевірки передбачає поділ набору даних на 10 окремих частин. Згодом виконується 10 ітерацій, в яких кожна складова використовується виключно для перевірки моделі, тоді як решта 9 складових використовуються для навчання моделі. Після цього результати валідації усереднюються [171].

Вона зазвичай використовується в прикладному навчанні для порівняння та вибору моделей для конкретної проблеми прогнозування. Також перехресна перевірка застосовується для підбору гіперпараметрів алгоритмів машинного навчання. У контексті маркетингових активностей запобігання як недостатньої,

так і надмірної відповідності має вирішальне значення для покращення виявлення сильних закономірностей, отриманих на основі даних про клієнтів. Це, в свою чергу, сприяє прийняттю обґрунтованих бізнес-рішень, пропонуючи більш надійний показник справжньої узагальненості моделі, а отже, уможлиблюючи точні прогнози майбутньої поведінки.

2.2 Застосування інтелектуальних технологій до моделювання поведінки бізнес-клієнтів консалтингової агенції

У дослідженні використані відкриті дані з української версії ресурсу Dealroom [39]. Dealroom.co – провідний постачальник даних про стартапи та технологічні екосистеми в Європі та світі. Українська версія дозволяє легко аналізувати дані про стартапи, які базуються або працюють в Україні. Через обмеженість ресурсів, вдалося отримати доступ до інформації лише про 566 підприємств, пов'язаних з Україною.

Під поведінкою бізнес-клієнтів в даній частині наукового дослідження розглядається активність українських підприємств з отримання додаткового фінансування у вигляді інвестицій. Розуміння факторів та закономірностей, що сприяють фінансовому успіху українських стартапів, надасть можливість консалтинговій агенції розробити маркетингові стратегії розвитку, що включають стратегії позиціонування та комунікацій, створення нових продуктів чи послуг, для власних клієнтів.

Згенерований набір даних містить 98 полів характеристик підприємств. Перелік усіх полів представлено на рисунку 2.3. Повний опис кожного поля подано в Додатку А.

Як можна побачити на рис. 2.3, багато полів є неприйнятними для використання в моделюванні. Наприклад, PROFILE URL (посилання на сайт підприємства), FOUNDERS (імена засновників) або LOGO (посилання для завантаження логотипу підприємства). Відповідно, першим кроком буде очищення наших даних від зайвих полів.

NAME	TAGS	FACEBOOK LIKES	FOUNDERS UNIVERSITIES
PROFILE URL	B2B/B2C	TWITTER FOLLOWERS	FOUNDERS COMPANY EXPERIENCE
WEBSITE	REVENUE MODEL	TWITTER TWEETS	FOUNDERS FIRST DEGREE
TAGLINE	LAUNCH DATE	TWITTER FAVORITES	FOUNDERS FIRST DEGREE YEAR
ADDRESS	CLOSING DATE	SW TRAFFIC 6 MONTHS	FOUNDERS LINKEDIN
HQ REGION	INDUSTRIES	SW TRAFFIC 12 MONTHS	FOUNDERS FOUNDED COMPANIES TOTAL FUNDING
HQ COUNTRY	SUB INDUSTRIES	ANGELLIST	LISTS
HQ CITY	DELIVERY METHOD	FACEBOOK	COMPANY STATUS
LATITUDE	GROWTH STAGE	TWITTER	APP DOWNLOADS LATEST (IOS)
LONGITUDE	DEALROOM TAG	LINKEDIN	APP DOWNLOADS 6 MONTHS (IOS)
LOCATIONS	YEARLY GROWTH (SIMILARWEB)	GOOGLE PLAY LINK	APP DOWNLOADS 12 MONTHS (IOS)
FOUNDING LOCATION	ALEXA GROWTH (ALL TIME)	ITUNES LINK	APP DOWNLOADS LATEST (ANDROID)
TEAM (DEALROOM)	EMPLOYEES	EACH ROUND TYPE	APP DOWNLOADS 6 MONTHS (ANDROID)
TEAM (EDITORIAL)	EMPLOYEES (2016,2017,2018,2019,2020,2021)	EACH ROUND AMOUNT	APP DOWNLOADS 12 MONTHS (ANDROID)
INVESTORS	EMPLOYEES IN HQ country (2016,2017,2018,2019,2020)	EACH ROUND CURRENCY	TRAFFIC COUNTRIES
EACH INVESTOR TYPES	LAST KPI DATE	EACH ROUND DATE	TRAFFIC SOURCES
LEAD INVESTORS	PROFIT (2016,2017,2018,2019)	TOTAL ROUNDS NUMBER	SIMILARWEB RANK 3/6/12 MONTHS
TOTAL FUNDING (EUR M)	EBITDA (2016,2017,2018,2019)	EACH ROUND INVESTORS	EMPLOYEE RANK 3/6/12 MONTHS
LAST ROUND	REVENUE (2016,2017,2018,2019)	LOGO	APP RANK 3/6/12 MONTHS
LAST FUNDING	FINANCIALS CURRENCY	FOUNDERS	NUMBER OF ALUMNI EUROPEAN FOUNDERS THAT RAISED > 10M
LAST FUNDING DATE	VALUATION	FOUNDERS STATUSES	TECHNOLOGIES
FIRST FUNDING DATE	VALUATION CURRENCY	FOUNDERS GENDERS	INCOME STREAMS
SEED YEAR	VALUATION (EUR)	FOUNDERS IS SERIAL	TECH STACK DATA (BY PREDICTLEADS)
OWNERSHIPS	VALUATION DATE	FOUNDERS BACKGROUNDS	TRADE REGISTER NUMBER
SDGS	CORE SIDE VALUE		

Рис. 2.3 Перелік усіх полів у наборі даних

Джерело: побудовано автором на основі [39]

Решта факторів, після виконання процесу очищення, показані на рисунку 2.4.

NAME	B2B/B2C	FACEBOOK LIKES	TWITTER FOLLOWERS
REVENUE MODEL	TWITTER TWEETS	LAUNCH DATE	TWITTER FAVORITES
SW TRAFFIC 6 MONTHS	SW TRAFFIC 12 MONTHS	EACH ROUND TYPE	APP DOWNLOADS LATEST (IOS)
HQ REGION	INDUSTRIES	EACH ROUND AMOUNT	APP DOWNLOADS 6 MONTHS (IOS)
HQ COUNTRY	SUB INDUSTRIES	EACH ROUND CURRENCY	APP DOWNLOADS 12 MONTHS (IOS)
HQ CITY	DELIVERY METHOD	EACH ROUND DATE	APP DOWNLOADS LATEST (ANDROID)
TOTAL FUNDING (EUR M)	GROWTH STAGE	TOTAL ROUNDS NUMBER	APP DOWNLOADS 6 MONTHS (ANDROID)
LAST ROUND	YEARLY GROWTH (SIMILARWEB)	PROFIT (2016,2017,2018,2019)	APP DOWNLOADS 12 MONTHS (ANDROID)
LAST FUNDING	ALEXA GROWTH (ALL TIME)	EBITDA (2016,2017,2018,2019)	SIMILARWEB RANK 3/6/12 MONTHS
LAST FUNDING DATE	EMPLOYEES	REVENUE (2016,2017,2018,2019)	APP RANK 3/6/12 MONTHS
FIRST FUNDING DATE	EMPLOYEES (2016,2017,2018,2019,2020,2021)	VALUATION	TECHNOLOGIES
SEED YEAR	EMPLOYEES IN HQ country (2016,2017,2018,2019,2020)	VALUATION CURRENCY	INCOME STREAMS

Рис. 2.4 Доступні поля після очищення даних

Джерело: побудовано автором на основі [39]

Ще однією проблемою цього набору даних є те, що в ньому відсутній критерій того, чи намагалося підприємство отримати фінансування. Відповідно, немає достовірної інформації про те, що підприємство, яке не залучило зовнішніх інвестицій, є неуспішним. Ймовірно, бізнес-клієнт не потребував фінансової підтримки інвесторів (рис. 2.5), розвиваючись за рахунок власних джерел фінансування.

Тому, щоб уникнути неточностей, підприємства, для яких інформація про фінансування не була відображена, були вилучені з масиву даних. Таким чином, у нашому масиві залишилося 358 клієнтів.

Наступна проблема полягає в тому, що дані про інтернет-активність (SW TRAFFIC 12 MONTHS, APP DOWNLOADS 12 MONTHS (IOS) та TWITTER

FAVORITES) доступні або на поточну дату, або за минулий рік. Відповідно, неможливо застосувати ці фактори для підприємств, які мали останній раунд до 2019 року (за винятком ситуації, що це був єдиний раунд фінансування).

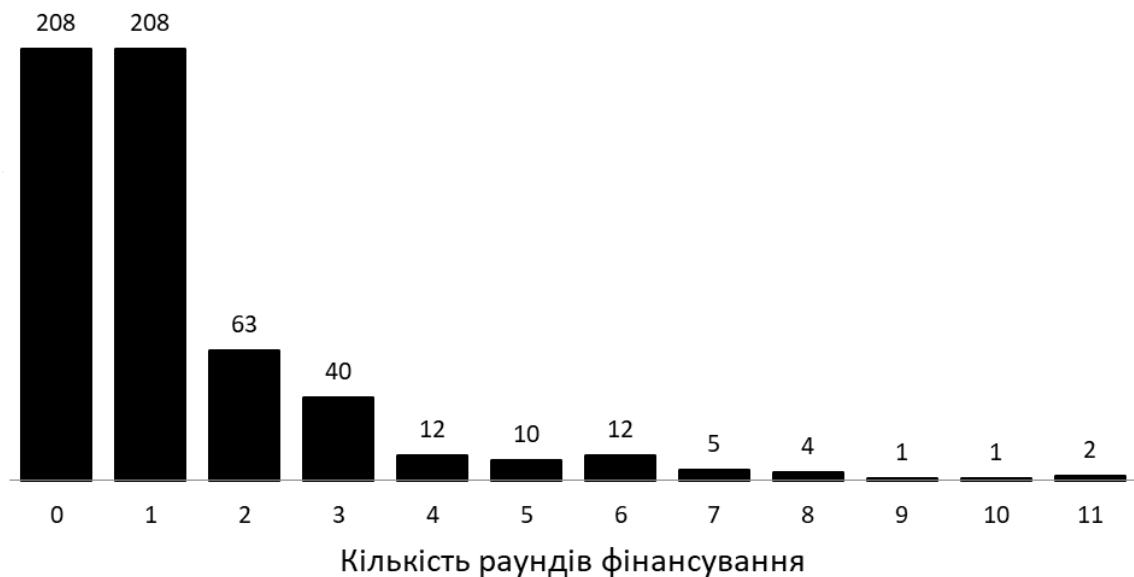


Рис. 2.5 Розподіл підприємств за кількістю раундів фінансування

Джерело: побудовано автором

Для достовірності висновків вибірка була обмежена підприємствами, які отримали перше фінансування в період з 2017 по 2019 рік. Після цих маніпуляцій залишилося лише 149 клієнтів. Розподіл за типом останнього раунду фінансування представлений у таблиці 2.4.

Таблиця 2.4

Розподіл підприємств за типом останнього фінансування

Тип останнього раунду	Кількість підприємств
SEED	82
ANGEL	18
ACQUISITION	13
EARLY VC	10
GRANT	10
SERIES A	5
ICO	4
SERIES C	2
LATE VC	2
CONVERTIBLE	1
DEBT	1
GROWTH EQUITY	1

Джерело: побудовано автором

У таблиці 2.4 лише 14 бізнес-клієнтів підпадають під класичне визначення успішності - ACQUISITION or GROWTH EQUITY, чого недостатньо для побудови моделі [1]. Враховуючи це, було прийнято рішення обмежити дослідження лише інвестиційним успіхом, критерієм якого було обрано отримання повторного фінансування (Success). Оскільки частина інвестиційних угод є конфіденційною інформацією і не для кожного підприємства відома сума раніше залучених інвестицій, зростання фінансування не може бути використано як критерій успішності. Фактором, який також може вплинути на достовірність отриманих результатів, є наявність спостережень, які вже в першому раунді отримали фінансування A+ або здійснили вихід. Невідомо, що вплинуло на рішення інвесторів у цій ситуації, тому дані про такі підприємства також були вилучені з нашого масиву даних.

В результаті залишилося 123 спостереження, які були трансформовані бінарним методом (1 – фінансовий успіх (факт повторного інвестування), 0 – інші підприємства (відсутність факту повторного фінансування). Кількість та співвідношення успішних та неуспішних клієнтів можна побачити на рисунку 2.6.

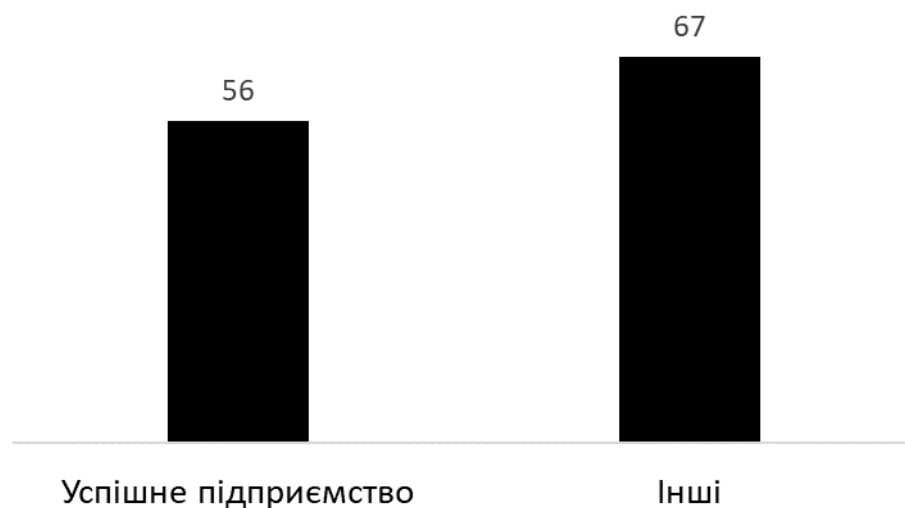


Рис. 2.6 Кількість успішних та неуспішних підприємств у наборі даних

Джерело: побудовано автором

Наступним кроком було створення нових змінних з існуючих. Так, наприклад, нова змінна, **Час до першої інвестиції** (Time_to_first_funding), була

створена шляхом віднімання Дати запуску стартапу (LAUNCH DATE) та року першої інвестиції з поля EACH ROUND DATE (Дата кожного інвестиційного раунду) .

Для створення змінної **Тип першої інвестиції** (First_round_type) перше значення було взято з поля EACH ROUND TYPE (Тип кожного інвестиційного раунду). З поля EACH ROUND AMOUNT (Сума кожного інвестиційного раунду) було виділено **Суму першої інвестиції** - First_round_amount.

Для **інтегральної оцінки обсягу трафіку на сайт** (SW_traffic) за основу було взято значення кількості відвідувачів сайту за останні 12 місяців (SW TRAFFIC 12 MONTHS). Для оцінки **кількості завантажень додатку клієнтів** (APP_downloads) було обчислено суму кількості завантажень додатку за 12 місяців IOS та Android (APP DOWNLOADS 12 MONTHS (IOS) + APP DOWNLOADS 12 MONTHS (ANDROID)).

Присутність клієнтів в топ-рейтингах Similar Web оцінювалася шляхом створення нової змінної TOP_Rank_SW, в якій під значенням "TOP" вказувалася наявність підприємства в будь-якому рейтингу SIMILAR WEB RANK 3/6/12 MONTHS або APP RANK 3/6/12 MONTHS.

Наступним кроком буде перевірка пропусків факторів. Якщо значення поля відсутнє для більшості спостережень – його було видалено. Якщо відсутня лише частина – значення доповнювалося вручну. Якісна інформація була отримана з сайту підприємства та відкритих джерел, а кількісна інформація заповнювалася методом усереднення непропущених спостережень. Для відсутніх значень, на основі заповнених, було визначено середнє значення для кожного співвідношення B2B_B2C (Тип взаємодії з клієнтами), INDUSTRIES (Ринок), First_round_type, INCOME STREAMS (Тип надходжень), REVENUE_MODEL (Модель доходу) та TECHNOLOGIES (Технології).

Оскільки алгоритм побудови моделі машинного навчання передбачає поділ даних на тестову та навчальну вибірки (25% / 75%), а навчальної вибірки на 5 рівних груп для перехресної перевірки, потрібно, щоб кількість кожного з категорійних значень була більшою за 6 (по 5 на кожен перехресну перевірку та

по одному на навчальну вибірку). Тому необхідно згрупувати значення параметрів моделі відповідно до заданого порогового значення.

Результати групування наведено нижче:

- **HQ_COUNTRY** – перетворено на бінарну змінну. Характеризує наявність головного офісу в Україні або в іншій країні
- **TIME_TO_FIRST_FUNDING** – тепер характеризує час до першої інвестиції в розрізі трьох груп: отримання фінансування в рік відкриття, в наступний рік, більше ніж через 1 рік.
- **FIRST_ROUND_TYPE** – змінна згрупована у дві категорії SEED_EARLY_VC та ANGEL_GRANT.
- **INDUSTRIES** – групування в 4 категорії: Heals_Food_Education, Enterprise_Software_Manufacturing, Creative_Services, Home_Travel_Hobbies.
- **TECHNOLOGIES** – групування в 4 категорії: App_programs_development, New_product_development, Online_services, та VR_AR_ML_AI.

Опис полів представлено в таблиці 2.5.

Таблиця 2.5

Характеристики змінних моделі

Назва змінної	Опис
SUCCESS	Залежна змінна, що характеризує фактор фінансового успіху (Так / Ні)
HQ_COUNTRY	Країна розташування головного офісу
TIME_TO_FIRST_FUNDING	Час до першої інвестиції
B2B_B2C	Тип взаємодії з клієнтами (бізнес або фізична особа)
INDUSTRIES	Сфера діяльності
YEARLY_GROWTH_SIMILARWEB	Щорічний приріст цифрової активності в рейтингу SW
SW_TRAFFIC	Кількість відвідувань сайту (середнє за рік)
FIRST_ROUND_TYPE	Тип першого раунду фінансування
FIRST_ROUND_AMOUNT	Сума першого раунду фінансування
APP_DOWNLOADS	Середня кількість завантажень додатку за останній рік
TOP_RANK_SW	Присутність у топ-рейтингу за цифровою активністю у своїй сфері діяльності
INCOME_STREAMS	Канал доходу
REVENUE_MODEL	Модель доходу
TECHNOLOGIES	Які технології/інновації були запропоновані

Джерело: побудовано автором

Детальний опис кожної категорійної змінної подано у Додатку Б. В результаті отримано 14 полів, які характеризують стартап.

Останнім кроком буде перетворення всіх категорійних змінних в бінарні. Це означає, що за наявності x рівнів факторів буде створено $x-1$ фіктивних змінних, а всі фактори, крім першого, будуть перетворені в нові стовпчики. Всі числові фактори були нормалізовані.

Основним програмним пакетом для налаштування алгоритмів машинного навчання було обрано пакет Tidymodels мови програмування R.

Як уже зазначалося, першим етапом, спільним для всіх моделей, є поділ початкової вибірки на тестову та навчальну. До навчальної вибірки було включено 93 клієнти, до тестової – 30.

Після реалізації алгоритму CART значення Ассигасу було встановлено на рівні 0.633, Roc_Auc – 0.665. Оцінювання проводилося на основі тестової вибірки. Цей рівень означає, що модель дозволяє передбачити успіх підприємства з більшою ймовірністю, ніж це було б зроблено випадковим чином.

Але однієї ітерації оцінки недостатньо, щоб з повною впевненістю стверджувати про ефективність моделі. Тому для остаточних результатів необхідно було провести перехресну перевірку ефективності моделі. Результати перехресної перевірки представлені в таблиці 2.6.

Таблиця 2.6

Результати перехресної перевірки дерева рішень CART

Метрика	Min	Median	Mean	Max
Roc_Auc	0.444	0.581	0.550	0.644
Accuracy	0.444	0.632	0.590	0.722
Sensitivity	0.333	0.625	0.572	0.778
Specificity	0.300	0.500	0.602	0.909
F	0.462	0.500	0.553	0.667

Джерело: побудовано автором

Після перехресної перевірки були отримані нижчі результати ефективності моделі. Особливе занепокоєння викликають мінімальні значення для Roc_Auc та Ассигасу. Вони є меншими за порогове значення для ефективної моделі – 0.5.

Для підвищення ефективності прогнозування було проведено процедуру налаштування гіперпараметрів моделі. Тому для кожної моделі були підібрані оптимальні гіперпараметри. І для кожного гіперпараметру була проведена 5-кратна перехресна перевірка, щоб знайти їх найоптимальніше співвідношення.

У таблиці 2.7 наведено гіперпараметри, які забезпечать моделі найбільші мінімальні значення для Roc_Auc та Accuracy.

Таблиця 2.7

Оптимальні значення гіперпараметрів

Гіперпараметр	Значення за замовчуванням	Оптимальне значення
cost_complexity	0.01	9.45078958391999e-08
tree_depth	30	14
min_n	20	33

Джерело: побудовано автором

У таблиці 2.8 наведено результати прогнозування з урахуванням нових гіперпараметрів. Налаштування гіперпараметрів моделі CART дозволило збільшити мінімальні значення точності прогнозування за метриками Roc_Auc та Accuracy до 0.512 та 0.5 відповідно.

Таблиця 2.8

Результати перехресної перевірки дерева рішень CART після налаштування гіперпараметрів

Метрика	Min	Median	Mean	Max
Roc_Auc	0.512	0.581	0.576	0.644
Accuracy	0.500	0.632	0.612	0.722
Sensitivity	0.125	0.625	0.547	0.875
Specificity	0.300	0.800	0.662	0.909
F	0.182	0.636	0.523	0.667

Джерело: побудовано автором

На рисунку 2.7 показано криву ROC після реалізації алгоритму дерева класифікації та регресії. На рис. 2.7 ROC-крива знаходиться вище пунктирної лінії, що означає, що модель є більш ефективною, ніж випадковий розподіл на класи. Після застосування алгоритму логістичної регресії Accuracy була встановлена на рівні 0.467, Roc_Auc – 0.473. Оцінювання проводилося на основі тестової вибірки. Результати перехресної перевірки логістичної регресії представлені в таблиці 2.9.

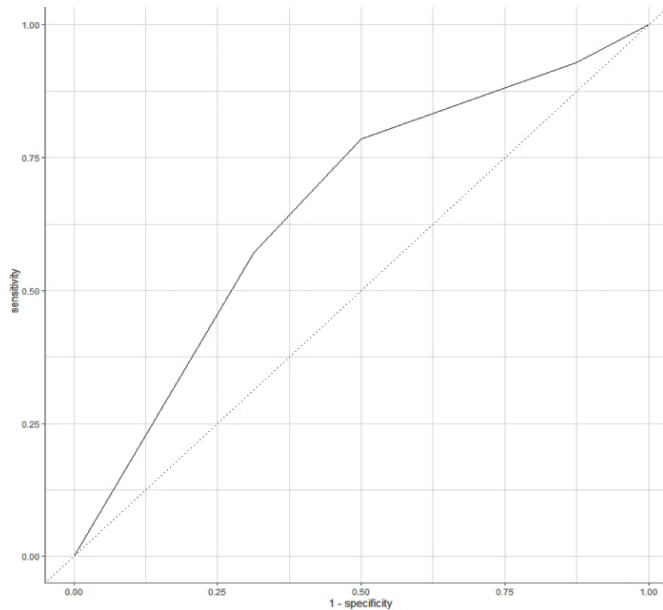


Рис. 2.7 Крива ROC для дерева класифікації та регресії

Джерело: побудовано автором

Як видно з таблиці 2.9, логістична регресійна модель показала неприйнятні результати для того, щоб стверджувати, що ця модель є ефективною – вона не пододала порогових значень 0.5.

Таблиця 2.9

Результати перехресної перевірки логістичної регресії

Metric	Min	Median	Mean	Max
Roc_Auc	0.462	0.495	0.544	0.689
Accuracy	0.389	0.550	0.504	0.579
Sensitivity	0.333	0.375	0.450	0.667
Specificity	0.300	0.500	0.545	0.727
F	0.375	0.421	0.445	0.600

Джерело: побудовано автором

Наступним кроком було покращення результатів прогнозування шляхом налаштування гіперпараметрів моделі. У таблиці 2.10 наведено гіперпараметри, які забезпечать моделі найбільші мінімальні значення для Roc_Auc та Accuracy.

Таблиця 2.10

Оптимальні значення гіперпараметрів

Гіперпараметр	Значення за замовчуванням	Оптимальне значення
penalty	0.1	0.0621016941891562
mixture	1	1

Джерело: побудовано автором

У таблиці 2.11 наведено результати прогнозування з урахуванням нових гіперпараметрів.

Таблиця 2.11

Результати перехресної перевірки логістичної регресії після налаштування

Метрика	Min	Median	Mean	Max
Roc_Auc	0.575	0.606	0.645	0.733
Accuracy	0.500	0.611	0.600	0.684
Sensitivity	0.125	0.500	0.450	0.625
Specificity	0.500	0.800	0.724	0.818
F	0.182	0.533	0.486	0.625

Джерело: побудовано автором

У таблиці 2.11 наведено результати прогнозування з урахуванням нових гіперпараметрів: Roc_Auc - 0.575 та Accuracy - 0.5. На рисунку 2.8 показано криву ROC після реалізації алгоритму логістичної регресії.

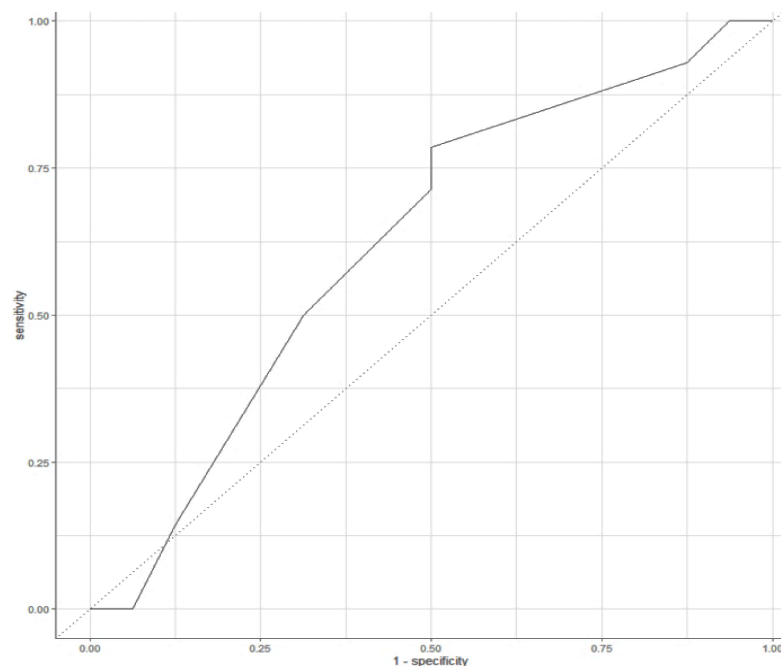


Рис. 2.8 Крива ROC для логістичної регресії

Джерело: побудовано автором

На рисунку 2.8 ROC-крива знаходиться вище пунктирної лінії, що означає, що модель є більш ефективною, ніж випадковий розподіл на класи.

Після реалізації алгоритму Random Forest значення Accuracy було встановлено на рівні 0.467, Roc_Auc – 0.5. Оцінювання проводилося на тестовій вибірці. Результати перехресної перевірки методу випадковий ліс наведено в таблиці 2.12.

Таблиця 2.12

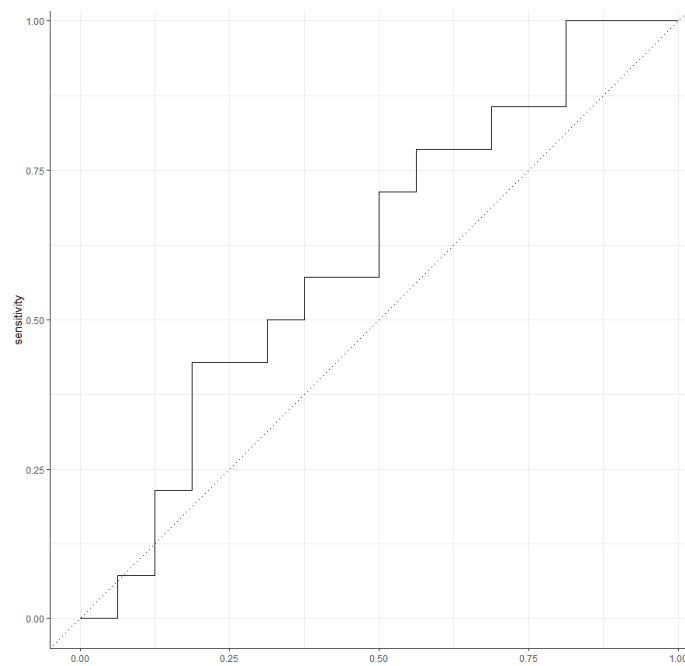
Результати перехресної перевірки випадкового лісу

Metric	Min	Median	Mean	Max
Roc_Auc	0.562	0.626	0.645	0.800
Accuracy	0.500	0.556	0.570	0.684
Sensitivity	0.111	0.500	0.358	0.556
Specificity	0.5	0.800	0.744	0.900
F	0.167	0.471	0.399	0.625

Джерело: побудовано автором

Як бачимо, мінімальні значення параметрів Roc_Auc та Accuracy після перехресної перевірки моделі випадкового лісу були встановлені на рівні більше 0.5, що свідчить про прийнятність моделі для прийняття рішень. Середні значення були встановлені на рівні 0.645 та 0.57 відповідно.

Процес оптимізації гіперпараметрів не дав суттєвих результатів, тому висновки були зроблені на основі їхніх значень за замовчуванням. На рисунку 2.9 зображено криву ROC після реалізації алгоритму «випадковий ліс».

**Рис. 2.9 Крива ROC для випадкового лісу**

Джерело: побудовано автором

На рис. 2.9 крива ROC знаходиться вище пунктирної лінії, що означає, що модель є більш ефективною, ніж випадковий розподіл на класи. Для порівняння моделей було розраховано середні значення характеристик ефективності прогнозування (табл. 2.13).

Таблиця 2.13

Порівняльна таблиця ефективності моделей

Метрика	CART	Logistic Regression	Random Forest
Roc_Auc	0.576	0.645	0.645
Accuracy	0.612	0.600	0.570
Sensitivity	0.547	0.450	0.358
Specificity	0.662	0.724	0.744
F	0.523	0.486	0.399

Джерело: побудовано автором

Оскільки для всіх моделей найвищі показники точності спостерігаються за показником Specificity, це означає, що моделі краще розпізнають неуспішні підприємства, ніж фінансовий успіх. Найбільш описовими показниками є Sensitivity та Accuracy, оскільки вони орієнтовані на вимірювання точності прогнозування успіху підприємства. Тому, навіть якщо зважений показник точності Roc_Auc в логістичній регресії (0.645) та випадковому лісі (0.645) буде вищим, але за рахунок більш точного передбачення негативних факторів, дерево класифікації та регресії буде вважатися найбільш ефективною моделлю (Sensitivity – 0.547, Accuracy – 0.612).

В результаті проведеної роботи розкрито підходи до очищення та підготовки даних для моделювання методами машинного навчання. Вперше описано алгоритм роботи з відсутніми даними шляхом ручного додавання інформації з відкритих джерел та методом заміни пропущених спостережень на усереднені значення непропущених спостережень.

Лістинг R-коду для реалізації всіх етапів моделювання в підрозділі 2.2 подано в додатку В.

2.3 Побудова системи моделей поведінки клієнтів в електронній комерції на підґрунті методів машинного навчання

На сучасному етапі розвитку економіки України інтернет-технології проникли майже у всі сфери життя суспільства, одним з проявів чого стала широка популярність онлайн-торгівлі. Все частіше клієнти користуються

послугами інтернет-магазинів, тому що це має безліч переваг: зручність вибору, нижча ціна, економія часу на покупку тощо. Водночас, завдяки розвитку цифрових технологій боротьба за клієнтів переміщується у глобальний простір. У найближчий час, на додачу до жорсткої внутрішньої конкуренції, очікується поява на українському ринку нових міжнародних гравців. Це ставить перед вітчизняними інтернет-магазинами нові завдання, щодо покращення умов реалізації продукції. Відповідно для українського ритейлу зростає важливість інновацій, які сприяють підвищенню ефективності ведення бізнесу онлайн-продавцями. Інноваційний розвиток даних комерційних організацій стає основним фактором їх конкурентоспроможності.

У більшості онлайн-ритейлерів значну долю витрат становлять видатки на маркетингову діяльність, а проблема їх оптимізації стала серйозним викликом для всіх гравців ринку. Розрізнення клієнтів, для яких очікується висока активність у майбутньому, і клієнтів, для яких прогнозується пасивна поведінка, та застосування індивідуального підходу має суттєво зменшити маркетингові витрати та підвищити їх ефективність.

Таким чином, моделювання активності клієнтів є актуальною проблемою, розв'язання якої дасть значний поштовх до росту конкурентоспроможності українських інтернет-торговців. Зокрема, важливою задачею при побудові ефективної маркетингової стратегії є прогнозування часу до наступного замовлення.

Застосування методів машинного навчання у прогнозуванні часу до наступного замовлення клієнтів інтернет-магазину, може бути досить ефективним підходом, щодо аналізу очікуваної активності клієнтів та управлінні маркетинговою діяльністю інтернет-магазину.

Для вирішення поставленої бізнес-задачі та для урахування особливостей наявних первинних даних було запропоновано реалізацію прогнозування поведінки клієнтів інтернет магазину у два етапи. На першому етапі алгоритму, при проведенні кластеризації, аналізується клієнтська база та здійснюється поділ клієнтів на певну кількість кластерів. На другому етапі за допомоги методів

машинного навчання модель тренують автоматично прогнозувати час до наступної покупки на основі даних, щодо минулої активності клієнта та з урахуванням кластеру, до якого його віднесено [117, 171].

У дослідженні використано дані стосовно продажів інтернет магазину [27], які містять інформацію, щодо 1 067 371 транзакцій купівлі товарів в інтернет магазині за період з 1.09.2009 по 9.12.2011. Приклад початкової вибірки подано в таблиці 2.14.

Вхідні дані містять такі поля:

- Invoice – унікальний код операції;
- StockCode – унікальний код товару;
- Description – назва товару;
- Quantity – кількість купленою/поверненої продукції;
- InvoiceDate – дата здійснення операції;
- Price – ціна товару;
- Customer ID – унікальний код покупця;
- Country – країна ініціатор операції.

Таблиця 2.14

Перші 5 рядків вибірки з транзакціями

Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	12/1/2010	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010	3.39	17850	United Kingdom

Джерело: [27]

Для прогнозування часу до наступного замовлення початкову вибірку було поділено на «минулу» і «майбутню» активність. За «минулу» активність клієнта

вважався період з 01.12.2009 по 31.05.2011 (18 місяців). Прогнозувалась активність клієнтів у період з 01.06.2011 по 09.12.2011 (6 місяців).

Першим завданням, яке необхідно вирішити в процесі дослідження, є вибір критеріїв для оцінки рівня купівельної активності клієнтів. Для вирішення цієї проблеми було обрано класичний підхід до вимірювання купівельної активності – RMF-модель (Recency - Frequency - Monetary).

Отже, спочатку ці показники купівельної активності були розраховані для кожного покупця у вибірці. Після цього були розраховані характеристики статистичних розподілів Recency, Frequency, Monetary, а саме: середнє, мінімальне та максимальне значення показників, а також 1, 2 та 3 квантили. Значення цих характеристик наведено в таблиці 2.15.

Таблиця 2.15

Розподіл показників купівельної активності

	Min	1 st Q	Median	Mean	3 rd Q	Max
Recency	0.0	25.85	98.1	202.86	381.04	738.08
Frequency	1.0	1.00	3.0	6.27	7.00	336.00
Monetary	2.9	336.10	848.8	2720.50	2211.70	608821.70

Джерело: побудовано автором

Як бачимо, середньостатистичний клієнт обраного інтернет-магазину зробив останню покупку 202 дні тому. В середньому клієнти замовляли 6 товарів протягом останніх 18 місяців, витрачаючи при цьому 2 720 доларів.

Вибірка для перевірки була обрана розміром у півроку не випадково. Як вже відзначалося клієнт в середньому робить одну покупку один раз у 202 дні. Тому, з огляду на наявний об'єм інформації, нам було необхідно мати хоча б 180 днів для перевірки, щоб прогноз виявився реалістичним. У «минулій» активності було зафіксовано 4 518 клієнтів.

Наступник кроком у даному дослідженні буде реалізація кластеризації клієнтів інтернет-магазину відповідно рівня їхньої купівельної активності, базуючись на щойно-обчислених показниках RFM.

Як вже зазначалося в пункті 2.1 алгоритм k-means, який був обраний для реалізації кластеризації, вимагає попереднього визначення кількості груп схожості, на який буде розподілятися наша вибірка даних. Отже, наступним

кроком буде обчислення критерію середнього силуету. Використовуючи формули (2.8) – (2.10), розрахуємо значення «силуету» для кожного варіанту кількості кластерів (від 1 до 10). Графічне представлення результатів розрахунків наведено на рис. 2.10.

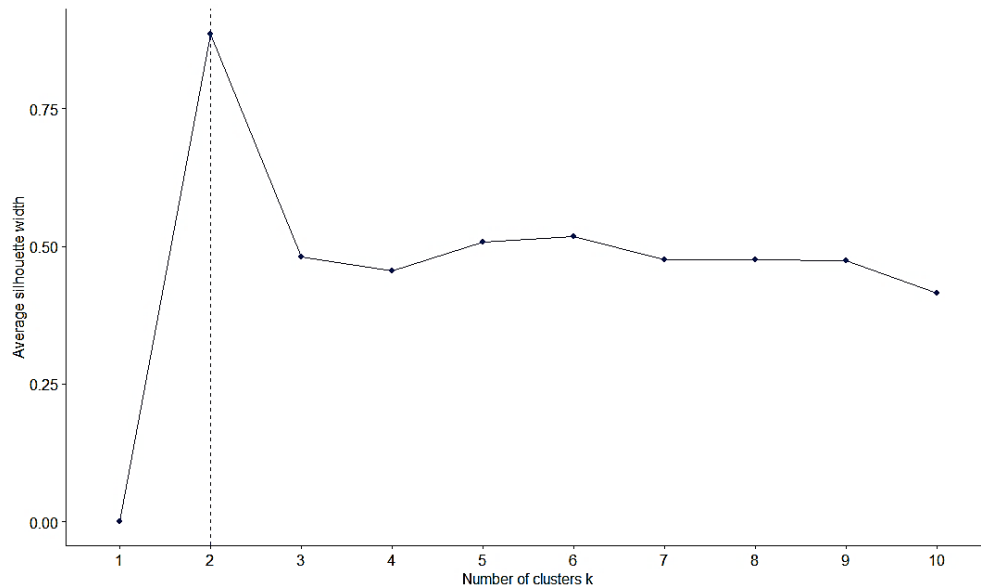


Рис. 2.10 Графічна реалізація методу «середнього силуету»

Джерело: побудовано автором

За цим методом ми шукаємо найбільше значення цього показника. Як бачимо на рис. 2.10, оптимальна кількість кластерів – 2 (на другому місці - 6). Варіант «2» було одразу відкинуто, оскільки він не принесе ніякої цінності в подальших розрахунках.

Тому основними варіантами залишаються 6 кластерів. Наступним кроком буде практична реалізація методу k-середніх та перевірка результатів на бізнес-логіці. Клієнтів було розподілено на 6 кластерів. У таблиці 2.16 наведено середні значення метрик RFM для шести кластерів.

Таблиця 2.16

Результати розподілу на 6 кластерів

Кластер	Recency	Frequency	Monetary	Кількість клієнтів
1	424	1.54	498	723
2	26	57.80	46 400	34
3	218	2.71	889	1 615
4	32	19.80	8 528	345
5	54	4.65	1 603	1 794
6	33	145.00	135 430	7

Джерело: побудовано автором

Для кращого розуміння характеристик кожного кластеру було реалізовано візуалізацію розподілів показників RFM в межах кожного кластеру (рис. 2.11-2.13).

Як бачимо з рис. 2.11, Кластери 1, 2 та 4 мають найменший період з моменту останнього замовлення, що може вказувати на те, що клієнти в межах даних кластерів або нові, або регулярно роблять замовлення.

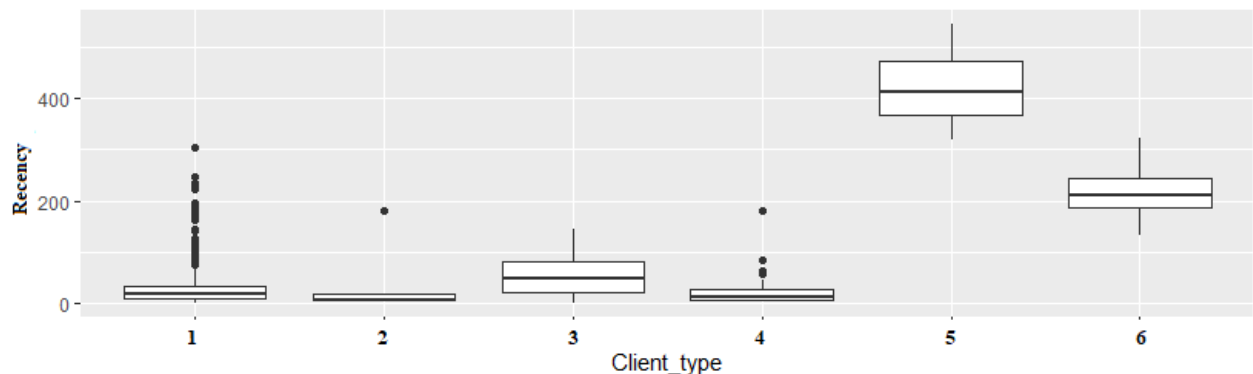


Рис. 2.11 Бокс-діаграма метрики Resency у розрізі кластерів клієнтів

Джерело: побудовано автором

Клієнти кластеру 6 в середньому робили замовлення 6 місяців назад, що вказує низьку регулярність покупок. Клієнти цього кластеру не робив замовлення уже більше року, що є попередньою ознакою втрати клієнта.

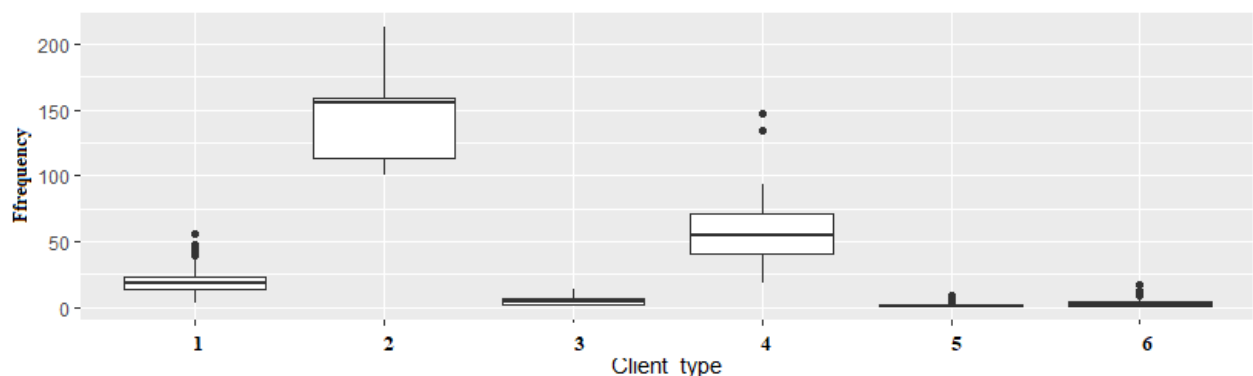


Рис. 2.12 Бокс-діаграма метрики Frequency у розрізі кластерів клієнтів

Джерело: побудовано автором

Як бачимо з рис. 2.12 більшість кластерів робило доволі мало замовлень за останні 18 місяців. Окремо варто лише виділити клієнтів в кластерах 2 та 4, які в середньому зробили від 75 до 175 замовлень.

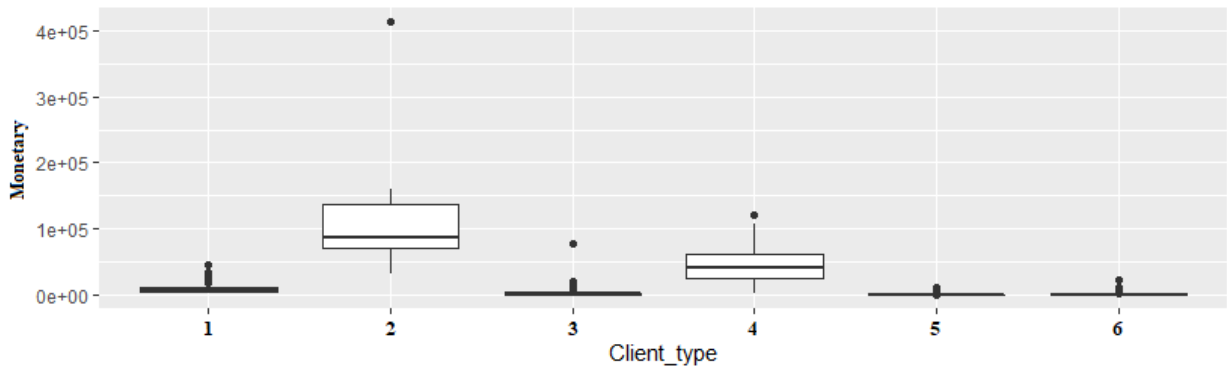


Рис. 2.13 Бокс-діаграма метрики Monetary у розрізі кластерів клієнтів

Джерело: побудовано автором

Проаналізувавши рис. 2.13, знову ж таки, варто відмітити лише кластери 2 та 4, які різко відрізняються об'ємом замовлень за останні 18 місяців. Очевидно, що це є результатом значної кількості покупок.

Зважаючи на значний вплив показника Frequency на показник Monetary, було вирішено додатково обрахувати показник середнього замовлення, як $\text{Monetary}/\text{Frequency}$. Результати обчислення подані на рис. 2.14.

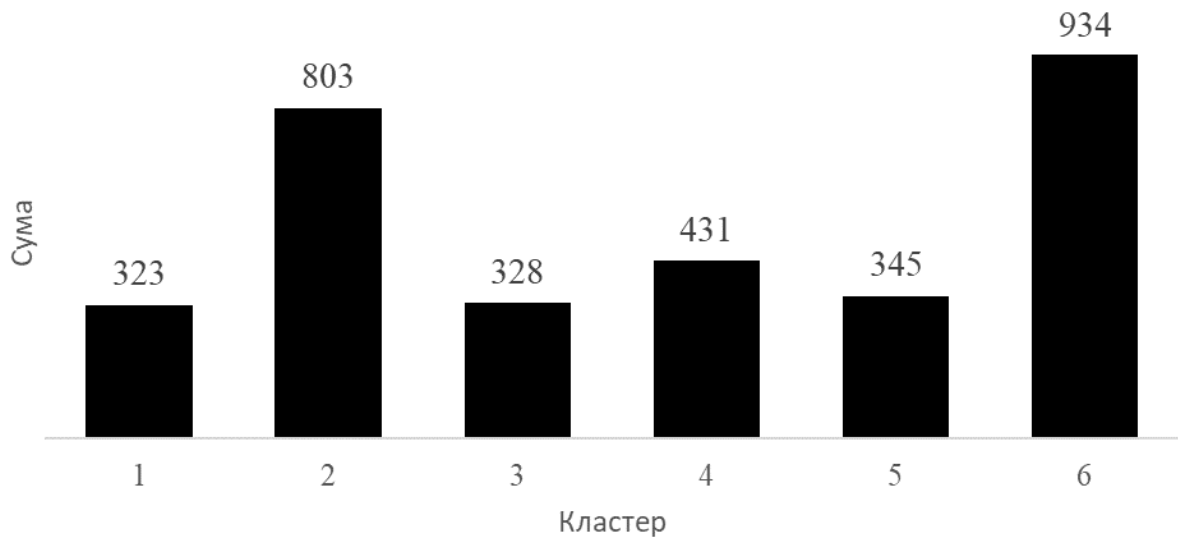


Рис. 2.14 Розподіл кластерів відповідно середньої суми замовлення

Джерело: побудовано автором

Як видно з таблиці 2.16 та з рис. 2.11-2.14, отримані кластери характеризують такі типи клієнтів:

Кластер 1. **Втрачені клієнти (Lost)** – клієнти, які не купують протягом довгого часу, здійснили менше 2 покупок за 18 місяців, остання з яких була більше року тому.

Кластер 2. **Новий оптовий покупець (New_wholesales)** - високий середній чек та активність, але пройшло небагато часу з першої покупки. Необхідно докласти зусиль для підвищення лояльності покупця до бізнесу.

Кластер 3. **Клієнти, яких ми скоро втратимо (Recent)**. Вони проявили середні показники активності, але пройшло багато часу з останньої покупки. Клієнти на яких варто звернути увагу і намагатися вплинути на здійснення більш частих операцій.

Кластер 4. **Активний роздрібний покупець (Act_retail)** - висока активність, висока частота та низька давність покупок, високий середній чек. Найцінніший і найлояльніший тип покупців для бізнесу.

Кластер 5. **Нові роздрібні покупці (New_retail)** - висока активність, але протягом короткого періоду, достатній рівень вартості замовлень. Необхідно докласти зусиль, щоб перетворити їх на постійних клієнтів.

Кластер 6. **Активний оптовий покупець (Act_wholesales)** - високий середній чек і активність, купують протягом тривалого періоду. Найприбутковіший тип клієнтів.

Далі, моделювання поведінки клієнтів інтернет магазину вимагає компіляції даних у формат, підходящий для реалізації алгоритмів машинного навчання. Першим кроком буде створення змінної інтересу, у нашому випадку – час до наступного замовлення клієнта.

За дату останньої покупки було прийнято максимальну дату по кожному клієнту у «минулій» активності. Датою наступної покупки була мінімальна дата по кожному з клієнтів у вибірці «майбутньої» активності.

Часом до наступної покупки була прийнята кількість днів між датою наступної та останньої покупок. Якщо наступного замовлення не відбулося у наступні 6 місяців, час до наступного замовлення встановлювався на рівні 999 днів. Візуалізація розподілу кількості клієнтів відповідно часу до наступного замовлення подано на рис. 2.15.

Атрибутами моделі прогнозування часу до наступного замовлення було обрано показники, які найкраще описують купівельну активність.

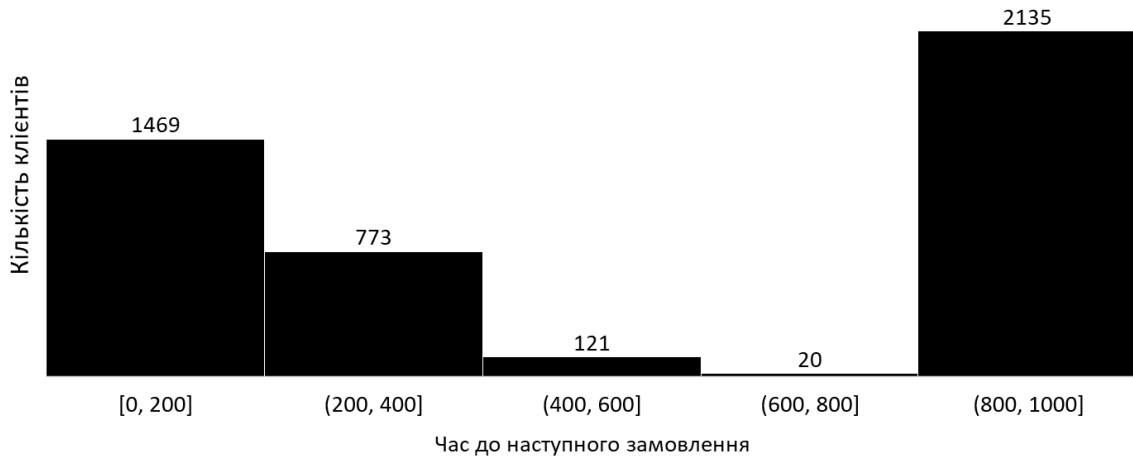


Рис. 2.15 Розподіл клієнтів відповідно часу до наступного замовлення

Джерело: побудовано автором

- Кількість днів між останніми чотирма замовленнями клієнта (Dif1, Dif2, Dif3). У ході обчислень було знехтувано тими клієнтами, які зробили за весь період менше 4 покупок. У фінальній вибірці залишилося 1795 клієнтів.
- Показники середньоквадратичного відхилення та середнє значення кількості днів між усіма покупками, які були здійснені клієнтом (StD, Ave).
- Сегмент клієнта (Client_Type), обчислений на попередньому етапі реалізації моделі поведінки клієнтів.

Так як для реалізації моделі прогнозування часу до наступного замовлення було обрано алгоритми класифікації, нашу залежну змінну необхідно з неперервного формату перетворити в категорійний. У розробленій моделі в якості класів будуть виступати періоди часу у майбутньому, в які клієнт з найбільшою ймовірністю здійснить наступну покупку. Дані періоди обираються, виходячи з бізнес потреб комерційної організації. Було обрано такі класи:

- здійснення покупки буде протягом наступних 2 місяців;
- здійснення покупки буде протягом наступних 6 місяців;
- період від 6 до 15 місяців;
- більше не буде здійснено жодної операції.

Фінальний розподіл клієнтів відповідно часу до наступного замовлення подано на рис. 2.16.

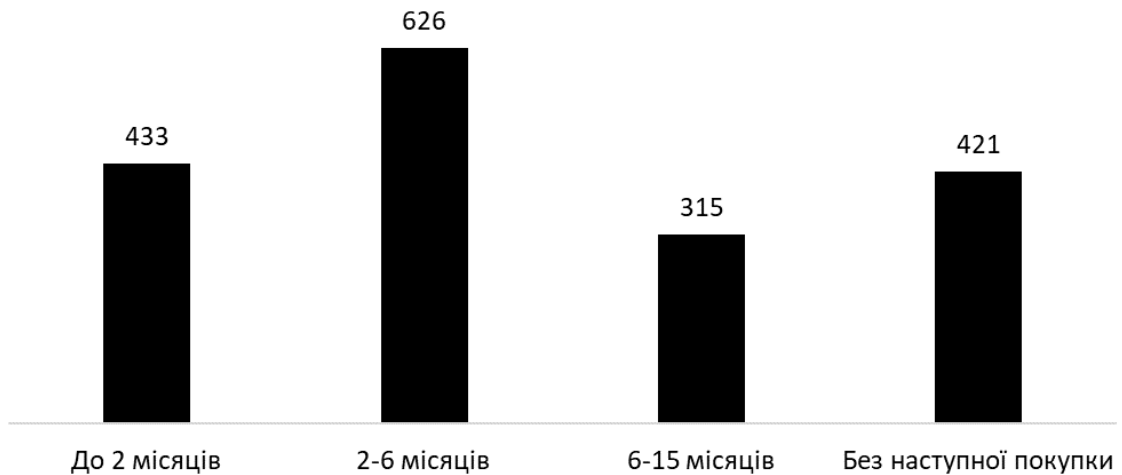


Рис. 2.16 Розподіл клієнтів відповідно часу до наступного замовлення

Джерело: побудовано автором

Приклад вхідної інформації для побудови моделей прогнозування поведінки клієнтів подано у таблиці 2.17.

Таблиця 2.17

Попередня вибірка для побудови моделей класифікації

Client_ID	Час до наступної покупки	Dif1	Dif2	Dif3	Ave	StD	Client_Type
13085	2-6 місяців	384	2	57	148	207	Act_retail
18102	До 2 місяців	2	1	1	1	1	Act_wholesales
18087	6-15 місяців	89	9	5	34	47	New_retail
15413	Без жодної операції	7	26	246	71	82	Lost
16167	Без жодної операції	1	24	179	68	97	Recent
13767	До 2 місяців	2	5	8	5	3	New_wholesales

Джерело: побудовано автором

Для оцінки якості моделювання використовувалась оцінка «Ассурасу» (2.14). У ході моделювання важливо переконатися, що результати моделювання не залежать від конкретного набору даних. Для вирішення цієї задачі під час впровадження алгоритму класифікації клієнтів було застосовано десятикратну перехресну перевірку.

У таблиці 2.18 наведено характеристики розподілу показника точності «Ассурасу»: його мінімальне, максимальне, середнє значення та 1, 2 і 3 квартилі. Дані характеристики отримані на навчальній вибірці для кожного з розглянутих методів. Як бачимо з таблиці 2.18, на етапі навчання найкраще себе показали алгоритми LDA (54%) та RF (53%).

Таблиця 2.18

Результати перехресної перевірки на тренувальній вибірці

Accuracy	Min,	1st Qu,	Median	Mean	3rd Qu,	Max,
LDA	0.48	0.52	0.54	0.54	0.55	0.59
CART	0.49	0.51	0.52	0.53	0.55	0.57
kNN	0.30	0.31	0.33	0.34	0.35	0.43
SVM	0.46	0.53	0.53	0.53	0.54	0.59
RF	0.47	0.51	0.53	0.53	0.54	0.59

Джерело: побудовано автором

Побудовані моделі було, також, перевірено на тестовій вибірці. Найкращим чином тут показав себе алгоритм RF (54%). Далі, для прогнозування часу до наступного замовлення будемо застосовувати саме цей алгоритм.

Результати реалізації на тестовій вибірці методу «випадковий ліс» в абсолютних показниках представлено у таблиці 2.19. У таблиці наведено значення кількості клієнтів, віднесених до визначених класів.

Таблиця 2.19

Результати прогнозування за методом «випадковий ліс»

Період наступної покупки		Реальні значення			
		До 2 місяців	2-6 місяців	6-15 місяців	Без жодної операції
Прогнозні значення	До 2 місяців	61	11	5	4
	2-6 місяців	53	126	24	45
	6-15 місяців	0	7	40	47
	Без жодної операції	0	1	10	15

Джерело: побудовано автором

Слід зазначити, що оцінка якості моделей за допомогою метрики «Ассигасу» необхідно використовувати дуже обережно, оскільки вона досить чутлива до екстремальних значень кількості об'єктів, які утворюють певний клас. Наприклад, якщо існує 2 класи об'єктів, де в першому класі 3 об'єкти, а в другому 97. Алгоритму достатньо всім об'єктам присвоїти другий клас для отримання точності в 97%. З огляду на дану проблему, для оцінки фінальної моделі було побудовано «матрицю невідповідності».

У таблиці 2.20 подано результати розрахунку критеріїв якості моделі за методом «матриця невідповідності» для всіх наявних класів. На основі матриці

невідповідності обраховано такі показники, як: Precision (2.15), Recall (2.11), та F (2.16).

Таблиця 2.20

Результати обчислення показників якості моделі

	Precision	Recall	F
До 2 місяців	75.31%	53.51%	62.56%
2-6 місяців	50.81%	86.90%	64.12%
6-15 місяців	42.55%	50.63%	46.24%
Без жодної операції	57.69%	13.51%	21.90%

Джерело: побудовано автором

Як можна побачити, не маючи ніякого уявлення про купівельні переваги клієнтів на початку дослідження, було досягнуто до 64% точних передбачень періоду до наступного замовлення. Найкраще алгоритму вдається передбачити поведінку клієнтів в межах “До 2 місяців” та “2-6 місяців”, показник F для яких встановився на рівні 62.5% і 64.1% відповідно.

Використовуючи розроблений підхід щодо моделювання поведінки клієнтів власники інтернет-магазину мають змогу: зосереджувати рекламні бюджети лише для тих клієнтів, які з великою ймовірністю зроблять наступну покупку; будувати стратегію управління клієнтським досвідом (активувувати клієнтів, які пішли, пропонувати знижки клієнтам, які збираються піти тощо).

Лістинг R-коду для реалізації всіх етапів моделювання в підрозділі 2.3 подано в додатку Г.

Висновки до розділу 2

В другому розділі дисертаційної роботи розроблено загальний концептуальний підхід до моделювання поведінки клієнтів за умови обмеженості даних.

Рекомендації щодо моделювання ґрунтуються на твердженні, що ключовим фактором, який суттєво впливає на точність і надійність моделей поведінки клієнтів, є якість даних. Різні фактори, такі як: наявність пропущених даних, невідповідність форматів даних або якщо наявні дані містять значну кількість шуму, упереджень та неточностей, можуть мати негативний вплив на процес моделювання. Було обґрунтовано, що наявність даних низької якості не тільки створює перешкоди для початкового аналізу, але й може призвести до помилкових інтерпретацій та неефективного вибору стратегії.

В контексті моделювання клієнтської поведінки на процес значною мірою впливають особливості доступних джерел даних. Було запропоновано класифікацію джерел даних, а саме: відкриті дані та комерційні дані, наведено унікальні характеристики, переваги та недоліки кожного типу.

Розроблено процес побудови моделей поведінки клієнтів в умовах обмеженості даних, який складається з 9 кроків, де перші чотири кроки описують підходи з підготовки даних до моделювання, наступні 4 кроки – це процес вибору, налаштування та реалізації алгоритмів машинного навчання, останній – процес оцінювання результатів моделювання.

Проведено експериментальне дослідження прогнозування фінансового успіху бізнес-клієнтів консалтингової агенції на основі запропонованого процесу. Реалізовано очищення та підготовку даних для моделювання методами машинного навчання, використано підходи для підвищення ефективності моделювання в умовах обмежених даних. Застосовано 3 методи машинного навчання: дерево класифікації та регресії, логістична регресія та випадковий ліс. Було реалізовано процес налаштування гіперпараметрів моделей, проведено множинні перехресні перевірки за п'ятьма різними метриками оцінювання якості моделей. Найкращим алгоритмом було обрано дерево класифікації та регресії. На основі проведеного аналізу підходів до прогнозування поведінки бізнес-клієнтів доведено ефективність запропонованого методу, шляхом порівняння отриманих прогнозних результатів з пороговими значеннями використовуваних метрик.

Спроектовано та реалізовано систему моделей, що дозволила передбачити час до наступного замовлення клієнта онлайн-магазину. Спочатку було проведено сегментацію клієнтів інтернет-магазину методом k-середніх за показниками купівельної активності RFM. На підставі розрахованих значень показників кластеризації виявлено 6 груп клієнтів: 1. втрачені клієнти— клієнти, які не купують протягом довгого часу, здійснили менше 2 покупок за 18 місяців, остання з яких була більше року тому; 2. новий оптовий покупець - високий середній чек та активність, але пройшло небагато часу з першої покупки; 3. клієнти, яких ми скоро втратимо - проявили середні показники активності, але пройшло багато часу з останньої покупки; 4. активний роздрібний покупець - висока активність, висока частота та низька давність покупок, високий середній чек; 5. нові роздрібні покупці - висока активність, але протягом короткого періоду, достатній рівень вартості замовлень; 6. активний оптовий покупець - високий середній чек і активність, купують протягом тривалого періоду.

На другому етапі було реалізовано 5 алгоритмів класифікації: лінійний дискримінантний аналіз, дерево класифікації та регресії, «випадковий ліс», метод опорних векторів, k - найближчих сусідів. Найкращу прогностну силу показав алгоритм «випадковий ліс». Для уточнення якості моделювання було побудовано «матрицю невідповідності». Зважена точність «F» в групах «До 2 місяців» та «2-6 місяців» досягла 62.5% і 64.1% відповідно. Результати моделювання та подальшого оцінювання свідчать, що запропонований підхід щодо моделювання поведінки клієнтів демонструє свою здатність отримувати значні знання з первинних даних клієнтів, незалежно від масштабу, повноти та якості даних. Ці висновки дають практичну інформацію про поведінку клієнтів, яка може безпосередньо впливати на маркетингові рішення.

Основні результати Розділу 2 опубліковано у наукових працях автора [118, 117, 171, 169].

РОЗДІЛ 3

УДОСКОНАЛЕННЯ МАРКЕТИНГОВОГО МЕНЕДЖМЕНТУ В БАНКУ НА ПІДГРУНТІ МОДЕЛЮВАННЯ ПОВЕДІНКИ КЛІЄНТІВ

3.1 Особливості моделювання поведінки клієнтів за наявності якісної інфраструктури даних на прикладі клієнтів банку.

У сучасну епоху, банки взаємодіють з клієнтами через багато точок дотику, що призводить до регулярного генерування та накопичення значних обсягів даних в рамках їх операційної діяльності. Системи банку збирають різноманітні набори даних починаючи від клієнтських транзакцій та історії взаємодії з клієнтами, і закінчуючи продуктовими портфелями, демографічними показниками та маркетинговою активністю.

Велика кількість даних ставить банки у вигідне становище, порівняно з іншими сферами економічної активності, та слугує складовою їхньої діяльності з розробки маркетингових стратегій. Здатність банків використовувати різні методи моделювання поведінки клієнтів надає їм значну перевагу і робить їх проактивними суб'єктами у сфері маркетингу, заснованого на даних. Використовуючи ці дані, банки мають можливість класифікувати клієнтів, аналізувати їхню унікальну поведінку, передбачати їхні потреби та надавати персоналізовані послуги.

Звісно, у банках зберігаються певні обмеження даних, які все ще ускладнюють процес управління клієнтським досвідом. Хоча банківські дані містять цінні поведінкові характеристики, їм бракує важливої персональної інформації клієнтів, необхідної для більш повного їх розуміння. Очевидною, є відсутність конкретної інформації про життєві події, потреби, самореалізацію, уподобання чи ставлення клієнтів.

Також банківська сфера піддається сильному впливу зовнішніх факторів, які можуть кардинально змінити підходи до співпраці між банком та клієнтами.

Непередбачуваність, що спостерігається в банківському середовищі, виникає з багатьох джерел:

Економічні фактори, включаючи коливання економічних показників, таких як процентні ставки, інфляція та безробіття, мають потенціал для значного впливу на поведінку клієнтів. Наприклад, стрімке зростання облікової ставки може спонукати більшу кількість клієнтів до заощаджень, тоді як падіння облікової ставки може стимулювати запозичення та інвестиційну діяльність.

Ринкові тенденції можуть суттєво впливати на поведінку клієнтів, охоплюючи інвестиційні потреби, так і буми на ринку нерухомості, тенденції на ринку валютнообміну. Потенційно на клієнтів можуть впливати засоби масової інформації або перспектива отримання швидких прибутків.

Регуляторні зміни можуть впливати на поведінку клієнтів через їхній вплив на доступність та умови певних фінансових продуктів та послуг. Наприклад, запровадження більш жорстких стандартів кредитування може створити додаткові труднощі для клієнтів, які прагнуть отримати позику. Обмеження у сфері валютнообміну може спричинити значне уповільнення міжнародної активності клієнтів.

У сфері маркетингу на основі даних у банківському секторі ключовими факторами успіху є адаптивність та універсальність. Банки не обмежені у використанні єдиної методики, коли йдеться про моделювання клієнтської поведінки. Натомість вони можуть гнучко адаптуватися і використовувати різноманітні підходи, що походять з різних дисциплін.

Інтелектуальний маркетинг в банку базується на таких даних:

Демографічні дані. Використання соціально-демографічної інформації, часто, слугує основою для моделювання поведінки клієнтів. Ця інформація охоплює декілька соціально-демографічних характеристик, таких як вік, стать, дохід та географічний регіон. Вищезазначені дані мають важливе значення в процесі сегментації клієнтів та в отриманні уявлення про вплив різних демографічних факторів на фінансову поведінку та тривалість активності.

Інформація про рахунки. Інформація про рахунки клієнтів, що охоплює такі фактори, як тип рахунку, баланс та історія транзакцій, може дати цінне уявлення про їхній фінансовий добробут та моделі фінансової активності. Вищезгадані дані можуть бути корисними в різних методах моделювання, включаючи, прогнозування відтоку клієнтів або виявлення потенційних клієнтів з високими доходами.

Інформація про кредити. Для банків, що надають кредитні послуги, інформація про клієнтів часто включає кредитні скоринги, суми кредитів, розмір заборгованості, відсоткові ставки та історію погашення. Ці дані відіграють ключову роль у моделюванні кредитного ризику та оцінці кредитоспроможності клієнтів.

Продуктове навантаження. Дана інформація стосується багатьох продуктів і послуг, які пропонує банк, включаючи кредитні картки, ощадні рахунки, поточні рахунки в національній та іноземних валютах, кредити готівкою, кредити на авто, та іпотечні кредити. Ці знання дозволяють банку брати участь у перехресних продажах та цільових маркетингових стратегіях.

Фінансова інформація – стосується доходів у банківському секторі, а саме чистого банківського доходу, чистого маржинального доходу, комісійного доходу тощо. Статистика доходів у банківському секторі відіграє вирішальну роль в інформаційному забезпеченні процесів прийняття стратегічних рішень.

Чистий банківський дохід (NBI) слугує показником для оцінки загальної прибутковості банку (клієнту чи продукту) вимірюючи різницю між його загальними доходами (комісійні та маржинальні) та операційними витратами.

Чистий маржинальний дохід (NIM) – це показник, який слугує для вимірювання доходу отриманого виключно від різниці вартості пасивів та активів банку. Цей показник дозволяє банкам здійснювати сегментацію клієнтів, що дає змогу ідентифікувати та визначати пріоритетність клієнтів з найвищою дохідністю.

У сучасній банківській індустрії, що характеризується інтенсивною конкуренцією, розуміння та аналіз поведінки клієнтів стали ключовими

елементами для досягнення довгострокового зростання та утримання міцних позицій на ринку. Банки завжди шукають нові стратегії для підвищення ефективності своїх маркетингових активностей, управління витратами та розвитку більш ефективної комунікації зі своїми клієнтами. Виникнення такої необхідності можна пояснити кількома факторами, такими як висококонкурентний характер ринку, зростаюча схильність до підвищених очікувань і вимог клієнтів, та беззаперечна важливість збереження лояльності існуючої клієнтської бази. Щоб досягти успіху в постійно мінливому бізнес-середовищі, банки повинні прийняти стратегію персоналізації, використовуючи клієнтські інсайти, отримані в результаті аналізу даних та моделювання.

Епоха міцних, ексклюзивних зв'язків з клієнтами минає. У наш час клієнти практикують диверсифікацію своїх операцій між багатьма банками, стратегічно обираючи найоптимальніші товари, зручні канали та заохочення, які пропонує кожна установа.

Серед особливостей банківської діяльності слід виділити те, що універсальні банки обслуговують широкий спектр індивідуальних і корпоративних клієнтів, які мають різні фінансові потреби. Стратегічне узгодження відповідних продуктів і послуг з унікальними профілями і поведінкою кожного клієнтського сегмента має першорядне значення для досягнення організаційного успіху. Основними клієнтськими сегментами є індивідуальні клієнти, малий та середній бізнес, та великі корпоративні організації, кожен з яких має свої особливості.

Два основні підходи моделювання, які привернули значну увагу в цій галузі – це сегментація [26] та моделювання виживання [6]. Ці методи вирішують окремі, але взаємопов'язані завдання, кожна з яких сприяє досягненню головної мети – підвищенню рівня залучення, утримання та задоволеності клієнтів. При ретельному застосуванні ці підходи дозволяють банкам ефективно аналізувати складні закономірності клієнтських взаємодій, отримуючи таким чином цінні дані для цільового маркетингу, адаптованих продуктів та проактивних заходів з утримання клієнтів.

Сегментація клієнтів – це практика розподілу клієнтської аудиторії на окремі кластери, використовуючи спільні атрибути, уподобання та поведінку як визначальні критерії. Цей метод виходить за рамки базового демографічного аналізу, досліджуючи тонкощі залучення клієнтів, моделі використання продуктів та фінансові характеристики. Завдяки процесу сегментування клієнтів банки отримують можливість адаптувати маркетингові повідомлення, продуктові пропозиції та методи утримання клієнтів.

Як вже зазначалося в пункті 1.3 **Моделювання виживання** – це статистичний алгоритм, який зосереджений на оцінці тривалості часу до настання певної події. У сфері моделювання поведінки клієнтів банку основна увага часто зосереджується на виникненні плинності або відтоку клієнтів. Використання цієї методики є надзвичайно важливим для розробки ефективної стратегії утримання клієнтів, оскільки вона пропонує цінне розуміння факторів, які впливають на відтік клієнтів, а також заходів, які можуть бути застосовані для подовження життєвого циклу клієнтів.

Інтеграція сегментації та моделювання виживання пропонує комерційним організаціям повний набір інструментів для утримання клієнтів. Завдяки процесу сегментації клієнтської бази банки можуть розпізнати та виокремити цінних клієнтів, можливі ризики відтоку клієнтів, а також перспективи перехресних продажів. Моделювання виживання поглиблює ці знання, роблячи прогнози щодо відтоку клієнтів, що дозволяє вживати проактивних заходів для збереження важливих клієнтів.

Розглянемо клієнтську базу банку з двох найбільш критичних перспектив:

- клієнти які приносять найбільший дохід і є найбільш цінними – корпоративні бізнес-клієнти;
- клієнти з найменшою лояльністю і найменшим життєвим циклом – індивідуальні клієнти залучені через мережу партнерів, в нашому випадку через мережу автодиллерів.

Розробка стратегій утримання даних сегментів клієнтів є надважливою для банку. Якщо досліджувати передумови становлення поведінки корпоративних

клієнтів банку, то через різке зростання конкуренції, дерегуляцію банківської діяльності та впровадження нових технологій, банки все частіше звертаються до практики оптимізації витрат, з метою досягнення максимальної ефективності корпоративного бізнесу. Банки, які забезпечують максимальну прибутковість для кожного клієнта, отримують найбільш суттєву конкурентну перевагу. Відповідно, концепція клієнтської цінності та її максимізації набуває життєво важливого значення.

Для більшості універсальних банків, які надають широкий спектр послуг для роздрібного та корпоративного сегментів, частка прибутку, отримана від корпоративного бізнесу, є системоутворюючою. **Корпоративний банкінг** – це частина банківської діяльності, яка працює з бізнес-клієнтами. Зазвичай він обслуговує широке коло клієнтів - від малого та середнього бізнесу з кількомільйонним доходом до великих корпорацій з мільярдними продажами.

Банки потребують більш своєчасної та адекватної інформації, щоб краще обслуговувати своїх корпоративних клієнтів. На відміну від роздрібного бізнесу, традиційна модель продажів у корпоративному банкінгу пов'язана з особистою взаємодією між співробітниками банку та власниками бізнесу, з подальшим встановленням індивідуальних умов обслуговування. Такий підхід часто виявляється неефективним, коли перед банком постає питання максимізації ефективності бізнес-діяльності, особливо коли в його портфелі тисячі корпоративних клієнтів. Непросто забезпечити індивідуальний підхід до кожного клієнта, не втратити при цьому ефективність управління вартістю клієнтів і не витратити багато людських ресурсів одночасно. З іншого боку, конкурентне середовище не дозволяє забезпечити однакову цінову політику для всіх корпоративних клієнтів.

В Україні економічні кризи 2008-2010 та 2014-2015 років призвели до зупинки банківського кредитування в іноземній валюті. Як наслідок, депозити в євро стали збитковим видом послуг для банків (вартість обслуговування становить 1% на рік). Відповідно, якщо клієнт має лише один продукт – рахунок в євро, і не здійснює жодних банківських операцій, за які банк отримувал би

комісійний дохід – такий клієнт є повністю збитковим. З огляду на це, виникає необхідність мінімізації операційних витрат за рахунок:

- зменшення залишків на рахунках клієнтів в євро;
- збільшення кількості операцій з рахунками в євро;
- запровадження нової комісії за обслуговування рахунків в євро.

Одним з основних варіантів вирішення цієї проблеми є кластеризація корпоративних клієнтів за відповідними групами схожості – маркетинговими сегментами та розробка цінової стратегії окремо для кожного з них. Очікується, що такий підхід до сегментації, пов'язаний з рекомендаціями, дозволить оптимізувати операційні витрати без запровадження єдиної структури тарифів для всіх корпоративних клієнтів банку. В результаті, банки зможуть залишатися привабливими в конкурентному середовищі, уникаючи при цьому зайвих витрат. Крім того, практичні результати дослідження можуть слугувати основою для розробки автоматизованих маркетингових кампаній та персоналізованих комунікаційних стратегій, адаптованих до кожного маркетингового сегменту корпоративних клієнтів, що сприятиме підвищенню загального рівня задоволеності клієнтів.

У дослідженні поведінки бізнес-клієнтів банку пропонується використовувати лише метод кластеризації k -середніх, який завдяки своїй простоті вважається одним з найпоширеніших алгоритмів кластеризації. У пункті 2.1 було детально описано особливості реалізації даного алгоритму, тому ми не будемо зупинятися на цьому питанні ще раз. Єдине, що варто відмітити, що в даній частині дослідження для визначення оптимальної кількості кластерів (k) пропонується використовувати бібліотеку NbClust мови програмування R, яка дозволяє оцінити необхідну кількість кластерів за 26 критеріями, описаними в таблиці 2.1. Оптимальною кількістю кластерів ми будемо вважати ту кількість, за яку «проголосувала» більшість критеріїв. Однак остаточна кількість кластерів завжди обирається з урахуванням бізнес-потреб організації.

Останнім кроком розробки інтелектуальної сегментації та розробки маркетингової стратегії є опис побудованих кластерів. Кожен сегмент

відрізняється своїми середніми показниками діяльності, які необхідно розрахувати. Розрахунок середніх значень параметрів моделі дає можливість надати рекомендації, щодо необхідності проведення конкретної роботи з тією чи іншою групою клієнтів для зменшення операційних витрат банку [76].

Ще однією частиною цього дослідження є огляд особливостей поведінки клієнтів, першим продуктом яких був кредит на автомобіль. Банки по всьому світу активно використовують автокредитування як ефективний засіб залучення нових клієнтів, особливо з високим фінансовим потенціалом [116].

Однією з основних проблем роздрібного банківського обслуговування є утримання клієнтів. З посиленням конкуренції клієнти, які є часто мотивованими ціною послуг, готові з легкістю змінити банк, щоб отримати вигідніші умови. Це особливо правдиво у випадку автокредитування, де клієнти найчастіше мотивовані низькими процентними ставками. Таким чином, банк, який запропонував найнижчу ціну, отримує клієнтів, але водночас має майже нульовий прибуток від видачі автокредитів. За цією моделлю залучення клієнтів банк розраховує отримати прибуток від інших клієнтських активностей, таких як операції з картою чи поточними рахунками. Однак утримання клієнта після погашення автокредиту стає великою проблемою для банків, оскільки клієнт, який не бачить додаткової цінності в продуктах банку, може легко переключитися на конкурентну пропозицію. Основною проблемою, яку має вирішити банк, є пошук ефективних стратегій утримання клієнтів після повного погашення автокредиту. Це особливо важливо, оскільки передчасне закриття клієнтів може призвести до потенційної втрати доходів від інших банківських продуктів [116].

Деякі банки використовують стратегію збереження конкурентоспроможності на ринку, надаючи кредити на автомобіль за мізерними відсотковими ставками, наприклад, 0,01% річних. Суттєвим обмеженням цього продукту є те, що він видається на термін до 2 років, тому часу для заробітку на клієнті небагато. Стратегія цієї діяльності полягає в тому, що клієнти, які купують нові автомобілі, зазвичай є представниками заможного класу, а

автокредити є чудовим каналом для залучення таких клієнтів. Банк планує заробляти безпосередньо на активності клієнтів з іншими банківськими продуктами. Тому ефективність перехресних продажів додаткових продуктів є критично важливою.

Одним з варіантів утримання клієнтів після погашення автокредиту є пропозиція кредитної картки, яка дозволяє клієнту здійснювати повсякденні банківські операції з використанням кредитних коштів, а банку – отримувати дохід від наданих послуг. Однак банки стикаються з проблемами та перешкодами при продажу кредитних карток клієнтам з автокредитами. Зокрема, взаємодія клієнта з банком під час оформлення автокредиту є обмеженою, оскільки всі процедури оформлення здійснюються безпосередньо з автодилером в автосалоні. Це ускладнює встановлення емоційного зв'язку між банком та клієнтом. Це також знижує ймовірність використання клієнтом додаткових банківських продуктів після погашення автокредиту [116].

Іншу поведінку показують клієнти, які взяли кредит за стандартними ставками, тобто вище 0.01%. Так як банк не карає позичальника за дострокове погашення кредиту, клієнти мають схильність оформлювати позику на максимально можливий термін, але погашати борг швидше, ніж того вимагає банк. Відповідно, критерій «строк кредиту за договором» не є інформативним при аналізі тривалості клієнтської активності [172]. Відповідно важливо оцінити реальний період активності таких клієнтів, та фактори, які впливають на його підвищення.

З огляду на ці особливості продуктової пропозиції та поведінки клієнтів, було вирішено розглядати сегменти клієнтів з автокредитом за відсотковою ставкою в 0.01% та з автокредитом більше 0.01% окремо.

У банківській сфері «виживання» клієнтів має велике значення для довгострокової прибутковості. Банки використовують різні методи для утримання клієнтів та збільшення їхньої життєвої цінності, наприклад, пропонуючи персоналізовані послуги та конкурентоспроможні ціни,

покращуючи якість обслуговування. Проте виявлення чинників, які впливають на «виживання» клієнтів, залишається складним завданням [116].

Як вже зазначалося у пункті 1.3 **аналіз виживання** – це статистичний інструмент, який можна використовувати для оцінки часу до настання події-інтересу, такої як смерть, банкрутство чи закриття клієнта. У банківській справі аналіз виживання можна використовувати для аналізу поведінки клієнтів та визначення чинників, які впливають на їх утримання.

Аналіз виживання відрізняється від інших статистичних методів, оскільки частиною інформації, яка аналізується, є цензуровані дані, тобто дані, в яких подія ще не відбулася на момент закінчення дослідження або періоду спостереження.

Цензурування інформації може виникати з трьох основних причин:

1. Особа не стикається з подією до завершення дослідження.
2. Протягом періоду дослідження був випадок, коли учасник був втрачений для подальшого спостереження.
3. Вибуття учасника відбувається, коли особа припиняє свою участь у дослідженні через різні фактори, такі як смертність (якщо смерть не є основною подією, що становить інтерес) або інші причини, такі як побічні реакції на лікарські засоби або конкуруючі ризики [75].

Загалом аналіз виживання може бути цінним інструментом для банків, який дозволяє аналізувати «виживання» клієнтів і визначати фактори, що впливають на нього. Хоча аналіз виживання знайшов застосування в різних галузях банківської справи, у сфері роздрібних банківських послуг та маркетингу, все ще існує значна прогалина. Банки можуть отримати вигоду з використання аналізу виживання, щоб поліпшити свої стратегії утримання

В аналізі виживання час до настання події інтересу називається **часом виживання**, а ймовірність виживання після заданого часу називається **функцією виживання**.

Функція виживання $S(t)$ дає ймовірність того, що людина виживе довше певного часу t : тобто $S(t)$ дає ймовірність того, що випадкова величина T

перевищить вказаний час t . Функція виживання є фундаментальною для аналізу виживання, оскільки отримання ймовірностей виживання для різних значень t надає важливу підсумкову інформацію з даних про виживання.

Оскільки t змінюється від 0 до нескінченності, функція виживання може бути зображена у вигляді плавної кривої. На рис 3.1 відображено теоретичну функцію виживання, де t позначає вісь X .

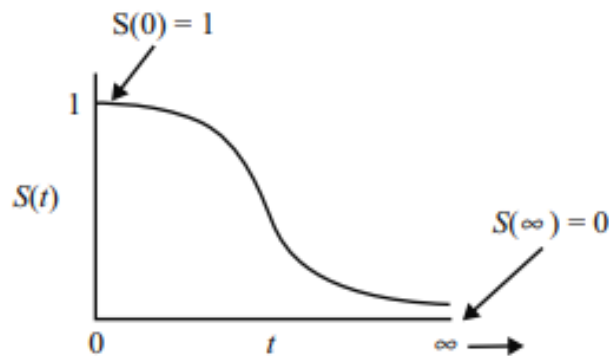


Рис 3.1 Функція виживання

Джерело: [75]

Як можна побачити з рис 3.1, всі функції виживання мають наступні характеристики:

- вони не зростають, тобто прямують донизу зі збільшенням t ;
- на початку дослідження, оскільки ніхто ще не отримав подію, ймовірність вижити після часу 0 дорівнює одиниці;
- теоретично, якби період дослідження збільшувався безмежно, врешті-решт ніхто б не вижив, тож крива виживання зрештою має впасти до нуля [75].

Функція небезпеки – ще один часто використовуваний показник у аналізі виживання, який представляє миттєву швидкість виникнення події в певний момент часу, з урахуванням того, що людина дожила до цього часу. **Рівень небезпеки** – це ймовірність виникнення події інтересу в короткому проміжку часу за умови, що вона не відбулася раніше.

Важливо визнати, що функція небезпеки відрізняється від функції виживання з точки зору їхніх відповідних фокусів. В той час як функція виживання наголошує на відсутності невдачі, функція небезпеки зосереджується

на виникненні невдачі або на самій події. Тому можна стверджувати, що функція небезпеки надає додаткову інформацію до тієї, що надається функцією виживання.

Щоб виконати аналіз виживання в R, дані мають бути структуровані у спеціальному форматі, який включає інформацію про час виживання, статус цензурування даних і будь-які додаткові змінні-предиктори. Час виживання, зазвичай, представлений як неперервна змінна, а статус цензування – як бінарна змінна, що вказує, чи сталася подія інтересу (1) чи була відцензурована (0) [102].

Моделювання виживання Вейбулла – це статистичний метод, який передбачає, що рівень небезпеки підпорядковується розподілу Вейбулла. Розподіл Вейбулла є гнучким розподілом, який може відтворювати різні форми функції небезпеки з плином часу.

У моделі виживання Вейбулла взаємозв'язок між рівнем небезпеки та змінними-предикторами моделюється за допомогою регресії, коефіцієнти якої представляють вплив кожної змінної-предиктора на рівень небезпеки, за умови, що всі інші змінні в моделі залишаються незмінними.

У моделі Вейбулла функція небезпеки описується наступним рівнянням:

$$h(t) = \lambda * \gamma * t^{(\gamma-1)}, \quad (3.1)$$

де t - це час,

λ - параметр масштабу,

γ - параметр форми розподілу Вейбулла [102].

Модель Вейбулла дозволяє збільшувати або зменшувати рівень небезпеки з часом, залежно від значення γ . Він визначає, чи рівень небезпеки збільшується ($\gamma > 1$), зменшується ($\gamma < 1$), чи залишається незмінним ($\gamma = 1$) з часом. Параметр масштабу визначає масштаб розподілу, або час, коли подія інтерес відбудеться для певної частки об'єктів спостереження, а параметр форми визначає форму розподілу. Для моделі Вейбулла позитивний коефіцієнт для змінної предиктора означає нижчий рівень небезпеки, а негативний коефіцієнт означає вищий рівень небезпеки. У моделі Вейбулла більш низьке значення параметра масштабу передбачає, що подія з більшою ймовірністю відбудеться

раніше, а більш високе значення передбачає, що вона з меншою ймовірністю відбудеться в короткостроковій перспективі.

Оцінка якості регресійної моделі виживання Вейбулла відбувається за рядом критеріїв:

- z-статистика для коефіцієнта змінної предиктора – є оцінкою коефіцієнта, поділеною на його стандартну похибку.
- p-значення для z-статистики – вимірює ймовірність спостереження z-статистики, настільки екстремальної, як відповідно нульової гіпотези, що коефіцієнт дорівнює нулю. Нижче значення p вказує на сильніші докази проти нульової гіпотези про нульовий коефіцієнт. Значення < 0.05 підтверджують статистичну значущість коефіцієнта для змінної предиктора.
- Лог-подібність моделі виживання Вейбулла є мірою того, наскільки добре модель відповідає спостережуваним даним виживання. Чим вище значення функції лог-подібності, тим краще модель відповідає спостережуваним даним про виживання.
- AIC (критерій Акаїке) - це статистична міра, що використовується для порівняння якості побудови різних моделей. AIC заснований на функції Лог-подібності та кількості параметрів у моделі. Алгоритм «карає» моделі з великою кількістю параметрів, щоб уникнути перенавчання. Чим нижче значення AIC, тим краще модель пояснює дані [172].

Для визначення найкращого поєднання змінних предикторів використовується функція `stepAIC` в R. У ході виконання алгоритму `stepAIC` відбувається покрокова процедура вибору моделі на основі AIC [102]. Поетапна процедура починається з повної моделі, що включає всі змінні-предиктори, і послідовно видаляє або додає змінні в модель, поки не буде отримана найкраща модель з найменшим значенням AIC. Покрокова процедура може бути як прямою (додавання змінних у модель), так і зворотною (видалення змінних з моделі).

Іншим підходом до моделювання рівня «виживання» є регресію пропорційних ризиків Кокса. Регресійний аналіз Кокса – це статистичний метод,

який досліджує зв'язок між часом до події, що нас цікавить, та однією або кількома змінними-предикторами. Регресійна модель є потужним інструментом для аналізу даних про виживання, перевага якого полягає в тому, що її коефіцієнти, по суті, представляють відношення ризиків (HR) для кожної предикторної змінної.

Формула регресійної моделі Кокса наступна:

$$h(t|x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p), \quad (3.2)$$

де $h(t|x)$ функція ризику в момент часу t із заданими предикторними змінними x ;

$h_0(t)$ – базисна функція ризику в момент часу t , яка представляє ризик, коли всі предикторні змінні дорівнюють нулю;

$\beta_1, \beta_2, \dots, \beta_p$ – коефіцієнти регресії, які представляють відношення лог-ризиків для кожної предикторної змінної;

$x_1, x_2, \dots, x_p$ – предикторні змінні.

Основними особливостями регресійної моделі Кокса є:

- є напівпараметричною, тобто вона не робить припущень щодо форми функції ризику в часі. Натомість, вона оцінює відношення ризиків для кожної змінної-предиктора.
- призначена для аналізу часових даних, тобто даних, які вимірюють час між початковою подією і наступною подією, наприклад, час між відкриттям і закриттям банківського рахунку клієнта.
- враховує цензурування, яке відбувається, коли подія ще не відбулася для деяких об'єктів до кінця досліджуваного періоду.
- може аналізувати як безперервні, так і категорійні змінні-предиктори.
- припускає, що рівні небезпеки є пропорційними в часі. Іншими словами, вплив предикторних змінних на функцію небезпеки не змінюється з часом [116].

Позитивне значення коефіцієнтів регресії Кокса (3.3) означає, що ризик відтоку клієнтів вищий для спостережень з вищими значеннями предикторних змінних або за наявності певної категорії змінної.

Відношення небезпек (HR) вимірює вплив кожної предикторної змінної на функцію небезпек і визначається за формулою:

$$HR = \exp(\beta_p), \quad (3.3)$$

де β_p - Коефіцієнт регресії Кокса.

Значення HR менше 1 вказує на те, що предикторна змінна пов'язана зі зменшенням ризику відтоку клієнтів. І навпаки, HR, що дорівнює 1, вказує на те, що предикторна змінна не впливає на рівень ризику відтоку клієнтів.

Статистика Вальда (w) оцінює, чи є β_p коефіцієнт даної змінної статистично значуще відрізняється від 0 і розраховується за формулою:

$$w = \beta_p / se(\beta_p), \quad (3.4)$$

де β_p – Коефіцієнт регресії Кокса;

$se(\beta_p)$ – стандартна похибка коефіцієнта регресії Кокса.

Значення $p - value$ для кожної предикторної змінної вказує на статистичну значущість тесту Вальда (3.4). Статистично значущими вважаються коефіцієнт β_p , для яких $p - value$ менше 0.05.

Також необхідно побудувати довірчі інтервали (CI 95%) для кожного коефіцієнта ризику. Якісний вплив – це змінна, для якої довірчий інтервал відношення ризиків не перетинає 1.

Останнім кроком у перевірці якості моделі Кокса є перевірка припущення про пропорційні небезпеки. Припущення про пропорційні небезпеки (РН) можна перевірити за допомогою статистичних тестів та графічної діагностики на основі масштабованих залишків Шенфельда. Для кожної предикторної змінної функція `cox.zph` співвідносить відповідний набір масштабованих залишків Шенфельда з часом, щоб перевірити незалежність між залишками і часом. Припущення про пропорційні небезпеки підтверджується незначним зв'язком між залишками та часом і відхиляється значним зв'язком.

Для інтерпретації результатів тесту `cox.zph` можна також побудувати графік залежності залишків Шенфельда від часу для кожного фактора в моделі. Припущення про пропорційні небезпеки виконується, якщо залишки випадково

розподілені навколо нуля. Однак, якщо в залишках спостерігаються будь-які помітні закономірності в часі, припущення про пропорційність небезпек порушується, і регресійну модель Кокса, можливо, доведеться переглянути.

Якщо припущення про пропорційність небезпек виконується, оцінений тренд залишків у часі повинен мати форму прямої горизонтальної лінії, що вказує на те, що очікувані та спостережувані значення змінної не змінюються систематично в часі. Якщо розрахунковий тренд залишків у часі має ступінчасту форму з додатним або від'ємним нахилом, це свідчить про порушення припущення про пропорційність ризиків. Криволінійний або U-подібний тренд вказує на те, що співвідношення небезпек змінюється з часом [116].

3.2 Моделювання тривалості клієнтської активності з використанням економетричних моделей виживання

Для виконання дослідження було використано статистику активності клієнтів банку, які були залучені у період з 2018 року по 2022 рік, і першим продуктом яких був кредит на автомобіль.

Основою дослідження слугував показник, який відображає тривалість активності клієнта (**time_to_deactivation** – кількість днів з моменту відкриття до моменту закриття клієнта), значення якого моделювалося. Ознакою того, чи спостереження є цензурованим, виступив фактор **deactivated** – ознака закриття клієнта (1 – закритий; 0 – діючий).

Іншими факторами, вплив яких на тривалість активності клієнтів було досліджено, є:

- gender – стать клієнта (F – жіноча, M – чоловіча);
- age_group – вікова група (до 35 років, 35-45, 45-55, старше 55 років).
- income_group – рівень доходів клієнта (до 35 тис. грн в місяць, 35-70, 70-150, більше 150 тис. грн в місяць);

- region_group – регіон проживання ("cv_if_km_lt_lv_rv_te_uz" – західний регіон, "kv" – Київ, "sm_kh_dp_dn_zp_mk_ks_od" – східний регіон, "cn_ck_pl_kr_vn_zt" – центральний регіон);
- credit_card – ознака продажу кредитної картки клієнту (0 - не продано, 1 - продано кредитну картку).
- active_credit_card - вказує на наявність у клієнта активованої кредитної картки (0 - відсутність активованої кредитної картки, 1 - наявність активованої кредитної картки).
- IR_0.01% - ознака зниженої відсоткової ставки (1 – відсоткова ставка по кредиту 0.01%, 0 – відсоткова ставка по кредиту вища за 0.01%).

Як зазначалося у пункті 3.1 поведінка клієнтів із відсотковою ставкою 0.01% та без кардинально відрізняється. Тому визначення факторів подовження клієнтського життя для цих груп будемо досліджувати окремо.

Для клієнтів з відсотковою ставкою більше 0.01% нам важливо оцінити реальний життєвий період, та які соціально-демографічні критерії на нього впливають. Тут вирішується більш прикладна задача, яка дасть зрозуміти саме тенденції, які сформувалися в портфелі клієнтів.

Для клієнтів з ставкою в 0.01% нам критично важливо зрозуміти, які додаткові продукти здатні зупинити деактивацію клієнтів після погашення автокредиту. Фактором впливу було обрано додатковий продаж кредитних карток. У ході роботи над цим завданням критично важливо не тільки переконатися у впливі кредитної картки на рівень виживання, але і оцифрувати цей вплив. Це важливо для донесення всієї значущості впливу до вищого керівництва банку, і надання пріоритету проектам з розвитку кампаній по продажу кредитних карток клієнтам з автокредитами.

Першим етапом дослідження буде аналіз поведінки клієнтів з відсотковою ставкою по автокредиту вище 0.01%. Для виконання цього завдання було обрано регресійну модель виживання Вейбулла (3.1).

Для початку оцінимо вплив вікової категорії клієнта на рівень «виживання». Візуальне відображення побудованої моделі можна побачити на рис. 3.2.

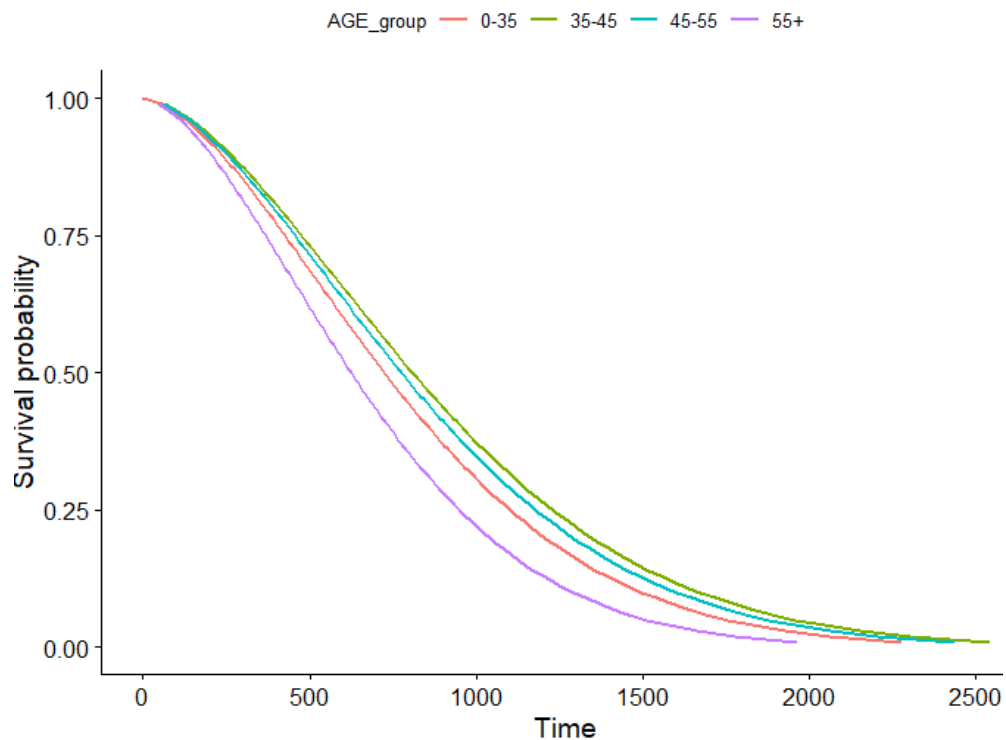


Рис. 3.2 Візуалізація кривих виживання для різних вікових груп

Джерело: побудовано автором

На рис. 3.2 можна спостерігати, що найвище розташована крива виживання вікової категорії клієнтів 35-45 років, що свідчить про їх довший рівень співпраці з банком. Поруч з цією кривою знаходиться також категорія клієнтів 45-55 років. Проаналізувавши результати моделювання та візуалізації, можна стверджувати, що цільовою категорією залучення клієнтів з автокредитами для банку є клієнти 35-55 років. І навпаки, клієнти у віці вище 55 років становлять високий ризик раннього закриття. Відповідно, маркетингова активність має бути сфокусована на спробах утримання, безпосередньо, саме цієї категорії клієнтів.

Але для визначення статистичної значущості впливу кожної з вікових категорій, необхідно дослідити коефіцієнти моделі. Характеристика якості коефіцієнтів моделі виживання з урахуванням віку клієнтів подана у таблиці 3.1. Кожне значення коефіцієнту представляє зміну рівня небезпеки для кожного

рівня змінної, в порівнянні з базовим референтним значенням (Перетин), яким в даній моделі виступає вікова група «вік до 35 років».

Таблиця 3.1

Характеристика якості коефіцієнтів моделі виживання з урахуванням віку клієнтів

Змінна	Значення коефіцієнту	Стандартна похибка	z-статистика	p-значення
Перетин	6.8070	0.0188	361.45	< 2e-16
Вік 35-45 років	0.1100	0.0236	4.67	3.1e-06
Вік 45-55 років	0.0675	0.0253	2.67	0.0076
Вік старше 55 років	-0.1479	0.0277	-5.35	8.9e-08

Джерело: побудовано автором

Всі коефіцієнти вікових груп є статистично значущими, так як їхні р-значення нижче прохідного 0.05. Існують вагомі докази того, що клієнти віком у діапазоні 35-55 років мають нижчий рівень ризику закриття, ніж ті, хто має вік до 35 років. Клієнти старші 55 років мають вищий рівень ризику, ніж клієнти до 35 років.

Оцінка параметру масштабу встановилася на рівні 0.605 (3.1). Менший параметр масштабу вказує на більш круту криву виживання, тоді як вищий параметр масштабу вказує на більш пологую. Параметр масштабу 0.605 вказує на те, що функція небезпеки відносно крута – ймовірність події (закриття клієнта) швидко зростає з часом. Це можна інтерпретувати як наявність відносно високого ризику закриття клієнта на ранньому етапі його життя.

Показник Лог-подібності побудованої моделі дорівнює -38940.3, що вище за рівень Лог-подібності моделі без врахування впливу змінних предикторів (-38994.6); р-значення для всієї моделі = 2.1e-23 є меншим за порогове 0.05. Відповідно, існує статистично значущий вплив факторів моделі на рівень «виживання» клієнтів.

Далі було оцінено вплив рівня доходів клієнта на рівень «виживання». Візуальне відображення побудованої моделі можна побачити на рис. 3.3. На рис. 3.3 можна спостерігати, що найвище розташована крива виживання категорії

0-35 тис. грн в місяць та 35-70 тис. грн в місяць, що свідчить про їх довший період співпраці з банком.

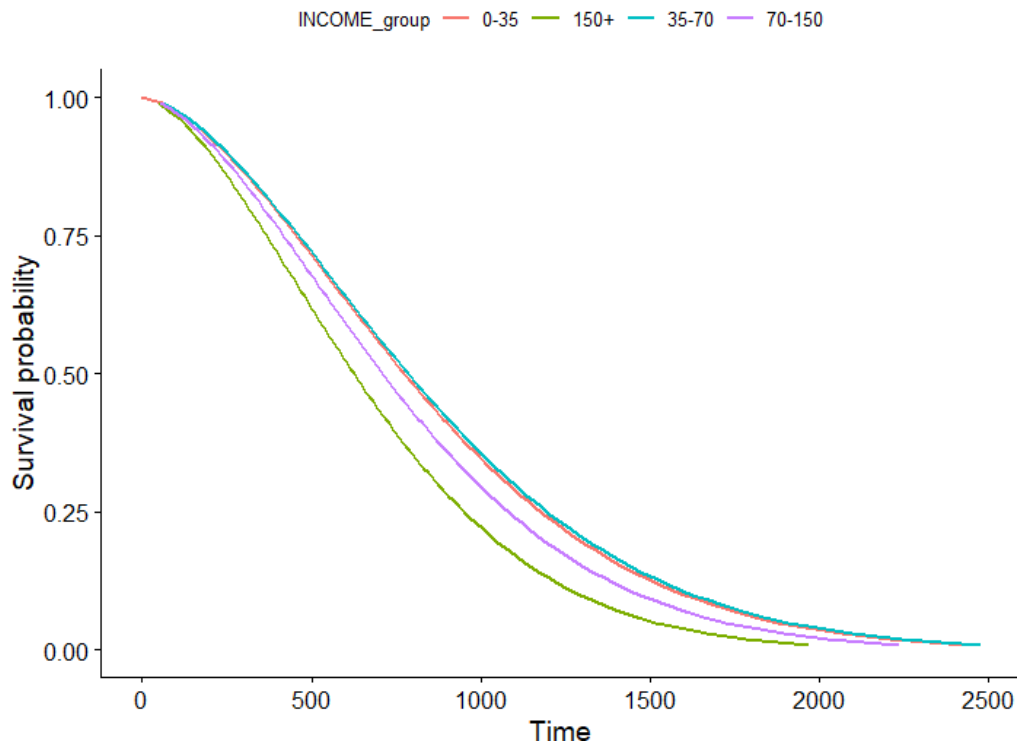


Рис. 3.3 Візуалізація кривих виживання для різних груп доходів

Джерело: побудовано автором

Проаналізувавши результати візуалізації можна стверджувати, що цільовою категорією залучення клієнтів з автокредитами для банку є клієнти з доходом до 70 тис грн в місяць. Іншим важливим висновком, який можна зробити після аналізу побудованої моделі, це те що клієнти з доходом вище 70 тис. грн на місяць становлять високий ризик раннього закриття. Відповідно, маркетингова активність має бути сфокусована на спробах утримання, безпосередньо, саме цієї категорії клієнтів.

Характеристика якості коефіцієнтів моделі виживання з урахуванням доходу клієнтів подана у таблиці 3.2.

Перетин представляє базовий рівень небезпеки для референтної групи предиктора, яким є група з доходом менше 35 тис. грн в місяць. Всі коефіцієнти вікових груп, окрім групи 35-70 тис. грн в місяць, є статистично значущими, так як їхні р-значення нижче прохідного 0.05. Це означає, що існують вагомі докази

того, що клієнти з доходом вище 70 тис. грн в місяць мають вищий рівень ризику закриття, ніж ті, хто має дохід до 35 тис. грн в місяць.

Таблиця 3.2

Характеристика якості коефіцієнтів моделі виживання з урахуванням доходу клієнтів

Змінна	Значення коефіцієнту	Стандартна похибка	z-статистика	p-значення
Перетин	6.8735	0.0122	562.22	< 2e-16
Дохід 35-70 тис. грн місяць	0.0156	0.0224	0.70	0.48449
Дохід 70-150	-0.0856	0.0237	-3.60	0.00031
Дохід вище 150 тис. грн місяць	-0.2137	0.0298	-7.16	7.9e-13

Джерело: побудовано автором

Оцінка параметру масштабу встановилася на рівні 0.606. Показник Лог-подібності побудованої моделі дорівнює -38964.4, що вище за рівень Лог-подібності моделі без врахування впливу змінних предикторів (- 38994.6). p-значення для всієї моделі = 4.9×10^{-13} є меншим за порогове 0.05, що в свою чергу означає статистично значущий вплив факторів моделі на рівень «виживання» клієнтів.

Ще одна модель буде стосуватися впливу регіону проживання на рівень «виживання» клієнтів. Візуальне відображення побудованої моделі можна побачити на рис. 3.4.

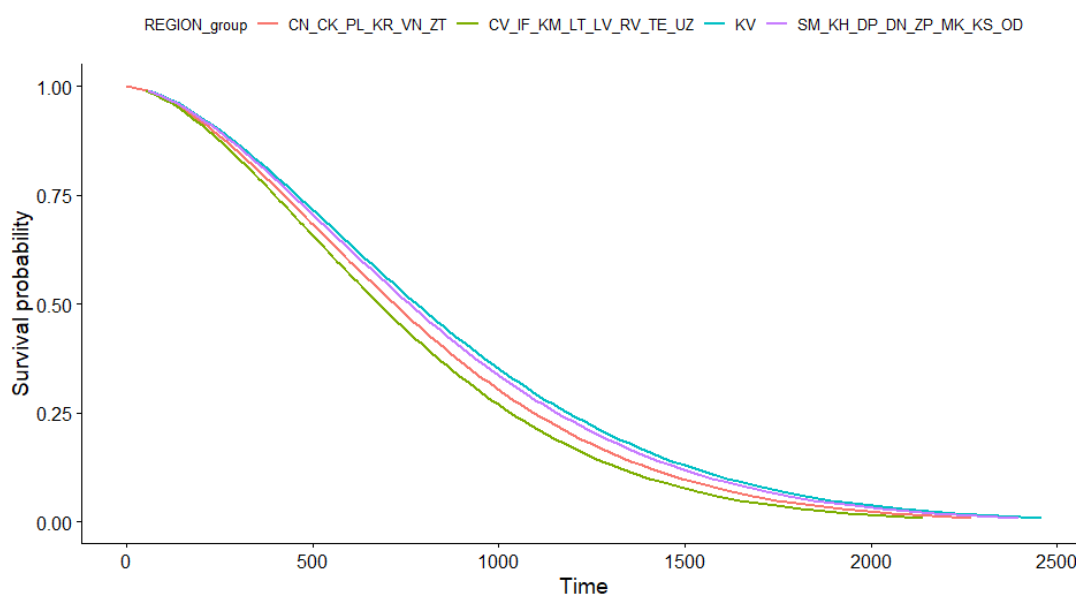


Рис. 3.4 Візуалізація кривих виживання для різних регіонів проживання

Джерело: побудовано автором

На рис. 3.4 можна спостерігати, що криві виживання різних регіонів розташовані майже поруч один до одного. Це може свідчити про відсутність статистичної різниці в ризику між різними регіонами країни. Найвище розташовані криві виживання клієнтів, які проживають в Києві та Східних регіонах, що свідчить про їх довший період співпраці з банком. Західний регіон характеризується найвищим ризиком передчасного закриття клієнтів.

Характеристика якості коефіцієнтів моделі виживання з урахуванням регіону проживання клієнтів подана у таблиці 3.3.

Таблиця 3.3

**Характеристика якості коефіцієнтів моделі виживання з урахуванням
регіону проживання**

Змінна	Значення коефіцієнту	Стандартна похибка	z-статистика	p-значення
Перетин	6.8007	0.0239	284.99	< 2e-16
Західний регіон	-0.0589	0.0328	-1.79	0.0727
Східний регіон	0.0551	0.0284	1.94	0.0525
Київ	0.0801	0.0273	2.94	0.0033

Джерело: побудовано автором

Перетин представляє базовий рівень небезпеки для референтної групи предиктора, яким є центральний регіон (cn_ck_pl_kr_vn_zt). Лише Київ (KV) показав статистично значущий коефіцієнт. Це означає, що існують вагомі докази того, що клієнти, які проживають в Києві, мають нижчий рівень ризику закриття, ніж ті, хто живе в центральних регіонах країни.

Оцінка параметру масштабу встановилася на рівні 0.607. Показник Лог-подібності побудованої моделі дорівнює -38979.2, що вище за рівень Лог-подібності моделі без врахування впливу змінних предикторів (- 38994.6). p-значення для всієї моделі = 9.3e-07 є меншим за порогове 0.05, що в свою чергу означає статистично значущий вплив факторів моделі на рівень «виживання» клієнтів.

Останнім фактором, який буде досліджений, є стать клієнта та її вплив на рівень «виживання». Візуальне відображення побудованої моделі можна побачити на рис. 3.5.

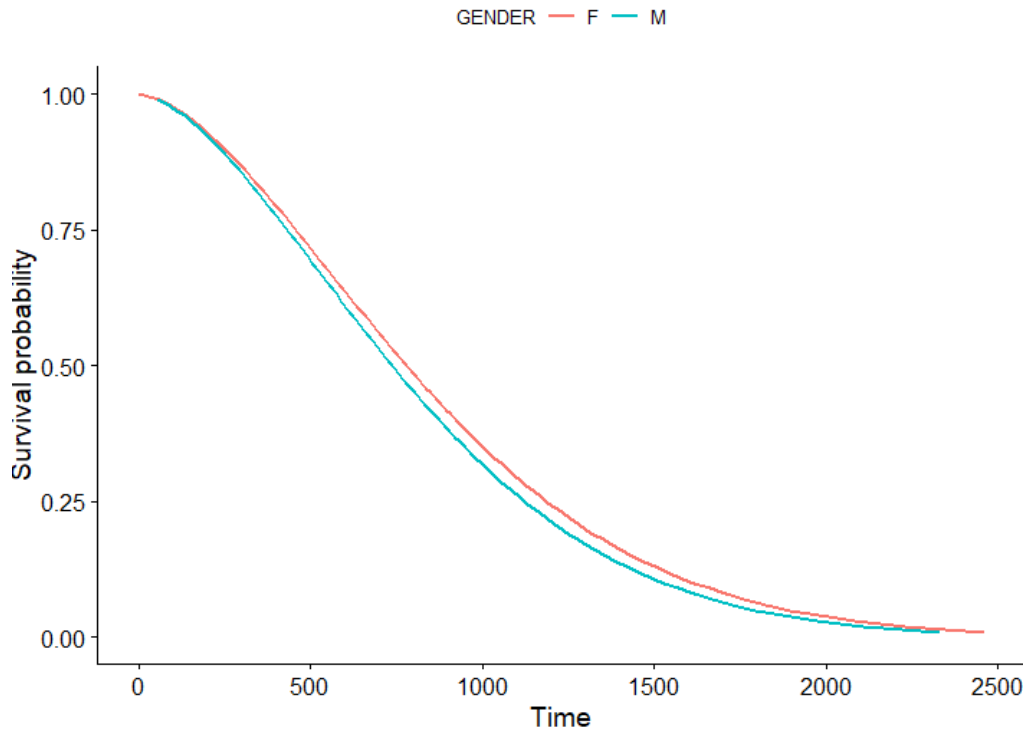


Рис. 3.5 Візуалізація кривих виживання в залежності від статі клієнта

Джерело: побудовано автором

На рис. 3.5 можна спостерігати, що крива виживання для жінок і чоловіків розташовані майже поруч один до одного. Це може свідчити про незначну відмінність коефіцієнта моделі від нуля.

Характеристика якості коефіцієнтів моделі виживання з урахуванням статі клієнта подана у таблиці 3.4.

Таблиця 3.4

Характеристика якості коефіцієнтів моделі виживання з урахуванням статі клієнта

Змінна	Значення коефіцієнту	Стандартна похибка	z-статистика	p-значення
Перетин	6.8802	0.0149	460.91	< 2e-16
Чоловік	-0.0550	0.0182	-3.03	0.0025

Джерело: побудовано автором

Перетин представляє базовий рівень небезпеки для референтної групи предиктора, яким є Жіноча стать. Чоловіча стать показала статистично значущий коефіцієнт. Це означає, що існують вагомі докази того, що Чоловіки мають вищий рівень ризику передчасного закриття, ніж Жінки. Оцінка параметру масштабу встановилася на рівні 0.608.

Показник Лог-подібності побудованої моделі дорівнює -38990, що вище за рівень Лог-подібності моделі без врахування впливу змінних предикторів (-38994.6). р-значення для всієї моделі = 0.0024 є меншим за порогове 0.05, що в свою чергу означає статистично значущий вплив факторів моделі рівень «виживання» клієнтів.

Зважаючи на неочевидні характеристики моделей статі та регіону, було вирішено обчислити сумісний вплив соціально демографічних характеристик клієнта на рівень «виживання» (таблиця 3.5).

Таблиця 3.5

Характеристика якості коефіцієнтів моделі виживання з урахуванням усіх соціально-демографічних характеристик клієнта

Змінна	Значення коефіцієнту	Стандартна похибка	z-статистика	р-значення
Перетин	6.8337	0.0328	208.5	< 2e-16
Західний регіон	-0.0495	0.0326	-1.52	0.12898
Київ	0.0900	0.0273	3.3	0.00097
Східний регіон	0.0645	0.0282	2.29	0.02219
Чоловік	-0.0350	0.0182	-1.93	0.05388
Вік 35-45 років	0.0968	0.0235	4.11	3.90e-05
Вік 45-55 років	0.0449	0.0254	1.77	0.07752
Вік старше 55 років	-0.1691	0.0280	-6.05	1.50e-09
Дохід вище 150 тис. грн в місяць	-0.2425	0.0302	-8.03	9.60e-16
Дохід 35-70 тис. грн в місяць	-0.0174	0.0226	-0.77	0.44004
Дохід 70-150 тис. грн в місяць	-0.1102	0.0240	-4.59	4.40e-06

Джерело: побудовано автором

Перетин представляє базовий рівень небезпеки для референтної групи предиктора, якою в даній моделі є Жінки, віком до 35 років, з доходом до 35 тис. грн в місяць, які проживають в центральному регіоні країни. Статистично значущими тепер виявилися ознаки: Київ та Східний регіон, вік 35-45 років та вище 55 років, дохід вище 70 тис. грн в місяць.

Оцінка параметру масштабу встановилася на рівні 0.601. Показник Лог-подібності побудованої моделі дорівнює -38887.8, що вище за рівень Лог-подібності моделі без врахування впливу змінних предикторів (-38994.6). р-значення для всієї моделі = 2.3e-40 є меншим за порогове 0.05, що в свою чергу означає статистично значущий вплив факторів моделі на рівень «виживання»

клієнтів. Для порівняння якості побудованих моделей було використано критерій AIC (таблиця 3.6).

Таблиця 3.6

Значення коефіцієнта AIC для усіх побудованих моделей

Модель	Значення AIC
Вплив віку	77890.51
Вплив доходу	77938.80
Вплив регіону проживання	77968.35
Вплив статі	77985.91
Загальна модель	77799.56

Джерело: побудовано автором

За критерієм AIC, найкращою моделлю є модель з урахуванням впливу усіх соціально демографічних характеристик клієнта. Але в моделі присутні групи клієнтів, які не несуть статистичного впливу на рівень виживання. Відповідно, цілком імовірно, що гарні показники «Загальної» моделі за критерієм AIC були досягнені за рахунок перенавчання моделі. Для виправлення можливої проблеми перенавчання було застосовано алгоритм stepAIC. Результати алгоритму подані в таблиці 3.7.

Таблиця 3.7

Характеристика якості коефіцієнтів моделі виживання з урахуванням усіх статистично значущих характеристик клієнта

Змінна	Значення коефіцієнту	Стандартна похибка	z-статистика	p-значення
Перетин	6.8642	0.0242	283.71	< 2e-16
Дохід до 35 тис. грн в місяць	0.1093	0.0240	4.56	5.2e-06
Дохід 35-70 тис. грн в місяць	0.0933	0.0278	3.36	0.00079
Дохід вище 150 тис. грн в місяць	-0.1300	0.0339	-3.83	0.00013
Західний регіон	-0.1286	0.0247	-5.20	2.0e-07
Центральний регіон	-0.0789	0.0258	-3.06	0.00219
Жінка	0.0370	0.0181	2.04	0.04086
Вік до 35 років	-0.0959	0.0235	-4.08	4.6e-05
Вік 45-55 років	-0.0526	0.0223	-2.36	0.01846
Вік старше 55 років	-0.2675	0.0251	-10.64	< 2e-16

Джерело: побудовано автором

Перетин представляє базовий рівень небезпеки для референтної групи предиктора, якою в даній моделі є Чоловіки віком 35-45 років, з доходом 70-150 тис. грн в місяць, які проживають в Києві або Східному регіоні.

Згідно результатів моделювання можна стверджувати, що клієнти жіночої статі, які проживають в Києві чи в Східних регіонах, віком 35-45 років, та доходом до 70 тис грн в місяць статистично мають довший життєвий період.

Найсильнішим впливом на рівень небезпеки (найсильніша різниця порівняно з нулем) відзначилися характеристики: вік більше 55 років (-0.27), західний регіон проживання (-0.13), дохід вище 150 тис. грн в місяць (-0.13). Клієнти цієї групи є джерелом найвищого ризику передчасного закриття. Відповідно, маркетингова активність має бути сфокусована на спробах утримання, безпосередньо, саме цієї категорії клієнтів.

Оцінка параметру масштабу встановилася на рівні 0.601. Показник Лог-подібності побудованої моделі дорівнює -38888.6, що вище за рівень Лог-подібності моделі без врахування впливу змінних предикторів (- 38994.6). р-значення для всієї моделі = $9.8e-41$ є меншим за порогове 0.05, що в свою чергу означає статистично значущий вплив факторів моделі рівень «виживання» клієнтів.

Показник АІС для оптимізованої моделі встановився на рівні 77799.11, що нижче показника АІС для «Загальної» моделі (77799.56). Відповідно, відбулося покращення якості побудови та навчання моделі виживання.

Результати аналізу коефіцієнтів та дослідження кривих виживання побудованих моделей дозволили зробити висновки, що усі характеристики, які досліджувалися, суттєво впливають на рівень «виживання» клієнтів. Підтвердженням цієї ідеї також стало найменше значення показника АІС для моделі, яка включала усіх фактори одночасно. Показник масштабу для всіх побудованих моделей встановився на рівні 0.6, що є доволі малим значенням. Це вказує на значний ризик закриття клієнтів в короткостроковій перспективі після відкриття, і ще раз підтверджує необхідність пошуку факторів, які забезпечують продовження періоду співпраці клієнта з банком, або ж виділення груп клієнтів, які несуть в собі високий ризик передчасного закриття.

Отримані результати можуть бути корисні для банківських організацій при розробці тарифів та продуктових пропозицій, для залучення цільової аудиторії

клієнтів. Також вони можуть бути використані при визначенні стратегії розвитку банку та управлінні його портфелем клієнтів. Для групи найбільш схильних до швидкого закриття клієнтів можуть бути розроблені постійні комерційні кампанії, направлені на утримання їх активності після погашення кредиту.

Другим етапом дослідження буде аналіз поведінки клієнтів з відсотковою ставкою по автокредиту на рівні 0.01%. Для виконання цього завдання було обрано регресійну модель виживання Кокса (3.2). Для клієнтів з лояльною відсотковою ставкою бізнес модель вибудована таким чином, щоб клієнт максимально швидко повернув кредит. Відповідно всі соціально – демографічні показники не впливають на життя клієнта. Основним важелем є продаж додаткових продуктів банку. Як вже зазначалося в пункті 3.1, основним додатковим продуктом, який пропонується клієнтам з автокредитом, є кредитна картка. Отже, для початку необхідно оцінити вплив продажу кредитних карток на рівень виживання клієнтів.

Візуальне представлення побудованої моделі можна побачити на рис. 3.6, де верхня пунктирна лінія – це крива виживання для клієнтів з кредитною картою, а нижня суцільна – для клієнтів без кредитної картки.

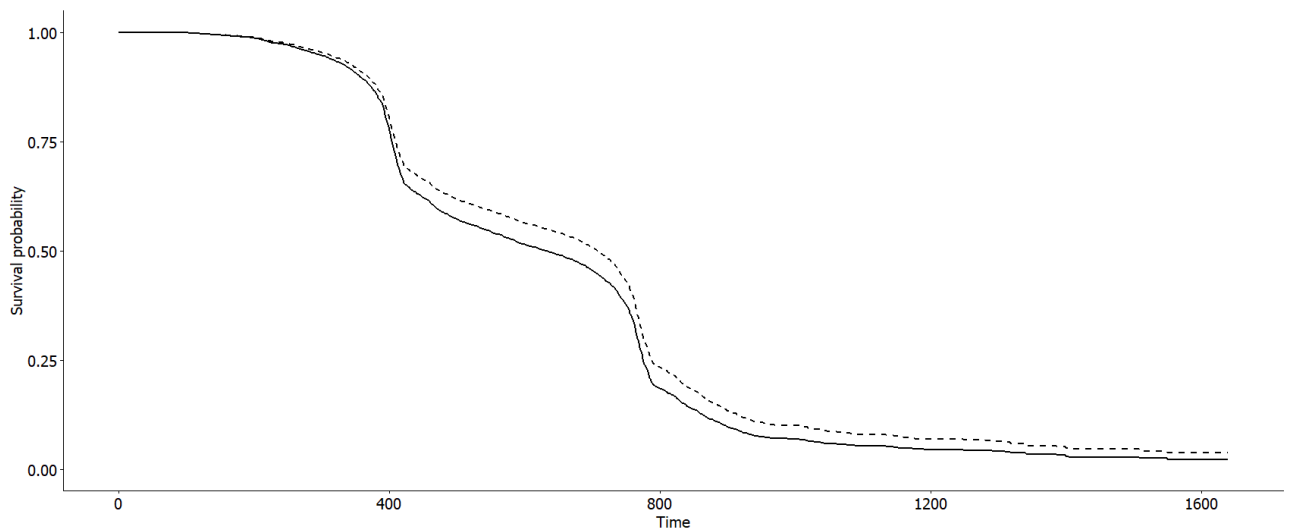


Рис. 3.6 Візуалізація кривої виживання для клієнтів з кредитною картою

Джерело: побудовано автором

З графіка видно, що клієнти з кредитною картою статистично мають більш тривалий період активності. Загалом, чим вища крива виживання клієнтів

з кредитними картками, тим значніший вплив цієї змінної на рівень ризику закриття клієнта. На рис. 3.6 відстань між кривими є незначною, що свідчить про незначний вплив. Також можна відзначити, що відстань між кривими виживання клієнтів з кредитною карткою та без неї зменшується до мінімуму в різні періоди.

Візуальний аналіз графіка не дозволяє виключити можливий перетин кривих, що означає порушення припущення про пропорційність ризиків у часі.

Отже, необхідний більш ретельний статистичний аналіз якості побудованої моделі. Статистична оцінка побудованої моделі представлена в таблиці 3.8 (3.3, 3.4).

Таблиця 3.8

Результати моделювання впливу кредитної картки на виживання клієнта

Характеристики якості моделі	Значення
β	-0.14890
HR	0.86166
HR-	1.161
CI 95% Lower	0.807
CI 95% Upper	0.9201
$se(\beta)$	0.03346
w	-4.45
$p - value$	<8.6e-06

Джерело: побудовано автором

У цій моделі $\beta = -0.15$, що означає, що наявність кредитної картки зменшує ризик закриття клієнта. Коефіцієнт небезпеки (HR) відображає ризик відтоку для другої групи ($credit_card = 1$) по відношенню до першої групи ($credit_card = 0$). У нашому випадку $HR = 0.86$, що означає, що клієнти з кредитною карткою закриваються з швидкістю 0.86 по відношенню до швидкості закриття клієнтів без кредитної картки. З іншого боку, якщо ми розглянемо $HR = 1.16$, ми можемо зробити висновок, що клієнти без кредитної картки закривають договори з швидкістю 1.16 відносно швидкості закриття угод клієнтами з кредитною карткою.

Для коефіцієнтів HR верхнє значення довірчого інтервалу (CI 95% Upper = 0.92) близьке до 1, що вказує на незначний статистичний вплив наявності кредитної картки на рівень виживання клієнтів. Результати в Таблиці 3.8

показують, що змінна `credit_card` має статистично значущий коефіцієнт регресії Кокса, оскільки $p - value < 0.05$.

Перевірка припущення про пропорційність ризиків є останнім кроком у перевірці якості побудованої моделі. На рис. 3.7 показано графік залишків Шенфельда з часом для змінної `credit_card`.

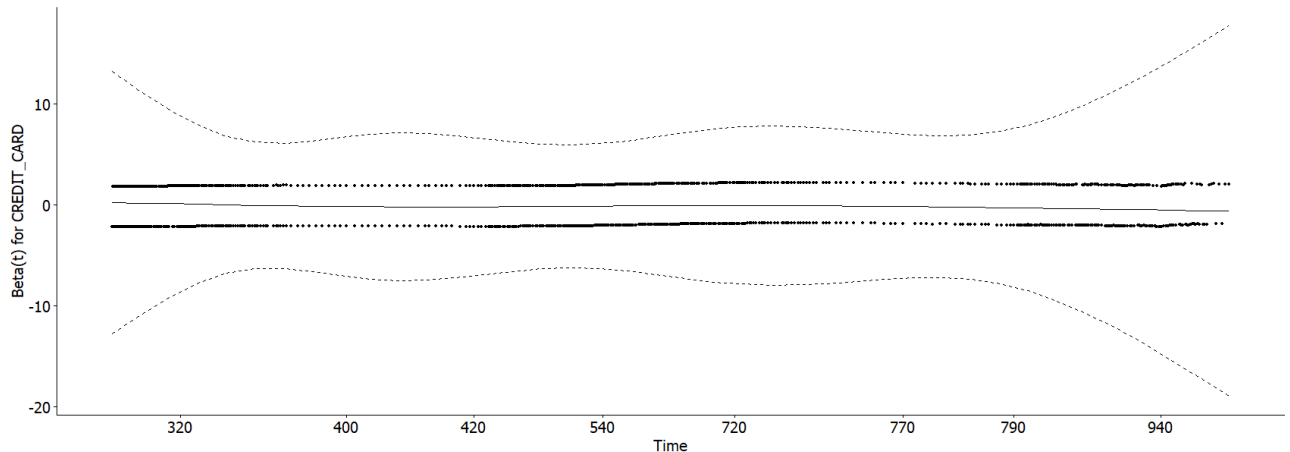


Рис 3.7 Візуалізація залишків Шенфельда для змінної `credit_card`

Джерело: побудовано автором

На рис. 3.7 видно ступінчасту суцільну лінію, яка пояснює попереднє відхилення припущення про пропорційність небезпек. Значення статистичної оцінки зв'язку між залишками Шенфельда та часом - глобальний тест Шенфельда встановився на рівні $p = 0.0008$. Тест є статистично значущим для змінної `credit_card`, оскільки $p - value < 0.05$. Отже, наше припущення про пропорційні ризики для цієї змінної відхиляється. Відповідно, постає завдання знайти фактор, вплив якого суттєво не змінюється з часом.

Для цього розглянемо не просто наявність кредитної картки, а оцінити її вплив, за умови активації. Активація означає, що клієнт скористався кредитною карткою хоча б один раз. Візуальне представлення побудованої моделі можна побачити на рис. 3.8, де верхня пунктирна лінія – крива виживання для клієнтів з активованою кредитною карткою, а нижня суцільна лінія – для клієнтів без активованої кредитної картки. З рис. 3.8 видно, що активна кредитна картка значно уповільнює процес закриття клієнта, тим самим подовжуючи період співпраці для значної частини клієнтів.

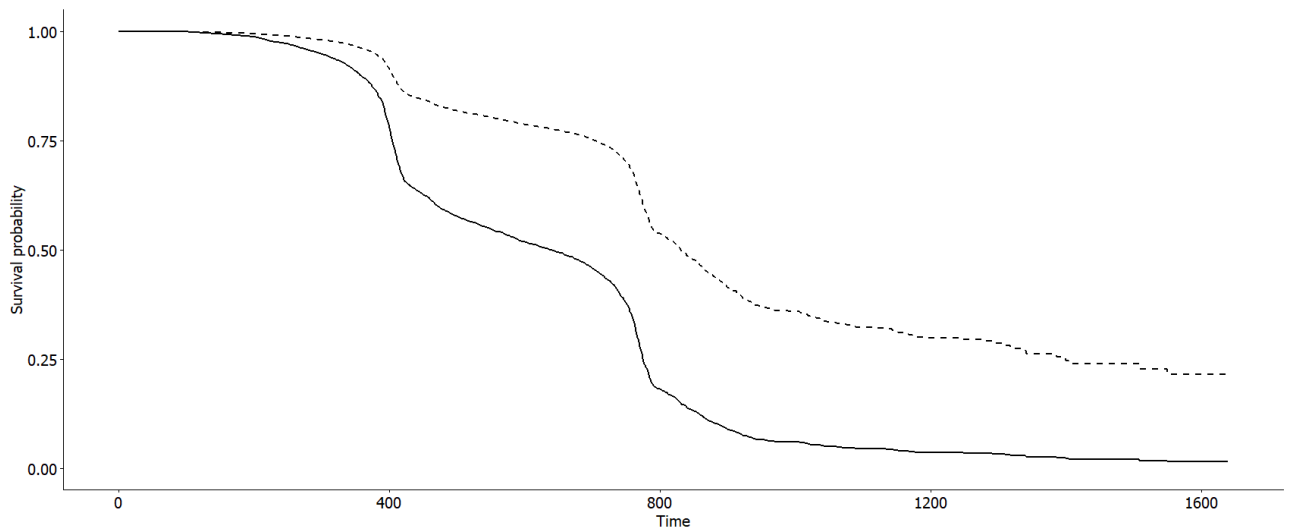


Рис 3.8 Візуалізація кривої виживання для клієнтів з активованою кредитною картою

Джерело: побудовано автором

Про це свідчить значна відстань між кривими, що вказує на значний вплив активних кредитних карток на виживання клієнтів. Також можна відзначити, що відстань між кривими виживання для клієнтів з активною кредитною картою та без неї змінюється з часом, але криві ніколи не перетинаються. Така ситуація свідчить про можливе порушення припущення про пропорційність ризиків. Тому необхідно провести статистичний аналіз якості побудованої моделі. Статистична оцінка побудованої моделі представлена в таблиці 3.9.

Таблиця 3.9

Результати моделювання впливу активованої кредитної картки на виживання клієнта

Характеристики якості моделі	Значення
β	-1.01209
HR	0.36346
HR-	2.751
CI 95% Lower	0.3125
CI 95% Upper	0.4227
$se(\beta)$	0.07702
w	-13.14
$p - value$	<2e-16

Джерело: побудовано автором

У цій моделі $\beta = -1.01$, що означає, що наявність активованої кредитної картки зменшує ризик закриття клієнта. Коефіцієнт небезпек (HR) відображає

ризик відтоку для другої групи ($\text{active_credit_card} = 1$) по відношенню до першої групи ($\text{active_credit_card} = 0$). У нашому випадку $\text{HR} = 0.36$, що означає, що клієнти з активованою кредитною картою закривають угоди з швидкістю 0.36 по відношенню до швидкості закриття договорів клієнтами без активованої кредитної картки.

З іншого боку, якщо ми розглянемо $\text{HR-} = 2.75$, ми можемо зробити висновок, що клієнти без активованої кредитної картки закривають договори з швидкістю 2.75 відносно швидкості закриття контрактів клієнтами з активованою кредитною картою.

Для коефіцієнтів HR верхнє значення довірчого інтервалу ($\text{CI } 95\% \text{ Upper} = 0.42$) є далеким від 1, що вказує на значний статистичний вплив наявності активованої кредитної картки на рівень виживання клієнтів. Результати в таблиці 3.14 показують, що змінна $\text{active_credit_card}$ має статистично значущий коефіцієнт регресії Кокса, оскільки $p\text{-value} < 0.05$.

Перевірка припущення про пропорційність ризиків є останнім кроком у дослідженні якості побудованої моделі. На рис. 3.9 показано графік залишків Шенфельда з часом для змінної $\text{active_credit_card}$.

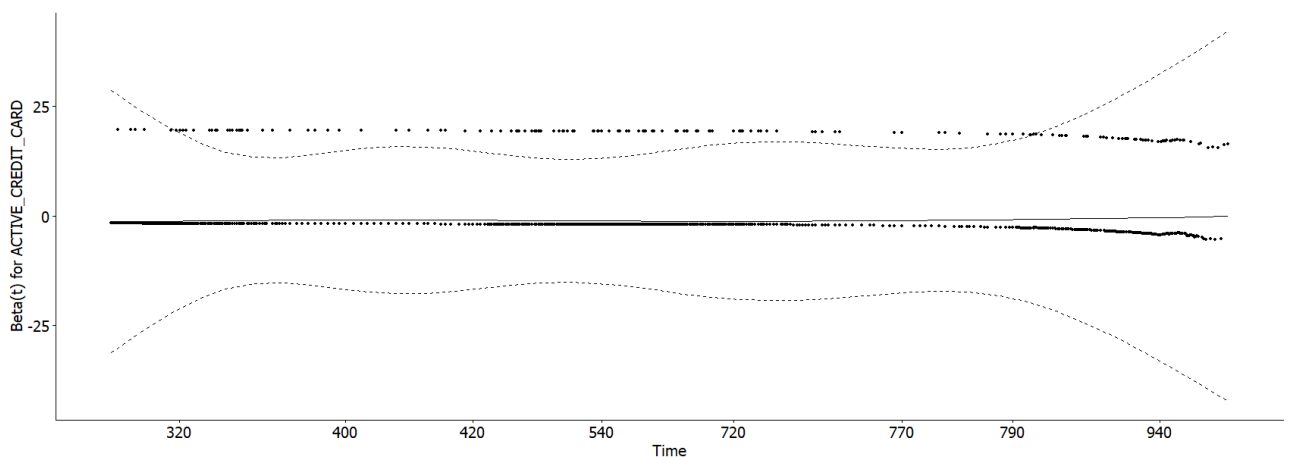


Рис 3.9 Візуалізація залишків Шенфельда для змінної $\text{active_credit_card}$

Джерело: побудовано автором

На рис. 3.9 видно ступінчасту суцільну лінію, що частково пояснює попереднє відхилення припущення про пропорційність ризиків. Проте, водночас, статистична оцінка зв'язку між залишками Шенфельда та часом -

глобальний тест Шенфельда встановилася на рівні $p = 0.2135$. Тест не є статистично значущим для змінної `active_credit_card`, оскільки $p - value < 0.05$.

Отже, припущення про пропорційні ризики не відхиляється для цієї змінної. Відповідно, побудована модель відповідає всім загальноприйнятим критеріям якості регресійної моделі Кокса.

Результати моделювання підтверджують гіпотезу про те, що продаж кредитної картки дозволяє продовжувати активність клієнтів банку, першим продуктом яких був автокредит. Однак критично важливим фактором успіху є продаж кредитної картки та її подальша активація. Активація картки дає якісне, статистично підтверджене твердження, що кредитна картка продовжує життя клієнта.

Отримані результати дослідження можуть бути використані для оптимізації маркетингової стратегії банку, забезпечення більш ефективного утримання клієнтів та збільшення прибутку банку. Подальші дослідження можуть бути зосереджені на вивченні причин активації кредитних карток.

Лістинг R-коду для реалізації всіх етапів моделювання в підрозділі 3.2 подано в додатку Д.

3.3 Удосконалення маркетингового менеджменту в банку шляхом побудови маркетингової сегментації бізнес-клієнтів

Основою дослідження є дані про активність користувачів поточних рахунків у євро за період з квітня 2020 року по квітень 2021 року. Вибірка включає 4 500 корпоративних бізнес-клієнтів українського банку (фізичних осіб-підприємців та юридичних осіб). Повний перелік полів вихідних даних наведено в таблиці 3.10.

Як зазначалося вище у пункті 2.1 першим кроком у здійсненні кластеризації корпоративних клієнтів банку є визначення оптимальної кількості груп схожості.

Таблиця 3.10

Характеристика полів вибірки вихідних даних

Змінні моделі	Опис
CL_ID	Внутрішній ідентифікатор клієнта
NOM_nb	Кількість місяців, коли на рахунку клієнта в євро був позитивний баланс
Oper_nb	Кількість місяців, коли клієнт здійснював операції з євро
NIM_nb	Кількість місяців, коли чистий маржинальний дохід клієнта був від'ємним (тобто чистий маржинальний збиток)
NBI_nb	Кількість місяців, коли чистий банківський дохід для цього клієнта був від'ємним (тобто чистий банківський збиток).

Джерело: побудовано автором

У таблиці 3.11 наведено результати розрахунку 26 критеріїв оцінки в пакеті NbClust (оптимальні значення для кожного критерію виділено).

Таблиця 3.11

Значення розрахованих індексів у пакеті NbClust

Індекс	Кількість кластерів								
	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

Джерело: побудовано автором

У таблиці 3.12 наведено результати «голосування» за оптимальну кількість кластерів.

Таблиця 3.12

Результати «голосування» за оптимальну кількість кластерів

Кількість кластерів	2	3	4	5	7	9	10
Кількість індексів, які проголосували за оптимум	4	1	3	4	8	1	2

Джерело: побудовано автором

Більшість індексів проголосували за 7 кластерів. Далі було розраховано середні значення параметрів моделі для оптимальної кількості кластерів (таблиця 3.13).

Таблиця 3.13

Середні характеристики кластерів при розподілі клієнтів на сім сегментів

Номер кластеру	Кількість клієнтів	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

Джерело: побудовано автором

Кластери 4 і 5 характеризують одну і ту саму групу клієнтів. Єдина відмінність полягає в інтенсивності використання рахунків у євро. Через це було вирішено реалізувати алгоритм лише з шістьма групами. Характеристики розподілу на шість сегментів представлені в таблиці 3.14.

Після аналізу активності клієнтів кожен маркетинговий сегмент можна коротко охарактеризувати наступним чином:

- **Кластер 1:** недохідні транзактори з накопиченням євро;
- **Кластер 2:** збиткові накопичувачі євро;
- **Кластер 3:** прибуткові накопичувачі євро;
- **Кластер 4:** прибуткові транзактори з євронакопиченням;

- **Кластер 5:** клієнти без євронакопичення;
- **Кластер 6:** прибуткові транзактори без євронакопичення.

Таблиця 3.14

Середні характеристики кластерів при розподілі клієнтів на шість сегментів

Номер кластеру	Кількість клієнтів	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

Джерело: побудовано автором

Щоб полегшити характеристику кожного визначеного маркетингового сегмента та надати персоналізовані рекомендації, щодо зменшення операційних витрат банку, дані про активність клієнтів у кожному сегменті були відображені у вигляді бокс-діаграми (рис. 3.10-3.15).

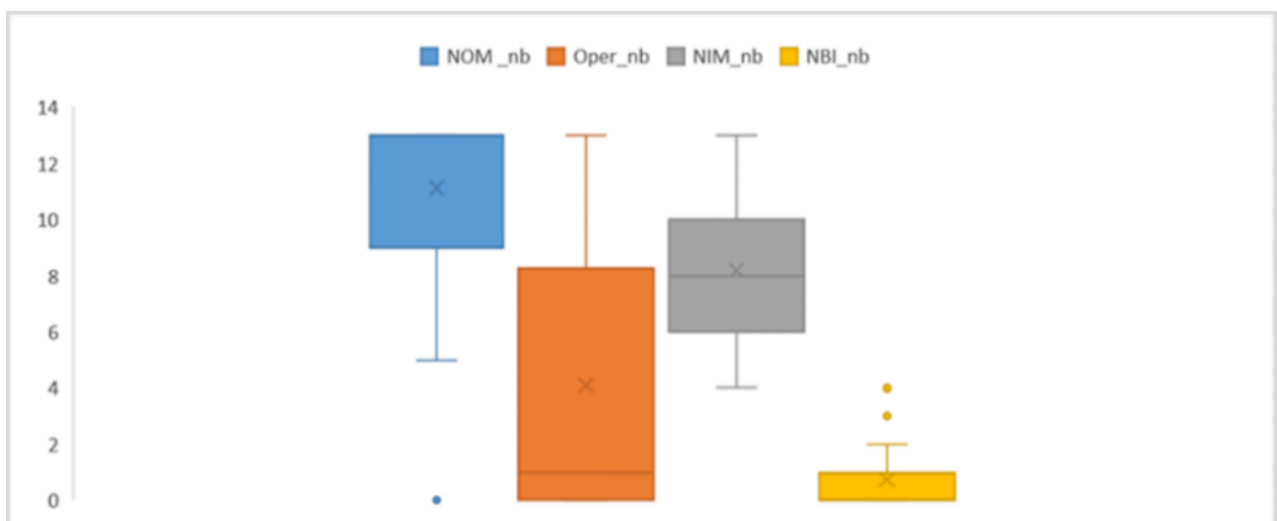


Рис 3.10 Активність у першому сегменті корпоративних клієнтів

Джерело: побудовано автором

На рис. 3.10 відображено активність клієнтів у кластері 1:

- постійний позитивний баланс на рахунках в євро;
- операції в євро здійснюються лише 4-6 разів на рік;

- ці клієнти демонструють чистий маржинальний збиток більшу частину часу, через недостатній залишок на рахунку в національній валюті або відсутність кредитів;
- операції в євро приносять додатковий комісійний дохід, який переважно покриває чистий маржинальний збиток.

Враховуючи вищезазначену характеристику та беручи до уваги невелику кількість клієнтів - 284 (таблиця 3.14), були розроблені наступні рекомендації для покращення прибутковості маркетингового сегменту:

- провести перемовини з клієнтами та рекомендувати купувати євро лише в разі потреби;
- запровадити комісію за обслуговування рахунків у євро в розмірі 1%.

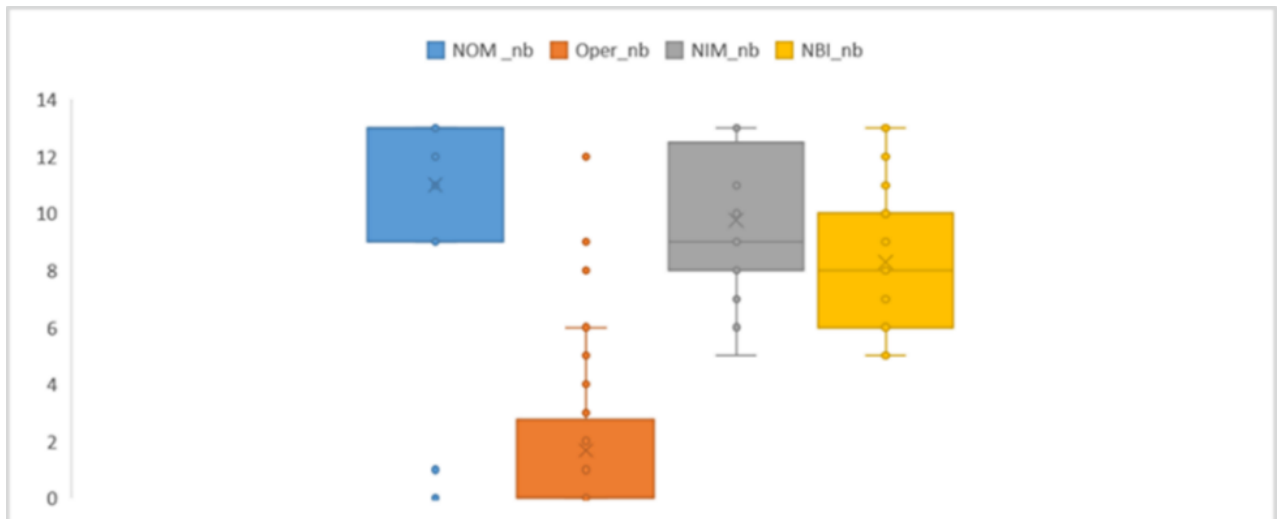


Рисунок 3.11 Активність у другому сегменті корпоративних клієнтів

Джерело: побудовано автором

На рис. 3.11 відображено активність клієнтів у кластері 2:

- постійний позитивний баланс на рахунках в євро;
- операції в євро здійснюються лише кілька разів на рік;
- ці клієнти завжди показують чистий банківський збиток.

Враховуючи вищезазначену характеристику та беручи до уваги невелику кількість клієнтів - 110 (таблиця 3.14), були розроблені наступні рекомендації для підвищення прибутковості маркетингового сегменту:

- запровадити комісію за обслуговування рахунків у євро в розмірі 1%;

- рекомендується провести опитування клієнтів для визначення причини постійного накопичення євро без подальшого використання.
- за результатами опитування пропрацювати шляхи задоволення потреб клієнта без додаткового фінансового навантаження для клієнтів та банку.

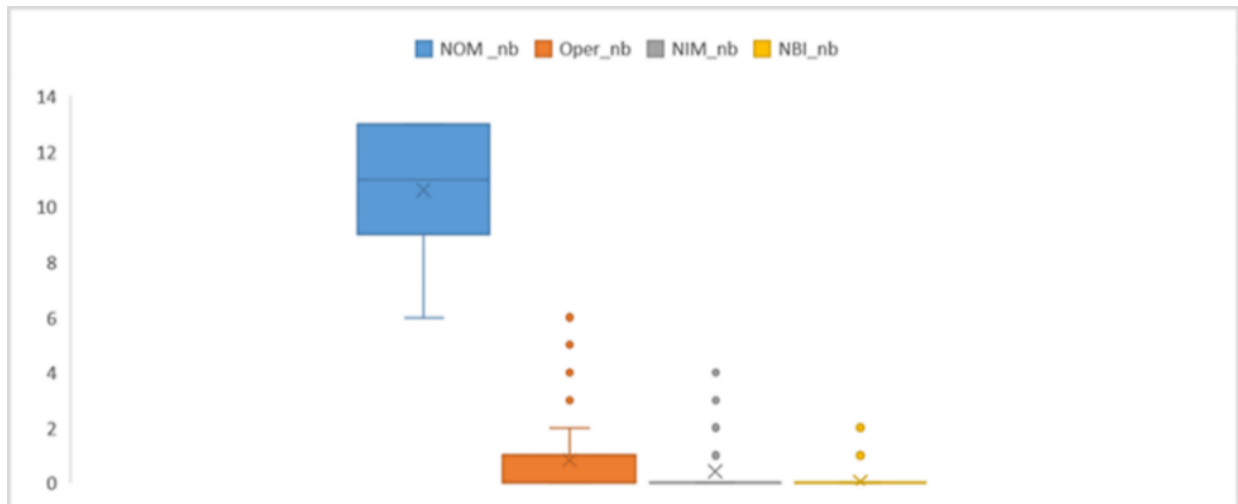


Рис 3.12 Активність у третьому сегменті корпоративних клієнтів

Джерело: побудовано автором

На рис. 3.12 відображено активність клієнтів у кластері 3:

- постійний позитивний баланс на рахунках в євро;
- операції в євро здійснюються лише кілька разів на рік;
- ці клієнти, завдяки іншим банківським продуктам, наразі є прибутковими для банку.

Зважаючи на вищезазначену характеристику та беручи до уваги велику кількість клієнтів - 1 071 (таблиця 3.14), було розроблено наступні рекомендації для підвищення прибутковості маркетингового сегменту:

- рекомендується провести опитування для визначення причини постійного накопичення євро без подальшого використання;
- провести комунікацію з клієнтами, щодо можливого запровадження 1% комісії за обслуговування євторахунків на другому етапі скорочення операційних витрат. Відповідно, дати час клієнтам на підготовку до цього впровадження.
- за результатами опитування пропрацювати шляхи задоволення потреб клієнта без додаткового фінансового навантаження для клієнтів та банку.

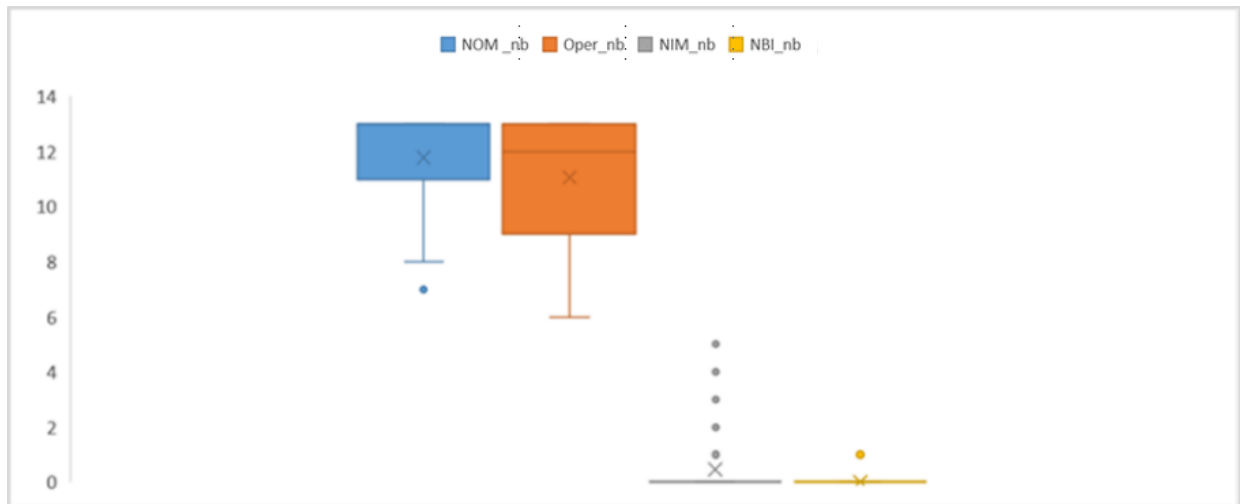


Рис. 3.13 Активність у четвертому сегменті корпоративних клієнтів

Джерело: побудовано автором

На рис. 3.13 відображено активність клієнтів у кластері 4:

- постійний позитивний баланс на рахунках в євро;
- накопичення постійно використовуються для операцій в євро;
- поточна активність дозволяє клієнтам залишатися повністю прибутковими для банку.

Зважаючи на вищезазначену характеристику та беручи до уваги велику кількість клієнтів - 665 (таблиця 3.14), були розроблені наступні рекомендації для підвищення прибутковості маркетингового сегменту:

- провести перемовини з клієнтами та рекомендувати купувати євро лише в разі потреби;
- провести комунікацію з клієнтами, щодо можливого запровадження 1% комісії за обслуговування єврорахунків на другому етапі скорочення операційних витрат. Відповідно, дати час клієнтам на підготовку до цього впровадження.

На рис. 3.14 та рис. 3.15 відображено активність клієнтів у кластерах 5 та 6:

- кілька разів на рік демонструють позитивний баланс на єврорахунках;
- клієнти шостого сегменту зазвичай купують валюту в момент здійснення операції, що дозволяє їм не накопичувати євро на поточних рахунках;

- накопичення постійно використовується для операцій з євро (різниця між кластерами - у кількості операцій);
- поточна активність дозволяє клієнтам залишатися повністю прибутковими для банку.

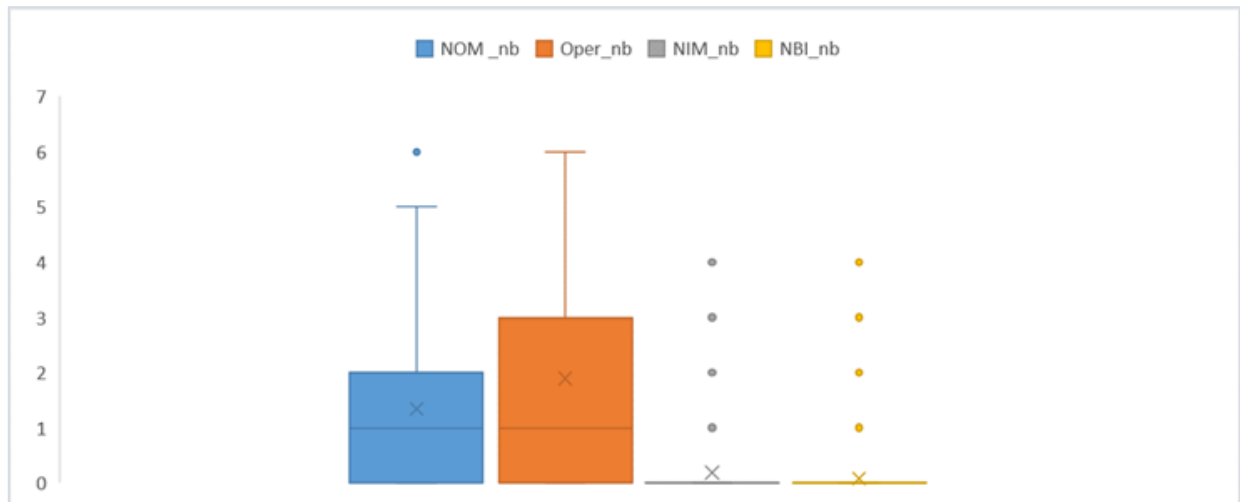


Рис. 3.14 Активність у п'ятому сегменті корпоративних клієнтів

Джерело: побудовано автором

З огляду на вищезазначену характеристику та беручи до уваги найбільш значну кількість клієнтів - 1 581 та 789 (таблиця 3.14), були розроблені наступні рекомендації для підвищення прибутковості маркетингових сегментів:

- не вводити жодних комісій за обслуговування рахунків у євро;
- моніторити подальшу поведінку клієнтів, стосовно активності з рахунками в євро, для запобігання виникнення додаткових операційних витрат.

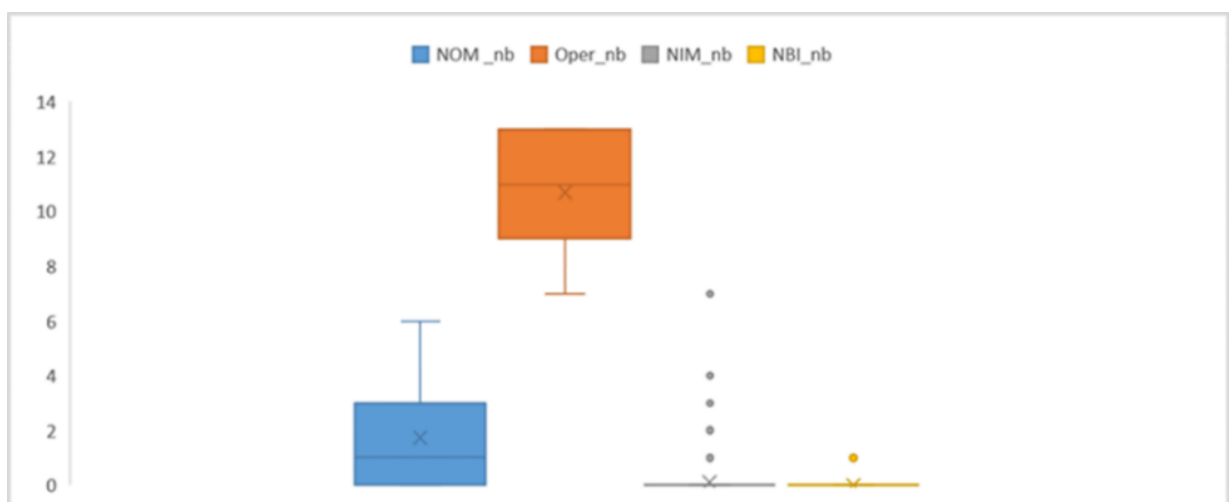


Рис. 3.15 Активність у шостому сегменті корпоративних клієнтів

Джерело: побудовано автором

Як бачимо, кластери 1 і 2 – це концентрація збиткових клієнтів, для яких необхідно терміново запровадити комісію за обслуговування. Маркетингові сегменти 3 і 4 – це клієнти, які не показують чистих збитків, але з якими необхідно працювати над підвищенням їхньої прибутковості. Крім того, щоб запобігти можливим збиткам, запровадити плату за обслуговування в майбутньому. Останні сегменти є прикладом здорового використання єврорахунків. Клієнти не накопичують євро, а купують його в момент потреби, не приносячи банку збитків. Таким клієнтам рекомендується не встановлювати додаткові комісії за обслуговування, що дозволяє економити людські ресурси, оскільки такі клієнти складають більше половини всіх користувачів єврорахунків.

Про ефективність розробленого алгоритму свідчить легкість опису та характеристики груп схожості, оскільки кожен сегмент має чітко визначені відмінності, які дозволяють розробити індивідуальні маркетингові рекомендації, щодо ціноутворення. Така сегментація та рекомендації дозволять оптимізувати операційні витрати без встановлення комісії для всіх корпоративних клієнтів банку. У першій хвилі пропонується опрацювати лише 9% клієнтів (ввести індивідуальну плату за обслуговування), що вплине на оптимізацію людських ресурсів банку.

Розроблений концептуальний підхід дозволяє банку залишатися привабливим у конкурентному середовищі, не несучи при цьому зайвих витрат. Практичні результати дослідження також можуть бути основою для побудови автоматичних маркетингових кампаній та комунікацій з індивідуальними пропозиціями для кожного маркетингового сегменту окремо. У подальших дослідженнях необхідно протестувати розширену кількість характеристик активності клієнтів для покращення якості моделі сегментації. Також на основі результатів розробки маркетингової сегментації корпоративних клієнтів банку можна розробити модель прогнозування потреб клієнтів у конкретному продукті чи послугі.

Лістинг R-коду для реалізації всіх етапів моделювання в підрозділі 3.3 подано в додатку Е.

Висновки до розділу 3

В третьому розділі дисертаційної роботи було проведено експериментальне дослідження стосовно особливостей моделювання поведінки клієнтів банку.

Обґрунтовано вигідне становище банків порівняно з іншими сферами економічної активності, за рахунок оперування великими об'ємами даних, що слугує фундаментальною складовою їхньої діяльності з розробки маркетингових стратегій.

У результаті наукового пошуку визначено, що банківська сфера піддається сильному впливу зовнішніх факторів, які можуть кардинально змінити підходи до співпраці між банком та клієнтами. Непередбачуваність стосовно взаємодії з клієнтами може виникати з багатьох джерел, таких як: регуляторні зміни, економічні та ринкові тенденції.

За результатами дослідження визначено, що два основні підходи економіко-математичного моделювання, які привернули значну увагу в цій галузі – це сегментація та моделювання виживання. Ці методи вирішують окремі, але взаємопов'язані завдання, кожна з яких сприяє досягненню головної мети – підвищенню рівня залучення, утримання та задоволеності клієнтів.

Розглянуто особливості поведінки банківських клієнтів, з фокусом на корпоративних бізнес-клієнтах з рахунками в євро та роздрібних клієнтах з автокредитом. Встановлено, що з точки зору прибутковості ці сегменти є найціннішими для банку. Однак ефективний маркетинговий менеджмент стає складним завданням через особливості їхньої поведінки. Це спонукало до розробки економіко-математичних моделей з метою покращення розуміння характеристик їх поведінки, на основі кластеризації та аналізу виживання.

Розроблено маркетингову сегментацію корпоративних клієнтів банку на основі алгоритму кластеризації k-середніх. Для забезпечення процесу кластеризації, запропоновано оптимальну кількість кластерів визначати шляхом зважування результатів розрахунку 26 індексів з пакету NbClust та бізнес-вимог банку. Під час розрахунку алгоритму було виявлено шість груп схожості: 1. недохідні транзактори з накопиченням євро; 2. збиткові накопичувачі євро; 3. прибуткові накопичувачі євро; 4. прибуткові транзактори з євронакопиченням; 5. клієнти без євронакопичення; 6. прибуткові транзактори без євронакопичення. Для кожного з виявлених сегментів було розроблено цінову стратегію стосовно вартості обслуговування рахунків в євро. На підставі опрацьованих даних визначено, що реалізована сегментація та рекомендації дозволять оптимізувати операційні витрати без встановлення комісії для всіх корпоративних клієнтів банку. У першій хвилі необхідно вплинути лише на 9% клієнтів (встановити індивідуальну плату за обслуговування).

Розроблено комплекс регресійних моделей виживання, що дозволило оцінити фактори, що впливають на тривалість активностей роздрібних клієнтів банку, першим продуктом яких був кредит на авто. На першому етапі було реалізовано окремі моделі виживання Вейбулла для кожної соціально-демографічної характеристики клієнта, щоб визначити вплив кожного із факторів на тривалість активності, та загальну модель, з урахуванням усіх факторів. Для покращення якості побудованих моделей, було застосовано функцію stepAIC для пошуку найкращої комбінації, як характеристик клієнта, так і рівнів цих характеристик. Найсильнішим ризиком передчасного закриття відзначилися характеристики: вік більше 55 років, західний регіон проживання, дохід вище 150 тис. грн в місяць. На другому етапі було побудовано модель пропорційних небезпек Кокса для визначення впливу продажів кредитних карток на виживання клієнтів. Аналіз кривих виживання показав, що клієнти з кредитними картками статистично мають більш тривалий період активності, але перевірка припущення про пропорційність небезпек для цієї змінної була відхилена. З огляду на це було реалізовано модель виживання для фактору

«активована кредитна картка». Аналіз кривих виживання за новою моделлю показав значний вплив активної кредитної картки на показники виживання клієнтів. Припущення про пропорційні небезпеки було підтверджено для цієї змінної. На основі запропонованого підходу щодо моделювання поведінки клієнтів визначено, що продаж кредитної картки дозволяє пролонгувати активність клієнтів банку, першим продуктом яких був автокредит. Однак критичним фактором успіху є продаж кредитної картки та її активація.

Основні результати Розділу 3 опубліковано у наукових працях автора [172, 76, 116].

ВИСНОВКИ

У дисертаційній роботі розроблено та реалізовано новий концептуальний підхід до вирішення наукової задачі моделювання поведінки клієнтів в системі маркетингового менеджменту. Цей підхід дозволяє перетворювати наявні первинні дані про клієнтів на знання про закономірності їх поведінки, що надає практичні переваги при розробці маркетингових стратегій.

Виконане наукове дослідження дозволило зробити наступні висновки:

1. Розглянуто сутність та фундаментальні принципи реалізації маркетингового менеджменту, а також виокремлено його основні сучасні тенденції, зокрема: впровадження клієнтоорієнтованого підходу в організаціях; увага до персоналізованих маркетингових стратегій; використання аналізу на основі даних для отримання уявлення про потреби та поведінку клієнтів; підвищення ролі ефективного управління витратами в маркетинговому менеджменті. Обґрунтовано, що усі дії, що їх здійснює комерційна організація, підпадають під маркетинг і відіграють вирішальну роль у визначенні загального клієнтського досвіду.

2. Проаналізовано перехід від фокусу на товарах до фокусу на клієнтах в маркетинговому менеджменті. Обґрунтовано зниження ефективності масового маркетингу та важливість процесів управління гетерогенністю клієнтів в сучасному бізнес-середовищі. Досліджено, що на процес формування лояльності клієнта можна впливати шляхом розробки поведінкової сегментації клієнтів, вивчення особливостей поведінки клієнтів в кожному із сегментів та реалізації стратегій перехресних і повторних продажів для кожної із наявних груп. Доведено, що організації можуть досягти справжнього успіху і встановити міцні зв'язки, якщо надаватимуть пріоритет розумінню поведінки клієнтів і персоналізації їхнього досвіду.

3. Проаналізовано розвиток аналітики в маркетингу, зумовлений стрімким зростанням обсягів даних та обчислювальних потужностей. Розглянуто останні тренди в галузі інтелектуального маркетингу, які включають великі дані,

діджиталізацію, життєву цінність клієнта, користувацький досвід, CRM та персоналізацію. Обґрунтовано важливість якості даних для прогнозування поведінки клієнтів, як у контексті з достатнім обсягом даних, так і в ситуаціях з обмеженою якістю даних. Додатково проаналізовано бар'єри, що перешкоджають потенціалу впровадження процесів моделювання в маркетинговий менеджмент, включаючи суб'єктивні судження осіб, що приймають управлінські рішення, суперечливі короткострокові та довгострокові результати впровадження, а також недостатній рівень володіння методологіями моделювання. Зроблено огляд передових технологій економіко-математичного моделювання, таких як: кластеризація, класифікація, аналіз виживання, з оцінкою їхньої ефективності та потенціалу в маркетинговому менеджменті, обґрунтовано їхні переваги та недоліки.

4. Розроблено концептуальні положення щодо моделювання поведінки клієнтів як в умовах обмеженості даних, так і в умовах якісних даних. За умови обмеженості даних розроблено моделі поведінки бізнес-клієнтів консалтингової агенції та моделі поведінки клієнтів інтернет-магазину. Сформовано бачення необхідності моделювання їх поведінки, з точки зору створення конкурентних переваг для підприємств, що зароджуються. З точки зору якісних даних, запропоновано моделювати поведінку клієнтів банку. Обґрунтовано необхідність дослідження двох найважливіших сегментів банківських клієнтів, а саме: роздрібних клієнтів, які взяли кредит на придбання автомобіля (з точки зору найнижчого рівня утримання), та корпоративних бізнес-клієнтів (з точки зору найбільшої комерційної цінності).

5. Охарактеризовано типологію даних, а саме поділ на відкриті та комерційні дані. Наведено переваги та недоліки кожного типу. Обґрунтовано застосування машинного навчання як найкращого підходу до роботи в ситуаціях з обмеженими даними. Розроблено підхід до економіко-математичного моделювання поведінки клієнтів в умовах обмеженості даних, який складається з 9 етапів, детально описано кожен з етапів.

6. Реалізовано очищення та підготовку даних для моделювання методами машинного навчання, використано підходи для підвищення ефективності моделювання в умовах обмежених даних. Було використано відкриті дані про активність українських підприємств з отримання додаткового фінансування у вигляді інвестицій. Процес налаштування даних знизив кількість спостережень до мінімальної – фінальна вибірка включала 123 стартапи. Визначено ключові фактори, що впливають на різні прояви фінансового успіху стартапу. Застосовано 3 методи машинного навчання: дерево рішень CART, логістична регресія та випадковий ліс. Було реалізовано процес налаштування гіперпараметрів моделі, проведено множинні перехресні перевірки за п'ятьма різними метриками оцінювання якості моделей. Підтверджено ефективність запропонованого методу, шляхом порівняння отриманих прогнозних результатів з пороговими значеннями використовуваних метрик.

7. Спроектовано та реалізовано систему моделей, що дозволила передбачити час до наступного замовлення клієнта онлайн-магазину. Аналіз різних підходів дозволив запропонувати вирішення цієї задачі у два етапи. Спочатку було проведено сегментацію клієнтів інтернет-магазину методом k-середніх за показниками купівельної активності RFM. Було виявлено 6 груп клієнтів: 1. втрачені клієнти – клієнти, які не купують протягом довгого часу, здійснили менше 2 покупок за 18 місяців, остання з яких була більше року тому; 2. новий оптовий покупець - високий середній чек та активність, але пройшло небагато часу з першої покупки; 3. клієнти, яких ми скоро втратимо - проявили середні показники активності, але пройшло багато часу з останньої покупки; 4. активний роздрібний покупець - висока активність, висока частота та низька давність покупок, високий середній чек; 5. нові роздрібні покупці - висока активність, але протягом короткого періоду, достатній рівень вартості замовлень; 6. активний оптовий покупець - високий середній чек і активність, купують протягом тривалого періоду. На другому етапі було реалізовано 5 алгоритмів класифікації: лінійний дискримінантний аналіз, дерево рішень, випадковий ліс, метод опорних векторів, k - найближчих сусідів. Найкращу

прогнозу силу показав алгоритм випадковий ліс, який дозволив передбачити до 64% майбутніх активностей клієнтів.

8. Визначено особливості інфраструктури даних банків, окреслено експоненціальне зростання обсягів даних у банківському секторі та інтенсивність конкуренції, які вимагають економіко-математичного моделювання мінливої поведінки клієнтів. Розглянуто особливості поведінки банківських клієнтів, з фокусом на корпоративних клієнтів з рахунками в євро та роздрібних клієнтів з автокредитом. Запропоновано комплексну методику моделювання поведінки банківських клієнтів, яка враховує універсальний характер банків, що обслуговують широке коло індивідуальних та корпоративних клієнтів на висококонкурентному ринку. Розкрито алгоритм аналізу виживання для моделювання тривалості активності клієнтів банку.

9. Розроблено маркетингову сегментацію корпоративних клієнтів банку на основі алгоритму кластеризації k-середніх. Оптимальна кількість кластерів визначалася шляхом зважування результатів розрахунку 26 індексів з пакету NbClust та бізнес-вимог банку. Під час розрахунку алгоритму було виявлено шість груп схожості: 1. недохідні транзактори з накопиченням євро; 2. збиткові накопичувачі євро; 3. прибуткові накопичувачі євро; 4. прибуткові транзактори з євронакопиченням; 5. клієнти без євронакопичення; 6. прибуткові транзактори без євронакопичення. Для кожного з виявлених сегментів було розроблено цінову стратегію, стосовно вартості обслуговування рахунків в євро. Сегментація та рекомендації дозволили оптимізувати операційні витрати без встановлення комісії для всіх корпоративних клієнтів банку. У першій хвилі було запропоновано опрацювати лише 9% клієнтів (встановити індивідуальну плату за обслуговування).

10. Розроблено комплекс регресійних моделей виживання, що дозволило оцінити фактори, що впливають на тривалість активності роздрібних клієнтів банку, першим продуктом яких був кредит на авто. На першому етапі було реалізовано окремі моделі виживання Вейбулла для кожної соціально-демографічної характеристики клієнта, щоб врахувати вплив кожного окремого

фактору на тривалість активності, та загальну модель з урахуванням усіх факторів. Для покращення якості побудованих моделей було застосовано функцію stepAIC для пошуку найкращої комбінації, як характеристик клієнта, так і рівнів цих характеристик. Найсильнішим ризиком передчасного закриття відзначилися характеристики: вік більше 55 років, західний регіон проживання, дохід вище 150 тис. грн в місяць. На другому етапі було побудовано модель пропорційних ризиків Кокса для визначення впливу продажів кредитних карток на виживання клієнтів. Аналіз кривих виживання показав, що клієнти з кредитними картками статистично мають більш тривалий період активності, але перевірка припущення про пропорційність небезпек для цієї змінної була відхилена. З огляду на це було реалізовано модель виживання для фактору «активована кредитна картка». Аналіз кривих виживання за новою моделлю показав значний вплив активної кредитної картки на показники виживання клієнтів. Припущення про пропорційні небезпеки було підтверджено для цієї змінної. Результати моделювання підтвердили гіпотезу про те, що продаж кредитної картки дозволяє пролонгувати активність клієнтів банку, першим продуктом яких був автокредит. Однак критичним фактором успіху є продаж кредитної картки та її активація.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ang, Y. Q., Chia, A., & Saghafian, S. (2020). Using machine learning to demystify startups funding, Post-Money valuation, and success. In: Babich V., Birge J.R., Hilary G. (eds). *Innovative Technology at the Interface of Finance and Operations*. Springer Series in Supply Chain Management. *Springer, Cham*. <https://doi.org/10.2139/ssrn.3681682>.
2. Armstrong, S., Esber, D., Heller, J., & Timelin, B. (2020). *Modern Marketing: What it is, what it isn't, and how to do it*. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/modern-marketing-what-it-is-what-it-isnt-and-how-to-do-it>.
3. Ascarza, E., Fader, P.S., & Hardie, B.G.S. (2017). Marketing Models for the Customer-Centric Firm. In: Wierenga, B., van der Lans, R. (eds) *Handbook of Marketing Decision Models*. *International Series in Operations Research & Management Science*. Springer, Cham, 254, 297–329. https://doi.org/10.1007/978-3-319-56941-3_10.
4. Aufreiter, N. A., Elzinga, D., & Gordon, J. W. (2003). Better branding. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/better-branding>.
5. Baardman, L., Boroujeni, S. B., Cohen-Hillel, T., Panchamgam, K., & Perakis, G. (2023). Detecting customer trends for optimal promotion targeting. *Manufacturing & Service Operations Management*, 25(2), 448–467. <https://doi.org/10.1287/msom.2020.0893>.
6. Bai, M., Zheng, Y., & Shen, Y. (2021). Gradient boosting survival tree with applications in credit scoring. *Journal of the Operational Research Society*, 73(1), 39–55. <https://doi.org/10.1080/01605682.2021.1919035>.
7. Baker, M. J. (1976). *Marketing*. Red Globe Press London. <https://doi.org/10.1007/978-1-349-15703-7>.

8. Baker, M. J. (2005). Marketing is marketing—everywhere! *Vikalpa: The Journal for Decision Makers*, 30(3), 1–10. <https://doi.org/10.1177/0256090920050301>.
9. Baron, S., & Harris, K. (1995). Services marketing: Text and cases. *Red Globe Press London*. <https://doi.org/10.1007/978-1-349-24174-3>.
10. Bartels, R. (1976). The History of Marketing Thought, 2nd edn. Columbus: Grid.
11. Bernstein, J. S. (2014). Standing room only: Marketing insights for engaging performing arts audiences. *Palgrave Macmillan*. <https://doi.org/10.1007/978-1-137-37569-8>.
12. Bibby, C., Gordon, J., Schuler, G., & Stein, E. (2021). The big reset: Data-driven marketing in the next normal. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-big-reset-data-driven-marketing-in-the-next-normal>.
13. Bickhoff, N., Hollensen, S., & Opresnik, M. O. (2014). Quintessence of Marketing: What you really need to know to manage your marketing activities. *Springer*. <https://doi.org/10.1007/978-3-642-45444-8>.
14. Blattberg, R. C., Kim, P., & Neslin, S. A. (2008). Database marketing analyzing and managing customers. *Springer Science + Business Media*. <https://doi.org/10.1007/978-0-387-72579-6>.
15. Bonacchi, M., & Perego, P. (2019). Customer accounting creating value with customer analytics. *Springer Cham*. <https://doi.org/10.1007/978-3-030-01971-6>.
16. Brei, V. A. (2020). Machine Learning in Marketing: Overview, learning strategies, applications, and future developments. *Foundations and Trends in Marketing*, 14(3), 173–236. <https://doi.org/10.1561/17000000065>.
17. Brinker, S., & Heller, J. (2015). Marketing technology: What it is and how it should work. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/marketing-technology-what-it-is-and-how-it-should-work>.

18. Brown, A. (2006). What is marketing? Bogers, R., Craker, L., & Lange, D. (Eds.), *Medicinal and Aromatic Plants. Wageningen UR Frontis series. Springer.*, 17, 213–221. Retrieved from <https://edepot.wur.nl/137157>.
19. Brown, B., Kanagasabai, K., Pant, P., & Pinto, G. S. (2017). Capturing value from your Customer Data. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/quantumblack/our-insights/capturing-value-from-your-customer-data>.
20. Brunswick, G. J. (2014). A chronology of the definition of marketing. *Journal of Business & Economics Research (JBER)*, 12(2), 105. <https://doi.org/10.19030/jber.v12i2.8523>.
21. Butt, C., Hakim, H., Schaffner, R., & Tas, R. (2020). The forgotten hero of marketing excellence: Spend management. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-forgotten-hero-of-marketing-excellence-spend-management>.
22. Cavallone, M. (2017). Marketing and customer loyalty: The Extra Step Approach. *Springer International Publishing*. <https://doi.org/10.1007/978-3-319-51991-3>.
23. Chagovets, L., Chahovets, V., Chernova, N. (2020). Machine Learning Methods Applications for Estimating Unevenness Level of Regional Development. In: Ageyev, D., Radivilova, T., Kryvinska, N. (eds) *Data-Centric Business and Applications. Lecture Notes on Data Engineering and Communications Technologies*, Springer, Cham. 42, 115 – 139. https://doi.org/10.1007/978-3-030-35649-1_6.
24. Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6), 1–36. <https://doi.org/10.18637/jss.v061.i06>.
25. Chatziefstathiou, D. (2007). The history of marketing an idea: The example of baron Pierre de Coubertin as a social marketer. *European Sport Management Quarterly*, 7(1), 55–80. <https://doi.org/10.1080/16184740701270337>.

26. Chawla, D., & Joshi, H. (2021). Segmenting mobile banking users based on the usage of mobile banking services. *Global Business Review*, 22(3), 689–704. <https://doi.org/10.1177/0972150918811257>.
27. Chen, D. (2019). Online retail II. *UCI Machine Learning Repository*. <https://doi.org/10.24432/C5CG6D>.
28. Chen, L. M., Su, Z., & Jiang, B. (2015). Mathematical problems in Data Science: Theoretical and Practical Methods. *Springer*. <https://doi.org/10.1007/978-3-319-25127-1>.
29. Chen, N. (2022). Research on E-commerce database marketing based on machine learning algorithm. *Computational Intelligence and Neuroscience*, 2022, 1–13. <https://doi.org/10.1155/2022/7973446>.
30. Chornous, G., Fareniuk, Y., Rolandas Giedraitis, V., Ulvidienė, E., & Kharlamova, G. (2023). A data science-based marketing decision support system for Brand Management. *Innovative Marketing*, 19(2), 38–50. [https://doi.org/10.21511/im.19\(2\).2023.04](https://doi.org/10.21511/im.19(2).2023.04).
31. Conway, D. (2013). The data science Venn diagram. Drew Conway. Retrieved from <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.
32. Cook, D. (1981). The evolution of the concept — marketing. *Industrial Management amp; Data Systems*, 81(7/8), 17–22. <https://doi.org/10.1108/eb057200>.
33. Court, D. C., Gordon, J. W., & Perrey, J. (2005). Boosting returns on marketing investment. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/boosting-returns-on-marketing-investment>.
34. Court, D., Perrey, J., McGuire, T., Gordon, J., & Spillecke, D. (2013). Big Data, analytics, and the future of Marketing & Sales. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/ebook-big-data-analytics-and-the-future-of-marketing--sales>.
35. Cundari, A. (2015). Customer-centric marketing: Build relationships, create advocates and influence your customers. *Wiley*, 176.

36. Curry, B., & Moutinho, L. (1994). Intelligent computer models for marketing decisions. *Management Decision*, 32(4), 30–35. <https://doi.org/10.1108/00251749410058653>.
37. Dam, N. A., Dinh, T. L., & Menvielle, W. (2022). The quest for customer intelligence to support marketing decisions: A knowledge-based framework. *Vietnam Journal of Computer Science*, 09(03), 349–368. <https://doi.org/10.1142/s2196888822500208>.
38. De Mauro, A., Sestino, A., & Bacconi, A. (2022). Machine learning and artificial intelligence use in marketing: A general taxonomy. *Italian Journal of Marketing*, 2022(4), 439–457. <https://doi.org/10.1007/s43039-022-00057-w>.
39. Dealroom Ukraine. *Official website of the Dealroom Ukraine*. Retrieved from <https://ukraine.dealroom.co/>.
40. Deighton, J. A., Mela, C. F., & Moorman, C. (2020). Marketing Thinking and Doing. *Journal of Marketing*, 85(1), 1-6. <https://doi.org/10.1177/0022242920977093>.
41. Derbentsev, V., Babenko, V., Khrustalev, K., Obruch, H., & Khrustalova, S. (2021). Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices. *International Journal of Engineering*, 34(1), 140–148. <https://doi.org/10.5829/ije.2021.34.01a.16>.
42. Derbentsev, V., Matviychuk, A., Soloviev, V.N. (2020). Forecasting of Cryptocurrency Prices Using Machine Learning. In: Pichl, L., Eom, C., Scalas, E., Kaizoji, T. (eds) *Advanced Studies of Financial Technologies and Cryptocurrency Markets*. Springer, Singapore. https://doi.org/10.1007/978-981-15-4498-9_12.
43. Derbentsev, V., Velykoivanenko, H., & Datsenko, N. (2019). Machine Learning Approach for forecasting cryptocurrencies time series. *Neuro-Fuzzy Modeling Techniques in Economics*, (8), 65–93. <https://doi.org/10.33111/nfmte.2019.065>.
44. Duhring, L. (2017). Reassessing the relationship between marketing and Public Relations: New Perspectives from the Philosophy of Science and history of thought. *Springer VS Wiesbaden*. <https://doi.org/10.1007/978-3-658-18034-8>.

45. D'Avino, M., De Simone, V., Iannucci, M., & Schiraldi, M. M. (2015). Guidelines for e-startup promotion strategy. *Journal of Technology Management & Innovation*, 10(1), 1–16. <https://doi.org/10.4067/s0718-27242015000100001>.
46. El-Ansary, A. I. (2005). Relationship Marketing Management. *Journal of Relationship Marketing*, 4(1–2), 43–56. https://doi.org/10.1300/j366v04n01_04.
47. Elandt-Johnson, R. C., & Johnson, N. L. (1999). Survival models and data analysis. *John Wiley & Sons*. <https://doi.org/10.1002/9781119011040>.
48. Fahrmeir, L., Tutz, G. (1994). Survival models. In: Multivariate Statistical Modelling Based on Generalized Linear Models. *Springer Series in Statistics*. Springer, New York, 305–344. https://doi.org/10.1007/978-1-4899-0010-4_9.
49. Foster, D. (1982). Mastering marketing. *Macmillan Master Series (Business)*. <https://doi.org/10.1007/978-1-349-86072-2>.
50. Foxall, G. (1984). Marketing's domain. *European Journal of Marketing*, 18(1), 25–40. <https://doi.org/10.1108/eum0000000004721>.
51. French, T., LaBerge, L., & Magill, P. (2011). We're all marketers now. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/were-all-marketers-now>.
52. Gaddam, S. R., Phoha, V. V., & Balagani, K. S. (2007). K-means+id3: A novel method for supervised anomaly detection by cascading K-means clustering and ID3 decision tree learning methods. *IEEE Transactions on Knowledge and Data Engineering*, 19(3), 345–354. <https://doi.org/10.1109/tkde.2007.44>.
53. Galchynska, J., Larina, Y., Varchenko, O., Struk, N., & Gryshchenko, O. (2021). Perspectives of Ukrainian Bioenergy Development: Estimation by means of cluster analysis and marketing approach. *Economic Annals-XXI*, 187(1–2), 63–74. <https://doi.org/10.21003/ea.v187-06>.
54. Galil, K., Hauptman, A., & Rosenboim, R. L. (2023). Prediction of corporate credit ratings with machine learning: Simple interpretative models. *Finance Research Letters*, 58, 104648. <https://doi.org/10.1016/j.frl.2023.104648>.
55. Giudici, P., & Figini, S. (2009). Applied Data Mining for business and industry. Wiley. <https://doi.org/10.1002/9780470745830>.

56. Goldsmith, R. E. (2004). Current and future trends in marketing and their implications for the discipline. *Journal of Marketing Theory and Practice*, 12(4), 10–17. <https://doi.org/10.1080/10696679.2004.11658527>.
57. Gomes, M. A., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and E-Business Management*, 21(3), 527–570. <https://doi.org/10.1007/s10257-023-00640-4>.
58. Gopagoni, D. R., Lakshmi, P. V., & Chaudhary, A. (2020). Evaluating machine learning algorithms for Marketing Data Analysis: Predicting Grocery Store Sales. *Communication Software and Networks*, 155–163. https://doi.org/10.1007/978-981-15-5397-4_17.
59. Gordon, J., Gruntges, V., Smith, V., & Staack, Y. (2016). New insights for new growth: What it takes to understand your customers today. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/new-insights-for-new-growth-what-it-takes-to-understand-your-customers-today>.
60. Grillo, H., Alemany, M. M. E., Ortiz, A., & Fuertes-Miquel, V. S. (2017). Mathematical modelling of the order-promising process for fruit supply chains considering the perishability and subtypes of products. *Applied Mathematical Modelling*, 49, 255–278. <https://doi.org/10.1016/j.apm.2017.04.037>.
61. Guo, Z. (2016). Intelligent decision-making models for production and retail operations. *Springer Berlin Heidelberg*. <https://doi.org/10.1007/978-3-662-52681-1>.
62. Guryanova, L., Yatsenko, R., Dubrovina, N., & Babenko, V. (2020). Machine learning methods and models, predictive analytics and applications. *CEUR Workshop Proceedings*, 2649, 1–5. Retrieved from <https://ceur-ws.org/Vol-2649/>.
63. Gust, B. (2020). User-oriented appropriateness a theoretical model of written text on Facebook for improved PR communication. *Springer Gabler Wiesbaden*. <https://doi.org/10.1007/978-3-658-29999-6>.

64. Gvozdytskyi, V. (2023). Transforming of traditional commerce into e-commerce: Trends in the world and in Ukraine. *Development Management*, 21(1). <https://doi.org/10.57111/devt/1.2023.36>
65. Harrell, F. E. (2015). Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis. *Springer Cham*. <https://doi.org/10.1007/978-3-319-19425-7>.
66. Homburg, C., Schafer, H., Schneider, J. (2012). Customers: The Focus of the Sales Strategy. In: Sales Excellence. Management for Professionals. *Springer, Berlin, Heidelberg*, 27–39. https://doi.org/10.1007/978-3-642-29169-2_2.
67. Huang, Z. (2022). Intelligent decision-making model of enterprise management based on Random Forest algorithm. *Mobile Information Systems*, 2022, 1–11. <https://doi.org/10.1155/2022/6184106>.
68. Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Machine learning towards Intelligent Systems: Applications, challenges, and opportunities. *Artificial Intelligence Review*, 54(5), 3299–3348. <https://doi.org/10.1007/s10462-020-09948-w>.
69. Izenman, A.J. (2013). Linear Discriminant Analysis. In: Modern Multivariate Statistical Techniques. *Springer Texts in Statistics*. Springer, New York, NY. 237–280. https://doi.org/10.1007/978-0-387-78189-1_8.
70. Kaminskyi, A., Nehrey, M., Babenko, V., & Zimon, G. (2022). Model of optimizing correspondence risk-return marketing for short-term lending. *Journal of Risk and Financial Management*, 15(12), 583. <https://doi.org/10.3390/jrfm15120583>.
71. Kang, M., & Gong, T. (2019). Dysfunctional customer behavior: Conceptualization and empirical validation. *Service Business*, 13(4), 625–646. <https://doi.org/10.1007/s11628-019-00398-1>.
72. Khoa Dam, N. A. (2019). Marketing intelligence from Data Mining Perspective — a literature review. *International Journal of Innovation, Management and Technology*, 10(5), 184–190. <https://doi.org/10.18178/ijimt.2019.10.5.859>.

73. Kim, J., Kim, M., Choi, J., & Trivedi, M. (2019). Offline social interactions and online shopping demand: Does the degree of Social Interactions matter? *Journal of Business Research*, 99, 373–381. <https://doi.org/10.1016/j.jbusres.2017.09.022>.
74. Kim, W. C., & Mauborgne, R. (2004). Blue Ocean Strategy. *Harvard Business Review*. Retrieved from <https://hbr.org/2004/10/blue-ocean-strategy>.
75. Kleinbaum, D. G., & Klein, M. (2020). Survival analysis: A self-learning text. *Springer*. <https://doi.org/10.1007/978-1-4419-6646-9>.
76. Klochko, R., & Piskunova, O. (2023). Marketing segmentation of Banks' corporate clients based on data mining technique. *Statistics in Transition New Series*, 24(5), 147–166. <https://doi.org/10.59170/stattrans-2023-068>.
77. Kofanov, O., & Zozul'ov, O. (2018). Successful development of startups as a global trend of innovative socio-economic transformations. *International and Multidisciplinary Journal of Social Sciences*, 7(2), 191. <https://doi.org/10.17583/rimcis.2018.3576>.
78. Kohli, S. (2023). The World of “Ands”: Consumers set the tone. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/the-world-of-ands-consumers-set-the-tone>.
79. Kotler, P., & Levy, S. J. (1969). Broadening the concept of marketing. *Journal of Marketing*, 33(1), 10–15. <https://doi.org/10.2307/1248740>.
80. Kotler, P., Pfoertsch, W., & Sponholz, U. (2020). H2H marketing: The genesis of human-to-human marketing. *Springer Cham*. <https://doi.org/10.1007/978-3-030-59531-9>.
81. Kumar, V. (2018). A theory of customer valuation: Concepts, metrics, strategy, and implementation. *Journal of Marketing*, 82(1), 1–19. <https://doi.org/10.1509/jm.17.0208>.
82. Kumar, V., & Reinartz, W. (2018). Customer relationship management concept, strategy, and Tools. *Springer Berlin Heidelberg*. <https://doi.org/10.1007/978-3-662-55381-7>.

83. Kundisch, D. (2003). New strategies for financial services firms: The life-cycle-solution approach. *Physica Heidelberg*. <https://doi.org/10.1007/978-3-642-57412-2>.
84. Kucuk, S.U. (2020). Consumer Voice. In: Consumer Voice. *Palgrave Macmillan, Cham*, 21–42. https://doi.org/10.1007/978-3-030-53983-2_2
85. Lancaster, G., & Reynolds, P. (1998). Marketing. *Macmillan Publishers Limited*. <https://doi.org/10.1007/978-1-349-14039-8>.
86. Landerman, L. R., Land, K. C., & Pieper, C. F. (1997). An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods & Research*, 26(1), 3–33. <https://doi.org/10.1177/0049124197026001001>.
87. Layton, R. (2015). Marketing, marketing systems, and the framing of marketing history. *Journal of Historical Research in Marketing*, 7(4), 549–572. <https://doi.org/10.1108/jhrm-02-2015-0008>.
88. Lecinski, J. (2011). ZMOT - think with google. Retrieved from https://www.thinkwithgoogle.com/_qs/documents/673/2011-winning-zmot-ebook_research-studies.pdf.
89. Lee, C.-C., Tseng, H.-C., Liu, C.-C., & Chou, H.-J. (2022). The analysis of intelligent marketing platform in high-tech products by data mining algorithm. *Wseas transactions on business and economics*, 19, 555–561. <https://doi.org/10.37394/23207.2022.19.50>.
90. Lee, W.-M. (2019). Python machine learning. *John Wiley & Sons*. <https://doi.org/10.1002/9781119557500>.
91. Liashenko, O., Kravets, T., & Prokopenko, M. (2021). Consumer behavior clustering of food retail chains by machine learning algorithms. *Access Journal - Access to Science, Business, Innovation in the Digital Economy*, 2(3), 234–251. [https://doi.org/10.46656/access.2021.2.3\(3\)](https://doi.org/10.46656/access.2021.2.3(3)).
92. Liu, X. (2012). Survival analysis: Models and applications. *John Wiley & Sons*. <https://doi.org/10.1002/9781118307656>.

93. Livingstone, J. M. (1976). *International Marketing Management*. Palgrave Macmillan London. <https://doi.org/10.1007/978-1-349-15653-5>.
94. Lv, Z., Wang, N., Ma, X., Sun, Y., Meng, Y., & Tian, Y. (2022). Evaluation standards of intelligent technology based on Financial Alternative Data. *Journal of Innovation & Knowledge*, 7(4), 100–229. <https://doi.org/10.1016/j.jik.2022.100229>.
95. Majid, K. A. (2020). Effect of interactive marketing channels on Service Customer Acquisition. *Journal of Services Marketing*, 35(3), 299–311. <https://doi.org/10.1108/jsm-08-2019-0282>.
96. Martinez, L. M., Neves, T. V., & Martinez, L. F. (2020). The effect of Web Advertising Visual Design on online purchase intention: Insights on generations Y and Z. *Advances in Digital Marketing and eCommerce*, 130–140. https://doi.org/10.1007/978-3-030-47595-6_17.
97. Meldrum, M., & McDonald, M. (1995). *Key Marketing Concepts*. Red Globe Press London. <https://doi.org/10.1007/978-1-349-13877-7>.
98. Mendonça, P., & Teixeira, S. (2022). *Digital Marketing Trends*. Porto. CEOS Edições. <https://doi.org/10.56002/ceos.0002b>.
99. Merritt, R. K., Kamin, T., Hussenöder, F., & Huibregtsen, J. (2017). The history of social marketing in Europe. *Social Marketing Quarterly*, 23(4), 291–301. <https://doi.org/10.1177/1524500417732771>.
100. Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the adoption of machine learning-based analytical tools in digital marketing. *IEEE Access*, 7, 85705–85718. <https://doi.org/10.1109/access.2019.2924425>.
101. Min, S., Zhang, X., Kim, N., & Srivastava, R. K. (2016). Customer acquisition and retention spending: An analytical model and empirical investigation in wireless telecommunications markets. *Journal of Marketing Research*, 53(5), 728–744. <https://doi.org/10.1509/jmr.14.0170>.
102. Moore, D. F. (2016). *Applied Survival Analysis Using R*. Springer Cham. <https://doi.org/10.1007/978-3-319-31245-3>.
103. Moscato, P., & de Vries, N. J. (2019). Business and consumer analytics new ideas. *Springer Cham*. <https://doi.org/10.1007/978-3-030-06222-4>.

104. Nai, W., Liu, L., Wang, S., & Dong, D. (2018). Modeling the trend of credit card usage behavior for different age groups based on singular spectrum analysis. *Algorithms*, 11(2), 1–11. <https://doi.org/10.3390/a11020015>.
105. Nauck, D., Ruta, D., Spott, M., & Azvine, B. (2008). Predictive Customer Analytics and Real-Time Business Intelligence. In: Voudouris, C., Lesaint, D., Owusu, G. (eds) *Service Chain Management*. Springer, Berlin, Heidelberg, 205–214. https://doi.org/10.1007/978-3-540-75504-3_14.
106. Nethravathi, P. S., Bai, G. V., Spulbar, C., Suhan, M., Birau, R., Calugaru, T., Hawaldar, I. T., & Ejaz, A. (2020). Business intelligence appraisal based on customer behaviour profile by using hobby based opinion mining in India: A case study. *Economic Research-Ekonomska Istrazivanja*, 33(1), 1889–1908. <https://doi.org/10.1080/1331677x.2020.1763822>.
107. Oberg, C. (2014). Customer relationship challenges following international acquisitions. *International Marketing Review*, 31(3), 259–282. <https://doi.org/10.1108/imr-10-2012-0166>.
108. Osborne, P., & Ballantyne, D. (2012). The paradigmatic pitfalls of customer-centric marketing. *Marketing Theory*, 12(2), 155–172. <https://doi.org/10.1177/1470593112441564>.
109. Oumayma, B. (2019). Social Media Made Me Buy it. *Proceedings of the 4th International Conference on Smart City Applications*, 1–7. <https://doi.org/10.1145/3368756.3369016>.
110. Ozawa, T., Sekiguchi, A., & Tsuda, K. (2017). A method for the construction of customer behavioral modeling knowledge for B2B event marketing strategy. *Procedia Computer Science*, 112, 1353–1360. <https://doi.org/10.1016/j.procs.2017.08.027>.
111. Pacheco, J., Chela, J., & Salome, G. (2023). Fraud detection with machine learning: Model comparison. *International Journal of Business Intelligence and Data Mining*, 22(4), 434–450. <https://doi.org/10.1504/ijbidm.2023.130587>.

112. Palmatier, R. W., & Crecelius, A. T. (2019). The “first principles” of marketing strategy. *AMS Review*, 9(1–2), 5–26. <https://doi.org/10.1007/s13162-019-00134-y>.
113. Parniangtong, S. (2017). Competitive advantage of customer centricity. *Springer Singapore*. <https://doi.org/10.1007/978-981-10-4442-7>.
114. Parowicz, I. (2019). Cultural Heritage Marketing: A relationship marketing approach to conservation services. *Palgrave Pivot Cham*. <https://doi.org/10.1007/978-3-030-00287-9>.
115. Pavlenko, V., Ponomarenko, I., Morhulets, O., Fedorchenko, A., Chorna, O., & Pylypenko, V. (2023). Creating educational products with using data science and Digital Marketing. *2023 IEEE 4th KhPI Week on Advanced Technology (KhPIWeek)*. <https://doi.org/10.1109/khpiweek61412.2023.10312906>.
116. Piskunova, O., Klochko, R., Bilyk, T., & Frolova, T. (2023). Modelling the bank customer activity duration based on the Cox Econometric Survival Model. *Financial and Credit Activity Problems of Theory and Practice*, 6(53), 21–31. <https://doi.org/10.55643/fcaptp.6.53.2023.4244>.
117. Piskunova, O., & Klochko, R. (2020). Classification of e-commerce customers based on Data Science techniques. *CEUR Workshop Proceedings*, 2649, 6 – 20. Retrieved from <https://ceur-ws.org/Vol-2649/>.
118. Piskunova, O., Ligonenko, L., Klochko, R., Frolova, T., & Bilyk, T. (2021). Applying machine learning approach to start-up success prediction. *Scientific Horizons*, 24(11), 72–84. [https://doi.org/10.48077/scihor.24\(11\).2021.72-84](https://doi.org/10.48077/scihor.24(11).2021.72-84).
119. Pyle, D., & Jose, C. S. (2015). An executive’s Guide to Machine Learning. *McKinsey & Company*. Retrieved from <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/an-executives-guide-to-machine-learning>.
120. Rajagopal. (2020). Transgenerational marketing: Evolution, expansion, and experience. *Palgrave Macmillan / Springer Nature Switzerland AG*. <https://doi.org/10.1007/978-3-030-33926-5>.

121. Ramasubramanian, K., & Singh, A. (2018). Machine learning using r: With time series and industry-based use cases in R. *Apress Berkeley, CA*. <https://doi.org/10.1007/978-1-4842-4215-5>.
122. Ratchford, B. T. (2020). The history of academic research in marketing and its implications for the future. *Spanish Journal of Marketing - ESIC*, 24(1), 3–36. <https://doi.org/10.1108/sjme-11-2019-0096>.
123. Ribeiro, R. B., & Soares, I. (2015). Marketing/Marketing Science. *The Wiley Blackwell Encyclopedia of Consumption and Consumer Studies*, 1–5. <https://doi.org/10.1002/9781118989463.wbeccs160>.
124. Rout, J. K., Rout, M., & Das, H. (2020). Machine Learning for Intelligent Decision Science. *Springer Singapore*. <https://doi.org/10.1007/978-981-15-3689-2>.
125. Ryan, A. M., & Ployhart, R. E. (2012). Customer service behavior. In *Handbook of Psychology, Second Edition*. *John Wiley & Son*. Retrieved from <https://doi.org/10.1002/9781118133880.hop212018>.
126. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and Research Directions. *SN Computer Science*, 2(3). <https://doi.org/10.1007/s42979-021-00592-x>.
127. Savych, O., & Shkoda, T. (2022). Cultural Relationship Marketing Prospects and potential. *Contributions to Management Science*, 269–278. https://doi.org/10.1007/978-3-031-14121-8_14.
128. Schlereth, C., Skiera, B., & Schulz, F. (2018). Why do consumers prefer static instead of dynamic pricing plans? an empirical study for a better understanding of the low preferences for time-variant pricing plans. *European Journal of Operational Research*, 269(3), 1165–1179. <https://doi.org/10.1016/j.ejor.2018.03.033>.
129. Schwarzkopf, S. (2015). Marketing history from below: Towards a paradigm shift in marketing historical research. *Journal of Historical Research in Marketing*, 7(3), 295–309. <https://doi.org/10.1108/jhrm-06-2015-0021>.
130. Shah, A. M., Zahoor, S. Z., & Qureshi, I. H. (2019). Social Media and purchasing behavior: A study of the mediating effect of customer relationships. *Journal of Global Marketing*, 32(2), 93–115. <https://doi.org/10.1080/08911762.2018.1497243>.

131. Shah, N., Engineer, S., Bhagat, N., Chauhan, H., & Shah, M. (2020). Research trends on the usage of machine learning and artificial intelligence in advertising. *Augmented Human Research*, 5(1), 1–15. <https://doi.org/10.1007/s41133-020-00038-8>.
132. Sharma, S., & Goyal, D. P. (2020). Entrepreneurial marketing strategies for small businesses : An exploratory study of start-up companies in India. *Indian Journal of Marketing*, 50(8–9), 48–65. <https://doi.org/10.17010/ijom/2020/v50/i8-9/154691>.
133. Shaw, E. H. (2009). Reflections on the history of marketing thought. *Journal of Historical Research in Marketing*, 1(2), 330–345. <https://doi.org/10.1108/17557500910974640>.
134. Shaw, E. H., & Jones, D. G. B. (2005). A history of schools of marketing thought. *Marketing Theory*, 5(3), 239–281. <https://doi.org/10.1177/1470593105054898>.
136. Shaw, E. H., & Tamlia, R. D. (2001). Robert Bartels and the History of Marketing Thought. *Journal of Macromarketing*, 21(2), 156–163. <https://doi.org/10.1177/0276146701212006>.
137. Sheth, J. (2021). New areas of research in marketing strategy, consumer behavior, and marketing analytics: The future is bright. *Journal of Marketing Theory and Practice*, 29(1), 3–12. <https://doi.org/10.1080/10696679.2020.1860679>.
138. Sheth, J. N., Sisodia, R. S., & Sharma, A. (2000). The antecedents and consequences of customer-centric marketing. *Journal of the Academy of Marketing Science*, 28(1), 55–66. <https://doi.org/10.1177/0092070300281006>.
139. Shin, D.-H., & Kim, W.-Y. (2008). Forecasting customer switching intention in mobile service: An exploratory study of predictive factors in mobile number portability. *Technological Forecasting and Social Change*, 75(6), 854–874. <https://doi.org/10.1016/j.techfore.2007.05.001>.
140. Sila, I. (2019). Antecedents of electronic commerce in developing economies. *Journal of Global Information Management*, 27(1), 66–92. <https://doi.org/10.4018/jgim.2019010104>.

141. Simmonds, K. (1986). Marketing as innovation the Eighth paradigm. *Journal of Management Studies*, 23(5), 479–500. <https://doi.org/10.1111/j.1467-6486.1986.tb00433.x>.
142. Skawińska, E., & Zalewski, R. I. (2020). Success factors of startups in the EU—A comparative study. *Sustainability*, 12(19), 8200. <https://doi.org/10.3390/su12198200>
143. Stauss, B., & Seidel, W. (2019). Effective complaint management: The Business Case for customer satisfaction. *Springer Cham*. <https://doi.org/10.1007/978-3-319-98705-7>.
144. Susilo, Y. O., Hanks, N., & Ullah, M. (2013). An exploration of shoppers travel mode choice in visiting convenience stores in the UK. *Transportation Planning and Technology*, 36(8), 669–684. <https://doi.org/10.1080/03081060.2013.851505>.
145. Swain, S. D., Berger, P. D., & Weinberg, B. D. (2014). The customer equity implications of using incentives in acquisition channels: A nonprofit application. *Journal of Marketing Analytics*, 2(1), 1–17. <https://doi.org/10.1057/jma.2014.1>.
146. Tamilia, R. D. (2009). An overview of the history of marketing thought. *Journal of Historical Research in Marketing*, 1(2), 346–360. <https://doi.org/10.1108/17557500910974659>.
147. Tongurai, J., & Vithessonthi, C. (2018). The impact of the banking sector on economic structure and growth. *International Review of Financial Analysis*, 56, 193–207. <https://doi.org/10.1016/j.irfa.2018.01.002>.
148. Usman, M. A., & Sun, X. (2023). The impact of digital platforms on new startup performance: Strategy as moderator. *Heliyon*, 9(12). <https://doi.org/10.1016/j.heliyon.2023.e22159>.
149. van Bruggen, G. H. (2010). Marketing decision making and decision support: Challenges and perspectives for Successful Marketing Management Support Systems. *Foundations and Trends in Marketing*, 4(4), 209–332. <https://doi.org/10.1561/17000000022>.
150. Velykoivanenko, H., & Korchynskyi, V. (2022). Application of clustering in the dimensionality reduction algorithms for separation of financial status of

commercial banks in Ukraine. *Universal Journal of Accounting and Finance*, 10(1), 148–160. <https://doi.org/10.13189/ujaf.2022.100116>.

151. Vieira, V. A., de Almeida, M. I., Agnihotri, R., da Silva, N. S., & Arunachalam, S. (2019). In pursuit of an effective B2B digital marketing strategy in an emerging market. *Journal of the Academy of Marketing Science*, 47(6), 1085–1108. <https://doi.org/10.1007/s11747-019-00687-1>.

152. Wagner, K. (2021). Drivers and barriers of consumer behavior regarding new technologies and digital channels: Investigating the phenomenon of anthropomorphism and new online consumption forms. *Springer Gabler Wiesbaden*. <https://doi.org/10.1007/978-3-658-35150-2>.

153. Wang, B., Wang, P., & Tu, Y. (2021). Customer satisfaction service match and service quality-based Blockchain Cloud Manufacturing. *International Journal of Production Economics*, 240, 108220. <https://doi.org/10.1016/j.ijpe.2021.108220>.

154. Wang, D. (2020). Research on bank marketing behavior based on machine learning. *Proceedings of the 2nd International Conference on Artificial Intelligence and Advanced Manufacture*, 150–154. <https://doi.org/10.1145/3421766.3421800>.

155. Waterworth, D. (1987). *Marketing for the Small Business*. Red Globe Press London. <https://doi.org/10.1007/978-1-349-18881-9>.

156. Webb, A.R. and Copsey, K.D. (2011). Linear Discriminant Analysis. In *Statistical Pattern Recognition*. John Wiley & Sons, 221–273. <https://doi.org/10.1002/9781119952954.ch5>.

157. Wierenga, B. (2011). Managerial decision making in marketing: The next research frontier. *International Journal of Research in Marketing*, 28(2), 89–101. <https://doi.org/10.1016/j.ijresmar.2011.03.001>.

158. Wierenga, B., & Gerrit, B. (2000). *Marketing Management Support Systems Principles, tools, and implentation*. Springer New York. <https://doi.org/10.1007/978-1-4615-4595-8>.

159. Wirtz, B.W. (2019). B2B Digital Business Models. In: *Digital Business Models*. Progress in IS. Springer, Cham, 161–173. https://doi.org/10.1007/978-3-030-13005-3_9.

160. Wright, L. T., Moutinho, L., Stone, M., & Bagozzi, R. P. (2021). *The routledge companion to marketing research*. Routledge.
161. Xu, X., Wang, X., Li, Y., & Haghighi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of Information Management*, 37(6), 673–683. <https://doi.org/10.1016/j.ijinfomgt.2017.06.004>.
162. Zhang, C., Wang, X., Cui, A. P., & Han, S. (2020). Linking big data analytical intelligence to customer relationship management performance. *Industrial Marketing Management*, 91, 483–494. <https://doi.org/10.1016/j.indmarman.2020.10.012>.
163. Zhang, J., & Zhu, J. (2014). Research intelligent precision marketing of e-commerce based on the big data. *Journal of Management and Strategy*, 5(1). <https://doi.org/10.5430/jms.v5n1p33>.
164. Zhao, H. (2020). Feature learning and understanding: Algorithms and applications. *Springer Cham*. <https://doi.org/10.1007/978-3-030-40794-0>.
165. Агютін, М. (2021). Моделювання процесу управління інструментами цифрового маркетингу. *Моделювання та інформаційні системи в економіці*, 101, 5–11. <https://doi.org/10.33111/mise.101.1>.
166. Архипова, Т., & Корюгін, В. (2023) Цифровізація управління маркетинговою діяльністю підприємства. *Інвестиції: практика та досвід*. 10. 65–69. <https://doi.org/10.32702/2306-6814.2023.10.65>.
167. Овсієнко, Н. (2021). Особливості маркетингової діяльності підприємств будівельної галузі. *Економіка та суспільство*, (26). <https://doi.org/10.32782/2524-0072/2021-26-26>.
168. Олексюк, О., & Шафалюк, О. (2022). Особливості використання відкритих даних у підготовці рішень маркетингового та комерційного менеджменту. *Стратегія економічного розвитку України*, 50, 45–55. <https://doi.org/10.33111/sedu.2022.50.045.055>.
169. Піскунова О., & Клочко, Р. (2020). Кластеризація клієнтів інтернет-магазину за рівнем їхньої купівельної активності. *Матеріали XII міжнародної*

науково-практичної конференції «Сучасні проблеми моделювання соціально-економічних систем» (09-10 квітня 2020 р.). Братислава – Харків, ВШЕМ – ХНЕУ ім. С. Кузнеця.

170. Піскунова О., & Клочко, Р. (2020). Нова парадигма вітчизняного маркетингу на підґрунті Big Data статистики. *Нові джерела та методи поширення даних у статистиці: матеріали XVIII Міжнародної науково-практичної конференції з нагоди Дня працівників статистики, Київ, «Інформаційно-аналітичне агентство»*, 37 – 41.

171. Піскунова, О., & Клочко, Р. (2020). Прогнозування часу до наступного замовлення клієнта інтернет-магазину на підґрунті методів машинного навчання. *Економічний простір*, 161, 147 – 152. <https://doi.org/10.32782/2224-6282/161-27>.

172. Піскунова, О., & Клочко, Р. (2023) Моделювання тривалості клієнтської активності в банку з використанням регресійної моделі виживання Вейбулла. Актуальні проблеми системного аналізу та моделювання процесів управління. За ред. В. Пономаренка, Л. Гур'янової, Я. Пеліової, Е. Ніжинського – Братислава-Харків, ВШЕМ – ХНЕУ ім. С. Кузнеця, 51 – 68.

173. Савич , О., Пилипчук, & В., Голіцин, А. (2023). Формування системи генерації додаткових продажів на принципах маркетингової аналітики. *Маркетинг і цифрові технології*, 7 (3), 52–65. <https://doi.org/10.15276/mdt.7.3.2023.4>.

174. Юнькова, О., & Володько, Т. (2021). Прогнозування популярності інтернет-курсів методами машинного навчання. *Моделювання та інформаційні системи в економіці*, 101, 5–11. <https://doi.org/10.33111/mise.101.14>.

ДОДАТКИ

ДОДАТОК А

ПЕРЕЛІК ХАРАКТЕРИСТИК СТАРТАПІВ ОТРИМАНИХ З РЕСУРСУ DEALROOM

Таблиця А.1

Перелік доступних характеристик стартапів

Поле	Значення
NAME	Назва клієнтаї
PROFILE URL	Посилання на сторінку клієнта
WEBSITE	Сайт клієнта
TAGLINE	Короткий опис
ADDRESS	Адреса
HQ REGION	Регіон головного офісу
HQ COUNTRY	Країна головного офісу
HQ CITY	Місто головного офісу
LATITUDE	Широта голвного офісу
LONGITUDE	Довгота головного офісу
LOCATIONS	Локація поточна
FOUNDING LOCATION	Локація заснування
TEAM (DEALROOM)	Команда
TEAM (EDITORIAL)	Команда
INVESTORS	Інвестори
EACH INVESTOR TYPES	Типи усіх інвесторів
LEAD INVESTORS	Основний інвестор
TOTAL FUNDING (EUR M)	Загальна сума інвестицій
LAST ROUND	Останній інвестиційний раунд (сума)
LAST FUNDING	Останній інвестиційний раунд (тип)
LAST FUNDING DATE	Останній інвестиційний раунд (дата)
FIRST FUNDING DATE	Перший інвестиційний раунд (даа\га)
SEED YEAR	Рік отримання інвестиції типу (Seed -один із перших етапів)
OWNERSHIPS	Власник
TAGS	Ключові слова
B2B/B2C	Тип взаємодії з клієнтами
REVENUE MODEL	Модель доходу
LAUNCH DATE	Дата заснування
CLOSING DATE	Дата закриття
INDUSTRIES	Ринок
SUB INDUSTRIES	Субкатегорія
DELIVERY METHOD	Метод доставки послуг
GROWTH STAGE	Етап росту
DEALROOM TAG	Ключові слова

Продовження таблиці А.1

YEARLY GROWTH (SIMILARWEB)	Річний приріст власних медіа каналів за даними платформи Alexa
DEALROOM TAG	Ключові слова
YEARLY GROWTH (SIMILARWEB)	Річний приріст власних медіа каналів за даними платформи Alexa
ALEXA GROWTH (ALL TIME)	Приріст власних медіа каналів за даними платформи Alexa
EMPLOYEES	Кількість співробітників
EMPLOYEES (2016,2017,2018,2019,2020,2021)	Кількість співробітників (2016,2017,2018,2019,2020,2021)
EMPLOYEES IN HQ country (2016,2017,2018,2019,2020,2021)	Кількість співробітників в головному офісі (2016,2017,2018,2019,2020,2021)
LAST KPI DATE	Остання дата постановки цілей росту
PROFIT (2016,2017,2018,2019)	Дохід
EBITDA (2016,2017,2018,2019)	EBITDA
REVENUE (2016,2017,2018,2019)	Прибуток
FINANCIALS CURRENCY	Валюта доходу
VALUATION	Оцінка вартості клієнта
VALUATION CURRENCY	Валюта оцінки
VALUATION (EUR)	Оцінка вартості клієнта в Євро
VALUATION DATE	Дата оцінки
FACEBOOK LIKES	Кількість лайків Facebook
TWITTER FOLLOWERS	Кількість підписників Twitter
TWITTER TWEETS	Кількість публікацій Twitter
TWITTER FAVORITES	Кількість вподобаних Twitter
SW TRAFFIC 6 MONTHS	Кількість відвідувачів сайту за останні 12 місяців
SW TRAFFIC 12 MONTHS	Кількість відвідувачів сайту за останні 6 місяців
ANGELLIST	Ангел (один із типів інвесторів)
FACEBOOK	Сторінка Facebook
TWITTER	Сторінка TWITTER
LINKEDIN	Сторінка LINKEDIN
GOOGLE PLAY LINK	Сторінка GOOGLE PLAY
ITUNES LINK	Сторінка ITUNES
EACH ROUND TYPE	Тип кожного інвестиційного раунду
EACH ROUND AMOUNT	Сума кожного інвестиційного раунду
EACH ROUND CURRENCY	Валюта кожного інвестиційного раунду
EACH ROUND DATE	Дата кожного інвестиційного раунду
TOTAL ROUNDS NUMBER	Кількість усіх інвестиційних раундів
EACH ROUND INVESTORS	Інвестор кожного інвестиційного раунду
LOGO	Логотип
FOUNDERS	Засновник
FOUNDERS STATUSES	Статус засновника
FOUNDERS GENDERS	Стать засновника
FOUNDERS IS SERIAL	Чи засновник є "серійним", тобто таким що постійно створює та продає бізнеси
FOUNDERS BACKGROUNDS	Рід занять засновника
FOUNDERS UNIVERSITIES	Університет засновника

Продовження таблиці А.1

FOUNDERS COMPANY EXPERIENCE	Досвід створення бізнесу у засновників
FOUNDERS FIRST DEGREE	Перший науковий ступінь засновника
FOUNDERS FIRST DEGREE YEAR	Перший науковий ступінь засновника (рік)
FOUNDERS LINKEDIN	LINKEDIN засновника
FOUNDERS FOUNDED COMPANIES TOTAL FUNDING	Сумарні інвестиції в бізнеси засновника
LISTS	Ключові слова
COMPANY STATUS	Статус клієнта
APP DOWNLOADS LATEST (IOS)	Кількість завантажень додатку (IOS)
APP DOWNLOADS 6 MONTHS (IOS)	Кількість завантажень додатку 6 місяців (IOS)
APP DOWNLOADS 12 MONTHS (IOS)	Кількість завантажень додатку 12 місяців (IOS)
APP DOWNLOADS LATEST (ANDROID)	Кількість завантажень додатку (ANDROID)
APP DOWNLOADS 6 MONTHS (ANDROID)	Кількість завантажень додатку 6 місяців (ANDROID)
APP DOWNLOADS 12 MONTHS (ANDROID)	Кількість завантажень додатку 12 місяців (ANDROID)
TRAFFIC COUNTRIES	Країни з яких заходять на сайт
TRAFFIC SOURCES	Джерела із яких заходять на сайт
SIMILARWEB RANK 3/6/12 MONTHS	Рейтинг сайту в системі Similar Web
EMPLOYEE RANK 3/6/12 MONTHS	Оцінка співробітниками
APP RANK 3/6/12 MONTHS	Рейтинг додатку в системі Similar Web
NUMBER OF ALUMNI EUROPEAN FOUNDERS THAT RAISED > 10M	Кількість засновників, що залучили 10M інвестицій
TECHNOLOGIES	Технології
INCOME STREAMS	Тип надходжень
TECH STACK DATA (BY PREDICTLEADS)	Ключові слова
TRADE REGISTER NUMBER	Номер реєстра торгівлі
SDGS	Цілі сталого розвитку
CORE SIDE VALUE	Цінності

Джерело: побудовано автором

ДОДАТОК Б

ОПИС КАТЕГОРІЙНИХ ЗМІННИХ ДЛЯ ПОБУДОВИ МОДЕЛЕЙ ПРОГНОЗУВАННЯ УСПІХУ СТАРТАПІВ

Таблиця Б.1

Групування категорій характеристики TECHNOLOGIES

Згруповані категорії	Початкові категорії	Опис
App_programs	Mobile app	Мобільні додатки
	Software	Програмне забезпечення
New_product	Connected device	Пристрої підключення
	Hardware	Програмне забезпечення
	Iot internetofthings	Інтернет речей
Online_services	Marketplace	Маркетплейси
	Other services	Інші сервіси
VR_AR_ML_AI	3d technology	3D технології
	Augmented reality	Доповнена реальність
	Deep tech	Технології глибокого навчання
	Virtual reality	Віртуальна реальність
	Artificial intelligence	Штучний інтелект
	Computer vision	Комп'ютерне бачення
	Machine learning	Машинне навчання
	Autonomous & sensor tech	Автономні та сенсорні технології
	Recognition technology	Технологія розпізнавання
	Big data	Великі дані
	Blockchain	Блокчейн
	Nanotech	Нанотехнології
	Natural language processing	Технології обробки природної мови

Джерело: побудовано автором

Таблиця Б.2

Групування категорій характеристики INDUSTRIES

Згруповані категорії	Початкові категорії	Опис
Heals_Food_Education	Health	Здоров'я
	Education	Освіта
	Food	Їжа
	Wellness beauty	Краса
	Kids	Діти
Enterprise_Software_Manufacturing	Telecom	Телекомунікації
	Fintech	Фінтех
	Robotics	Роботизація

Продовження таблиці Б.2

Enterprise_Software_Manufacturing	Energy	Енергетика
	Jobs recruitment	Підбір персоналу
	Enterprise software	Корпоративне програмне забезпечення
	Legal	Право
Creative_Services	Fashion	Мода
	Marketing	Маркетинг
	Media	Медіа
	Music	Музика
	Gaming	Ігри
Home_Travel_Hobbies	Real estate	Нерухомість
	Home living	Домашнє життя
	Security	Безпека
	Sports	Спорт
	Event tech	Івент-технології
	Travel	Подорожі
	Transportation	Транспорт
	Dating	Знайомства

Джерело: побудовано автором

Таблиця Б.3

Групування категорій характеристики HQ COUNTRY

Згруповані категорії	Початкові категорії
Ukraine	Ukraine
Other	United States
	Germany
	Australia
	United Kingdom
	Estonia
	Switzerland
	Cyprus
	Poland
	Montenegro
	Lithuania

Джерело: побудовано автором

Таблиця Б.4

Категорії характеристики B2B_B2C

Категорії	Опис
consumer	Споживач
business	Бізнес

Джерело: побудовано автором

Таблиця Б.5

Групування категорій характеристики FIRST_ROUND_TYPE

Згруповані категорії	Початкові категорії	Опис
SEED_EARLY_VC	SEED	Стартове фінансування
	EARLY VC	Ранній венчурний капітал
ANGEL_GRANT	ANGEL	Ангельські інвестиції
	GRANT	Гранти

Джерело: побудовано автором

Таблиця Б.6

Категорії характеристики TIME_TO_FIRST_FUNDING

Згруповані категорії	Опис
More_than_Year	Більше року
Same_year	Менше року
One_year	Один рік

Джерело: побудовано автором

Таблиця Б.7

Категорії характеристики REVENUE_MODEL

Категорії	Опис
marketplace_ecommerce	Маркетплейси та електронна комерція
saas	Програмне забезпечення як сервіс
manufacturing	Виробництво

Джерело: побудовано автором

Таблиця Б.8

Категорії характеристики INCOME_STREAMS

Категорії	Опис
commission	Комісії
subscription	Підписка
selling own inventory	Продаж продуктів

Джерело: побудовано автором

Таблиця Б.9

Категорії характеристики TOP_RANK_SW

Категорії	Опис
TOP	Присутність в рейтингах лідерів SimilarWeb
Other	Без присутності в рейтингах лідерів SimilarWeb

Джерело: побудовано автором

ДОДАТОК В

ЛІСТІНГ R-КОДУ ДЛЯ РЕАЛІЗАЦІЇ ВСІХ ЕТАПІВ МОДЕЛЮВАННЯ ПІДРОЗДІЛУ 2.2

#Підключення необхідних бібліотек

```
library(tidymodels)
```

```
library(readxl)
```

```
library(generics)
```

```
library(glmnet)
```

```
library(ranger)
```

#Завантаження та підготовка даних

```
Data <- read_excel("Data_model.xlsx")
```

```
Data$Success <- factor(Data$Success, levels = c("Yes", "No"))
```

```
startups_split <- initial_split(Data, strata = Success)
```

```
startups_training <- startups_split %>% training()
```

```
startups_test <- startups_split %>% testing()
```

#Створення процесу інженерії змінних

```
startups_recipe <- recipe(Success ~ ., data = startups_training)
```

```
%>% step_corr(all_numeric(), threshold = 0.95) %>%
```

```
step_normalize(all_numeric()) %>% step_dummy(all_nominal(), -  
all_outcomes())
```

#Модель CART

#Налаштування моделі

```
dt_model <- decision_tree() %>% set_engine('rpart') %>%
```

```
set_mode('classification')
```

```
startups_dt_wkfl <- workflow() %>% add_model(dt_model) %>%
```

```
add_recipe(startups_recipe)
```

#Тренування та тестування моделі

```
startups_dt_wkfl_fit <- startups_dt_wkfl %>% last_fit(split =  
startups_split)
```

```
startups_folds <- vfold_cv(startups_training, v = 5, strata =  
Success)
```

```
startups_metrics <- metric_set(roc_auc, sens, spec, f_meas,  
accuracy)
```

```
startups_dt_rs <- startups_dt_wkfl %>% fit_resamples(resamples =  
startups_folds, metrics = startups_metrics)
```

```

dt_rs_results <- startups_dt_rs %>% collect_metrics(summarize =
FALSE)
dt_rs_results %>% group_by(.metric) %>% summarize(min =
min(.estimate), median = median(.estimate), max = max(.estimate))

#Налаштування гіперпараметрів моделі
dt_tune_model <- decision_tree(cost_complexity = tune(), tree_depth
= tune(), min_n = tune()) %>% set_engine('rpart') %>%
set_mode('classification')
startups_tune_wkfl <- startups_dt_wkfl %>%
update_model(dt_tune_model)
dt_grid <- grid_random(parameters(dt_tune_model), size = 5)
dt_tuning <- startups_tune_wkfl %>% tune_grid(resamples =
startups_folds, grid = dt_grid, metrics = startups_metrics)
best_dt_model <- dt_tuning %>% select_best(metric = 'sens')

#Тренування та тестування моделі з новими гіперпараметрами
final_startups_wkfl <- startups_tune_wkfl %>%
finalize_workflow(best_dt_model)
startups_final_fit <- final_startups_wkfl %>% last_fit(split =
startups_split)
startups_dt_rs <- final_startups_wkfl %>% fit_resamples(resamples
= startups_folds, metrics = startups_metrics)
dt_rs_results <- startups_dt_rs %>% collect_metrics(summarize =
FALSE)
dt_rs_results %>% group_by(.metric) %>% summarize(min =
min(.estimate), median = median(.estimate), max = max(.estimate))

#Побудова кривої ROC
startups_final_fit %>% collect_predictions() %>% roc_curve(truth =
success , .pred_Yes) %>% autoplot()

```

#Модель Logistic regression

```

#Налаштування моделі
glm_model <- logistic_reg() %>% set_engine('glm') %>%
set_mode('classification')
startups_glm_wkfl <- workflow() %>% add_model(glm_model) %>%
add_recipe(startups_recipe)

#Тренування та тестування моделі

```

```

startups_glm_wkfl_fit <- startups_glm_wkfl %>%
last_fit(split = startups_split)
startups_glm_rs <- startups_glm_wkfl %>% fit_resamples(resamples =
startups_folds, metrics = startups_metrics)
glm_rs_results <- startups_glm_rs %>% collect_metrics(summarize =
FALSE)
glm_rs_results %>% group_by(.metric) %>% summarize(min =
min(.estimate), median = median(.estimate), max = max(.estimate))

#Налаштування гіперпараметрів моделі
glm_tune_model <- logistic_reg(penalty = tune(), mixture = 1) %>%
set_engine("glmnet")
glm_startups_tune_wkfl <- startups_glm_wkfl %>%
update_model(glm_tune_model)
glm_reg_grid <- tibble(penalty = 10^seq(-4, -1, length.out = 30))
glm_tuning <- glm_startups_tune_wkfl %>% tune_grid(startups_folds,
grid = glm_reg_grid, control = control_grid(save_pred = TRUE),
metrics = startups_metrics)
best_glm_model <- glm_tuning %>% select_best(metric = 'roc_auc')

#Тренування та тестування моделі з новими гіперпараметрами
glm_final_startups_wkfl <- glm_startups_tune_wkfl %>%
finalize_workflow(best_glm_model)
glm_startups_final_fit <- glm_final_startups_wkfl %>%
last_fit(split = startups_split)
startups_glm_rs <- glm_final_startups_wkfl %>%
fit_resamples(resamples = startups_folds, metrics =
startups_metrics)
glm_rs_results <- startups_glm_rs %>% collect_metrics(summarize =
FALSE)
glm_rs_results %>% group_by(.metric) %>% summarize(min =
min(.estimate), median = median(.estimate), max = max(.estimate))

#Побудова кривої ROC
glm_startups_final_fit %>% collect_predictions() %>%
roc_curve(truth = Success , .pred_Yes) %>% autoplot()

```

#Модель Random Forest

```

#Налаштування моделі

```

```

rf_mod <- rand_forest(trees = 1000) %>% set_engine("ranger",
importance = "impurity") %>% set_mode("classification")
startups_rf_wkfl <- workflow() %>% add_model(rf_mod ) %>%
add_recipe(startups_recipe)

#Тренування та тестування моделі
startups_rf_wkfl_fit <- startups_rf_wkfl %>% last_fit(split =
startups_split)
startups_rf_rs <- startups_rf_wkfl %>% fit_resamples(resamples =
startups_folds, metrics = startups_metrics)
rf_rs_results <- startups_rf_rs %>% collect_metrics(summarize =
FALSE)
rf_rs_results %>% group_by(.metric) %>% summarize(min =
min(.estimate), median = median(.estimate), max = max(.estimate))

#Налаштування гіперпараметрів моделі
rf_tune_model<- rand_forest(mtry = tune(), min_n = tune(), trees =
1000) %>% set_engine("ranger", num.threads = cores,importance =
"impurity") %>% set_mode("classification")
rf_startups_tune_wkfl <- startups_rf_wkfl %>%
update_model(rf_tune_model)
rf_tuning <- rf_startups_tune_wkfl %>% tune_grid(startups_folds,
grid = 25, control = control_grid(save_pred = TRUE), metrics =
startups_metrics)
best_rf_model <- rf_tuning %>% select_best(metric = 'sens')

#Тренування та тестування моделі з новими гіперпараметрами
rf_final_startups_wkfl <- rf_startups_tune_wkfl %>%
finalize_workflow(best_rf_model)
rf_startups_final_fit <- rf_final_startups_wkfl %>% last_fit(split
= startups_split)
startups_rf_rs <- rf_final_startups_wkfl %>%
fit_resamples(resamples = startups_folds, metrics =
startups_metrics )
rf_rs_results <- startups_rf_rs %>% collect_metrics(summarize =
FALSE)
rf_rs_results %>% group_by(.metric) %>% summarize(min =
min(.estimate), median = median(.estimate), max = max(.estimate))
#Побудова кривої ROC
rf_startups_final_fit %>% collect_predictions() %>%
roc_curve(truth = Success , .pred_Yes) %>% autoplot()

```

ДОДАТОК Г

ЛІСТІНГ R-КОДУ ДЛЯ РЕАЛІЗАЦІЇ ВСІХ ЕТАПІВ МОДЕЛЮВАННЯ ПІДРОЗДІЛУ 2.3

#Підключення необхідних бібліотек

```
library(tidyverse)
library(lubridate)
library(matrixStats)
library(Metrics)
library(caret)
library(doParallel)
library(modelr)
library(grid)
library(cluster)
library(NbClust)
```

#Завантаження та підготовка даних

```
Data_RFM <- read_excel("Data.xlsx")
Data_RFM <- Data_RFM %>% mutate(InvoiceNo=as.factor(Invoice),
StockCode=as.factor(StockCode), InvoiceDate=as.Date(InvoiceDate, '%Y
/%m/%d'), CustomerID=as.factor(CustomerID), Income =
Quantity*Price) %>% filter(!(Invoice %in% outliers))
train_rfm<- Data_RFM %>% filter(InvoiceDate < '2011/06/01')
test_rfm <- Data_RFM %>% filter(InvoiceDate > '2011/05/31')
```

#Обчислення показників RFM

```
max_date <- max(train_rfm$InvoiceDate)
df_RFM <- train_rfm %>% group_by(CustomerID) %>%
summarise(recency=as.numeric(max_date-max(InvoiceDate)),
frequenci=n_distinct(Invoice), monetary= sum(income))
```

#Модель кластеризації клієнтів k-means

#Підготовка даних для кластеризації

```
K_M <- df_RFM
row.names(K_M) <- K_M$CustomerID
K_M$CustomerID <- NULL
```

#Оцінка необхідної кількості кластерів методом середнього силуету

```
fviz_nbclust(scale(K_M), kmeans, method = "silhouette")
fviz_cluster(kmeans(scale(K_M) , centers = 6, nstart = 25), data =
scale(K_M))
```

#Реалізація кластеризації

```
KMEANS_member <- kmeans(scale(K_M) , 6, nstart = 25)
df_RFM <- df_RFM %>% mutate(K_M = KMEANS_member$cluster)
df_RFM$K_M[df_RFM$K_M==1] <- 'Lost'
df_RFM$K_M[df_RFM$K_M==2] <- 'Av_wholesales'
df_RFM$K_M[df_RFM$K_M==3] <- 'Recent'
df_RFM$K_M[df_RFM$K_M==4] <- 'Act_retail'
df_RFM$K_M[df_RFM$K_M==5] <- 'Av_retail'
df_RFM$K_M[df_RFM$K_M==6] <- 'Act_wholesales'
```

#Візуалізація утворених кластерів

```
df_RFM %>% ggplot() + geom_boxplot(aes(Client_type, recency))
df_RFM %>% ggplot() + geom_boxplot(aes(Client_type, frequency))
df_RFM %>% ggplot() + geom_boxplot(aes(Client_type, monetary))
```

#Прогнозування часу до наступного замовлення

#Визначення залежної змінної

```
a <- test_rfm %>% group_by(CustomerID) %>% summarise( min_day =
min(InvoiceDate)) %>% unique()
b <- train_rfm %>% group_by(CustomerID) %>% summarise( max_day =
max(InvoiceDate)) %>% unique()
c <- left_join(b,a) %>% mutate(d_diff = difftime(min_day,
max_day,units = c("days"))) %>% select(CustomerID,d_diff) %>%
unique()
user <- train_rfm %>% select(CustomerID) %>% unique() %>%
left_join(c)
user$d_diff[is.na(user$d_diff)] <- 999
```

#Інженерія ознак моделі

```
s <- train_rfm %>% group_by( CustomerID, InvoiceDate) %>% unique()
%>% summarise(n = n()) %>% select(CustomerID, InvoiceDate) %>%
group_by(CustomerID) %>% mutate(id = row_number())
id <- s %>% group_by(CustomerID) %>% summarise(Tmax= max(id)) %>%
mutate( T3 = Tmax -3 , T2= Tmax - 2, T1 = Tmax - 1) %>% filter(T3
> 0)
```

```

sw <- inner_join(s, id, by = c("id" = "Tmax", "CustomerID"))
names(sw)[names(sw) == "InvoiceDate"] <- "Max_d"
sw1 <- left_join(sw, s, by = c("T1" = "id", "CustomerID"))
names(sw1)[names(sw1) == "InvoiceDate"] <- "T1_d"
sw2 <- left_join(sw1, s, by = c("T2" = "id", "CustomerID"))
names(sw2)[names(sw2) == "InvoiceDate"] <- "T2_d"
sw3 <- left_join(sw2, s, by = c("T3" = "id", "CustomerID"))
names(sw3)[names(sw3) == "InvoiceDate"] <- "T3_d"
mod <- sw3[, c(1,2,7,8,9)]
mod <- mod %>% mutate(Dif1 = as.numeric(Max_d - T1_d), Dif2 =
as.numeric(T1_d - T2_d), Dif3 = as.numeric(T2_d - T3_d))
mod$Ave <- rowMeans(mod[,6:8])
StDev <- apply(mod[,6:8], 1, sd)
mod <- cbind(mod, StD = StDev)
Mod <- left_join(user, mod[, c(1,6:10)])

```

```

#Додавання обчислених, на попередньому етапі, кластерів
Mod <- left_join(Mod, df_RFM %>% select(CustomerID, K_M)) %>%
filter(d_diff > 0)
Mod <- drop_na(Mod)
Mod_1 <- Mod[, c(-8, -9, -10)]
row.names(Mod_1) <- Mod_1$CustomerID
Mod_1$CustomerID <- NULL
Mod_1$d_diff <- as.numeric(Mod_1$d_diff)
Mod_1$Client_Type <- as.factor(Mod_1$Client_Type)

```

```

#Перетворення залежної змінної з неперервної в категорійну
Mod_1$d_diff[Mod_1$d_diff < 60] <- 'Next_month'
Next_m <- Mod_1 %>% filter(d_diff == 'Next_month')
Mod_1 <- Mod_1 %>% filter(d_diff != 'Next_month')
Mod_1$d_diff[Mod_1$d_diff < 200] <- 'Next_quartel'
Next_qu <- Mod_1 %>% filter(d_diff == 'Next_quartel')
Mod_1 <- Mod_1 %>% filter(d_diff != 'Next_quartel')
Mod_1$d_diff[Mod_1$d_diff < 500] <- 'year'
Next_y <- Mod_1 %>% filter(d_diff == 'year')
Mod_1 <- Mod_1 %>% filter(d_diff != 'year')
Mod_1$d_diff[Mod_1$d_diff > 500] <- 'without'
Mod_1 <- rbind(Mod_1, Next_m, Next_qu, Next_y)

```

```

#Реалізація алгоритмів машинного навчання

```

```

control <- trainControl(method="cv", number=10)
metric <- "Accuracy"
index <- sample(1:nrow(Mod_1), round(0.75*nrow(Mod_1)))
train <- Mod_1[index,]
test <- Mod_1[-index,]

#Модель LDA
fit.lda <- train(d_diff ~ ., data = train, method="lda",
metric=metric, trControl=control)
#Модель CART
fit.cart <- train(d_diff ~ ., data = train, method="rpart",
metric=metric, trControl=control)
#Модель kNN
fit.knn <- train(d_diff ~ ., data = train, method="knn",
metric=metric, trControl=control)
#Модель SVM
fit.svm <- train(d_diff ~ ., data = train, method="svmRadial",
metric=metric, trControl=control)
#Модель Random Forest
fit.rf <- train(d_diff ~ ., data = train, method="rf",
metric=metric, trControl=control)

#Обчислення ефективності побудованих моделей
results <- resamples(list(lda=fit.lda, cart=fit.cart, knn=fit.knn,
svm=fit.svm, rf=fit.rf))
predictions <- predict(fit.rf, test)
confusionMatrix(predictions, test$d_diff)

```

ДОДАТОК Д

ЛІСТІНГ R-КОДУ ДЛЯ РЕАЛІЗАЦІЇ ВСІХ ЕТАПІВ МОДЕЛЮВАННЯ

ПІДРОЗДІЛУ 3.2

```
#Підключення необхідних бібліотек
library(tidyverse)
library(survival)
library(survminer)
library(reshape2)
library(rms)
library(survMisc)
library(ggplot2)

#Завантаження даних
clients_to_model_final <- read_excel("data.xlsx")

#Моделювання тривалості клієнтської активності методом регресії
виживання Вейбулла

#Підготовка даних до моделювання
article_table_new <- clients_to_model_final %>% filter(IR_001 ==
0) %>% filter(ACTIVE_CREDIT_CARD == 0)

#Моделювання впливу віку клієнтів на рівень виживання
wb_age <- survreg(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
AGE_group, article_table_new)
summary(wb_age)
AIC(wb_age)

#Візуалізація кривих виживання
newdat <- expand.grid( AGE_group =
levels(article_table_new$AGE_group))
surv <- seq(.99, .01, by = -.01)
t_2 <- predict(wb_age, type = 'quantile', p = 1-surv, newdata =
newdat)
surv_wbmod_wide <- cbind(newdat, t_2)
surv_wmod <- melt(surv_wbmod_wide, id.vars = c( "AGE_group"),
variable.name = "surv_id", value.name = "time")
surv_wmod$surv <- surv[as.numeric(surv_wmod$surv_id)]
surv_wmod[,c("upper", "lower", "std.err", "strata")] <- NA
```

```
ggsurvplot_df(surv_wmod, surv.geom = geom_line, color =
"AGE_group", legend.title = NULL)
```

```
#Моделювання впливу доходу клієнтів на рівень виживання
```

```
wb_income <- survreg(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
INCOME_group, article_table_new)
summary(wb_income)
AIC(wb_income)
```

```
#Візуалізація кривих виживання
```

```
newdat <- expand.grid( INCOME_group =
levels(article_table_new$INCOME_group))
t_2 <- predict(wb_income, type = 'quantile', p = 1-surv, newdata =
newdat)
surv_wbmod_wide <- cbind(newdat, t_2)
surv_wmod <- melt(surv_wbmod_wide, id.vars = c( "INCOME_group"),
variable.name = "surv_id", value.name = "time")
surv_wmod$surv <- surv[as.numeric(surv_wmod$surv_id)]
surv_wmod[,c("upper", "lower", "std.err", "strata")] <- NA
ggsurvplot_df(surv_wmod, surv.geom = geom_line, color =
"INCOME_group", legend.title = NULL)
```

```
#Моделювання впливу регіону проживання клієнтів на рівень
виживання
```

```
wb_region <- survreg(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
REGION_group, article_table_new)
summary(wb_region)
AIC(wb_region)
```

```
#Візуалізація кривих виживання
```

```
newdat <- expand.grid(REGION_group =
levels(article_table_new_new$REGION_group))
t_2 <- predict(wb_region, type = 'quantile', p = 1-surv, newdata =
newdat)
surv_wbmod_wide <- cbind(newdat, t_2)
surv_wmod <- melt(surv_wbmod_wide, id.vars = c( "REGION_group"),
variable.name = "surv_id", value.name = "time")
surv_wmod$surv <- surv[as.numeric(surv_wmod$surv_id)]
surv_wmod[,c("upper", "lower", "std.err", "strata")] <- NA
ggsurvplot_df(surv_wmod, surv.geom = geom_line, color =
"REGION_group", legend.title = NULL)
```

```

#Моделювання впливу статі клієнтів на рівень виживання
wb_gender <- survreg(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
GENDER, article_table_new)
summary(wb_gender)
AIC(wb_gender)

#Візуалізація кривих виживання
newdat <- expand.grid(GENDER =
levels(article_table_new_new$GENDER))
t_2 <- predict(wb_gender, type = 'quantile', p = 1-surv,
newdata = newdat)
surv_wbmod_wide <- cbind(newdat, t_2)
surv_wmod <- melt(surv_wbmod_wide, id.vars = c( "GENDER"),
variable.name = "surv_id", value.name = "time")
surv_wmod$surv <- surv[as.numeric(surv_wmod$surv_id)]
surv_wmod[,c("upper", "lower", "std.err", "strata")] <- NA
ggsurvplot_df(surv_wmod, surv.geom = geom_line, color = "GENDER",
legend.title = NULL)

#Моделювання рівня виживання клієнтів з урахуванням впливу усіх
факторів
wb_total <- survreg(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
REGION_group + GENDER + AGE_group+INCOME_group, article_table_new)
AIC(wb_total)
summary(wb_total)

#Моделювання рівня виживання клієнтів з урахуванням впливу
найбільш значущих класів наявних характеристик
dummies <- dummyVars(~ AGE_group+GENDER+REGION_group+INCOME_group,
data = article_table_new)
article_table_new_new <- cbind(article_table_new,predict(dummies,
newdata = article_table_new))
wb_total_t <- survreg(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
`INCOME_group70-150` + `INCOME_group0-35` + `INCOME_group150+` +
INCOME_group35-70` + REGION_groupSM_KH_DP_DN_ZP_MK_KS_OD +
REGION_groupCV_IF_KM_LT_LV_RV_TE_UZ + REGION_groupKV +
REGION_groupCN_CK_PL_KR_VN_ZT + GENDER.M + GENDER.F + `AGE_group0-
35` + `AGE_group35-45` + `AGE_group45-55` + `AGE_group55+`,
article_table_new_new)
weibull_model_step <- stepAIC(wb_total_t, direction = "both")

```

```
summary(weibull_model_step)
AIC(weibull_model_step)
```

#Моделювання тривалості клієнтської активності методом регресії виживання Кокса

```
#Підготовка даних до моделювання
```

```
article_table <- clients_to_model_final %>% filter(IR_001 == 1)
```

```
#Моделювання впливу продажу кредитної картки на рівень виживання
cxmod_CREDIT_CARD <- coxph(Surv(TIME_TO_DEACTIVATION, DEACTIVATED)
~ CREDIT_CARD, data = article_table)
coef(cxmod_CREDIT_CARD)
summary(cxmod_CREDIT_CARD)
```

```
#Перевірка виконання припущення про пропорційні небезпеки
cox.zph(cxmod_CREDIT_CARD)
ggcoxzph(test.ph)
```

```
#Візуалізація кривих виживання, з урахуванням впливу продажу
кредитної картки
fitCPH_CREDIT_CARD<- cph(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
CREDIT_CARD, data=article_table, x = TRUE, y = TRUE, surv = TRUE)
newdat <- with(article_table, data.frame(CREDIT_CARD = c(0, 1)))
rownames(newdat) <- letters[1:2]
cxsf <- survfit(fitCPH_CREDIT_CARD, data = article_table, newdata =
newdat, conf.type = "none")
surv_cxmod0 <- surv_summary(cxsf)
pid <- as.character(surv_cxmod0$strata)
m_newdat <- newdat[pid,]
surv_cxmod <- cbind(surv_cxmod0, m_newdat)
ggsurvplot_df(surv_cxmod, legend.title = "", legend = "none", censor
= FALSE)
```

```
#Моделювання впливу активації кредитної картки на рівень виживання
cxmod_active_creit_card <- coxph(Surv(TIME_TO_DEACTIVATION,
DEACTIVATED) ~ ACTIVE_CREDIT_CARD, data = article_table)
coef(cxmod_active_creit_card)
summary(cxmod_active_creit_card)
```

```

#Перевірка виконання припущення про пропорційні небезпеки
cox.zph(coxmod_active_creit_card)
ggcoxzph(test.ph)
#Візуалізація кривих виживання, з урахуванням впливу активації
кредитної картки

fitCPH_article <- cph(Surv(TIME_TO_DEACTIVATION, DEACTIVATED) ~
ACTIVE_CREDIT_CARD, data=article_table, x = TRUE, y = TRUE, surv =
TRUE)
newdat <- with(article_table, data.frame(ACTIVE_CREDIT_CARD = c(0,
1)))
rownames(newdat) <- letters[1:2]
cxsf <- survfit(fitCPH_article, data = article_table, newdata =
newdat, conf.type = "none")
surv_cxmod0 <- surv_summary(cxsf)
pid <- as.character(surv_cxmod0$strata)
m_newdat <- newdat[pid,]
surv_cxmod <- cbind(surv_cxmod0, m_newdat)
ggsurvplot_df(surv_cxmod, legend.title = "", legend = "none", censor
= FALSE)

```

ДОДАТОК Е

ЛІСТІНГ R-КОДУ ДЛЯ РЕАЛІЗАЦІЇ ВСІХ ЕТАПІВ МОДЕЛЮВАННЯ ПІДРОЗДІЛУ 3.3

```

#Підключення необхідних бібліотек
library(tidyverse)
library(cluster)
library(NbClust)
library(factoextra)
library(readxl)
library(NbClust)

#Завантаження та підготовка даних
data <- read_excel("data.xlsx")
row.names(data) <- data$CLID
data$CLID <- NULL
scale_d <- scale(data)

#Вибір оптимальної кількості кластерів k
nc <- NbClust(scale_d, min.nc = 2, max.nc = 10, method = "kmeans")
table(nc$Best.n[1,])
nb_results <- as.data.frame(nc$All.index)
nb_best <- as.data.frame(nc$Best.nc)

#Реалізація алгоритму k-means
k_members_7 <- kmeans(scale_d,centers=7, nstart=50)
k_members_6 <- kmeans(scale_d,centers=6, nstart=50)

#Додавання створених кластерів до початкових даних
data_7 <- data %>% mutate(K_M = k_members_7$cluster)
data_6 <- data %>% mutate(K_M = k_members_6$cluster)

#Калькуляція середніх характеристик утворених кластерів
cluster_7 <- as.data.frame(aggregate(data_1[,2:11], by = list(
  k_members_7$cluster),median))
cluster_6 <- as.data.frame(aggregate(data_1[,2:11], by = list(
  k_members_6$cluster),median))

```

ДОДАТОК Є

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Праці, в яких опубліковані основні наукові результати дисертації

Статті у наукових фахових виданнях України та зарубіжних виданнях, проіндексованих у базах даних Scopus та Web of Science

1. Піскунова, О., & Ключко, Р. (2020). Прогнозування часу до наступного замовлення клієнта інтернет-магазину на підґрунті методів машинного навчання. *Економічний простір*, 161, 147–152. <https://doi.org/10.32782/2224-6282/161-27> (0.7 д.а. , з них 0.6 д.а. авторські: автором розроблено та реалізовано прогнозування часу до наступного замовлення клієнтів за допомогою методів машинного навчання).

2. Piskunova, O., Ligonenko, L., Klochko, R., & Frolova, T., Bilyk, T. (2021). Applying machine learning approach to start-up success prediction. *Scientific Horizons*, 24(11), 72–84. [https://doi.org/10.48077/scihor.24\(11\).2021.72-84](https://doi.org/10.48077/scihor.24(11).2021.72-84) (1.2 д.а., з них 0.9 д.а. авторські: автором реалізовано попередній процес обробки даних, зроблена побудова та аналіз моделі, що ґрунтується на використанні методів машинного навчання).

3. Klochko, R., & Piskunova, O. (2023). Marketing segmentation of Banks' corporate clients based on data mining technique. *Statistics in Transition New Series*, 24(5), 147–166. <https://doi.org/10.59170/stattrans-2023-068> (1.2 д.а., з них 1.1 авторські: автором реалізовано модель кластеризації клієнтів банку та розроблено цінові стратегії для кожного із виявлених сегментів).

4. Piskunova, O., Klochko, R., Bilyk, T., & Frolova, T. (2023). Modelling the bank customer activity duration based on the Cox Econometric Survival Model. *Financial and Credit Activity Problems of Theory and Practice*, 6(53), 21–31. <https://doi.org/10.55643/fcaptp.6.53.2023.4244> (0.9 д.а., з них 0.8 авторські: автором зроблена побудова та аналіз моделей активності роздрібних клієнтів банку, що ґрунтуються на використанні методу регресії виживання Кокса).

Розділ у колективній монографії

5. Піскунова, О., & Ключко, Р. (2023) Моделювання тривалості клієнтської активності в банку з використанням регресійної моделі виживання Вейбулла. Актуальні проблеми системного аналізу та моделювання процесів управління. За ред. В. Пономаренка, Л. Гур'янової, Я. Пеліової, Е. Ніжинського – Братислава-Харків, ВШЕМ – ХНЕУ ім. С. Кузнеця, 51–68. (0.85 д.а., з них 0.75 авторські: автором зроблена побудова та аналіз моделей активності роздрібних клієнтів банку, що ґрунтуються на використанні методу регресії виживання Вейбулла)

Праці, які засвідчують апробацію матеріалів дисертації

6. Piskunova, O., & Klochko, R. (2020). Classification of e-commerce customers based on Data Science techniques. *CEUR Workshop Proceedings*, 2649, 6 – 20 (0.8 д.а., з них 0.7 авторські: автором зроблена побудова та порівняльний аналіз моделей кластеризації та класифікації клієнтів онлайн-магазину; форма участі – очна).

7. Піскунова О., & Ключко, Р. (2020). Нова парадигма вітчизняного маркетингу на підґрунті Big Data статистики. Нові джерела та методи поширення даних у статистиці: матеріали XVIII Міжнародної науково-практичної конференції з нагоди Дня працівників статистики, Київ, «Інформаційно-аналітичне агентство», 37–41 (0.25 д.а, з них авторські 0.2: автором зроблено аналіз ключових напрямків розвитку маркетингу в епоху Big Data статистики; форма участі – заочна).

8. Піскунова О., & Ключко, Р. (2020). Кластеризація клієнтів інтернет-магазину за рівнем їхньої купівельної активності. Матеріали XII міжнародної науково-практичної конференції «Сучасні проблеми моделювання соціально-економічних систем» (09-10 квітня 2020 р.). Братислава – Харків, ВШЕМ – ХНЕУ ім. С. Кузнеця (0.15 д.а., з них 0.1 авторські: автором запропоновано підхід до моделювання рівня купівельної активності клієнтів інтернет-магазину; форма участі – заочна).

«ЗАТВЕРДЖУЮ»:

Проректор з наукової роботи
Київського національного
економічного університету
імені Вадима Гетьмана,
д.е.н. проф. Антонюк Л.Л.



«11» грудня 2023 р.

«ЗАТВЕРДЖУЮ»:

Проректор з науково-педагогічної
роботи Київського національного
економічного університету
імені Вадима Гетьмана
д.е.н. проф. Колот А.М.



«11» грудня 2023 р.

ДОВІДКА

про впровадження у навчальний процес
результатів дисертаційної роботи

Клочка Ростислава Сергійовича на тему:

«Моделювання поведінки клієнтів у системі маркетингового менеджменту»

Результати дисертаційної роботи «Моделювання поведінки клієнтів у системі маркетингового менеджменту» здобувача наукового ступеня доктора філософії (PhD) Клочка Ростислава Сергійовича використовуються у навчальному процесі кафедри математичного моделювання та статистики Київського національного економічного університету імені Вадима Гетьмана.

Зокрема, розроблена Клочком Р.С. методика прогнозування та моделювання поведінки клієнтів у системі маркетингового менеджменту на підґрунті застосування сучасних економетричних моделей та алгоритмів машинного навчання використовується під час викладання навчальної дисципліни «Кількісні методи в економіці (в R)» для освітньо-професійної програми другого (магістерського) рівня «Економічна кібернетика і Дата Сайнс» за спеціальністю 051 «Економіка», а також навчальної дисципліни «Економетричне моделювання в ІТ-середовищі» для низки освітньо-професійних програм КНЕУ імені Вадима Гетьмана третього (освітньо-наукового) рівня.

Завідувач кафедри математичного
моделювання та статистики
КНЕУ імені Вадима Гетьмана,
к.ф.-м.н., професор

Галина ВЕЛИКОІВАНЕНКО



LEOPOLIS S.L.U
Carrer Esteve Albert, Bloc A, 1-2
AD500 Andorra la Vella,
Principality of Andorra

№ 3-6-11-23 , від 06.11.2023 р.

Довідка
Про впровадження результатів дисертаційної роботи
Клочко Ростислава Сергійовича
на тему:
«Моделювання поведінки клієнтів у системі маркетингового менеджменту»

На здобуття наукового ступеню доктора філософії (PhD) за спеціальністю 051 – «Економіка»

Видана здобувачу кафедри математичного моделювання та статистики Київського національного економічного університету імені Вадима Гетьмана Клочко Ростиславу Сергійовичу та засвідчує, що результати його дисертаційного дослідження були використані компанією Leopolis Group при розробці стратегії маркетингового менеджменту для українського банку, зокрема застосовано розроблені економетричні моделі виживання для визначення факторів, що впливають на тривалість активності клієнтів з автокредитом .

Завдяки теоретичним та практичним результатам та рекомендаціям Клочко Р.С., що висвітлені в результатах дослідження, вдалося підвищити рівень утримання клієнтів банку шляхом розробки стратегії додаткових продажів для цільових сегментів, які характеризуються підвищеним ризиком короточасного закриття.

Окрім цього, практичний інтерес представляє, запропонований в дисертаційному дослідженні, підхід до розробки стратегії ціноутворення для бізнес клієнтів банку на підґрунті методів машинного навчання. Розроблені практичні рекомендації дозволили залучити нові проекти по консультуванню українських банків стосовно шляхів підвищення ефективності їх маркетингової діяльності.

Директор,

B. Davidko
LEOPOLIS S.L.U.
C/ Esteve Albert Ed. La Roureda, Bl. A 1-2
AD500 Santa Coloma, PRINCIPAT D'ANDORRA
L-716625-U

Давидко В.А.



Товариство з обмеженою відповідальністю
«АДАПТИВ МАРКЕТИНГ ІННОВЕЙШЕНЗ»

01135, Київ, вул. Павлівська, 29

Код ЄДРПОУ 41564468

№ 1411-23 , від 14.11.2023 р.

Довідка

Про впровадження результатів дисертаційної роботи

Клочко Ростислава Сергійовича

на тему:

«Моделювання поведінки клієнтів у системі маркетингового менеджменту»

На здобуття наукового ступеню доктора філософії (PhD) за спеціальністю 051 – «Економіка»

Видана аспіранту кафедри математичного моделювання та статистики Київського національного економічного університету імені Вадима Гетьмана Клочко Ростиславу Сергійовичу та засвідчує, що результати проведеного дослідження були використані при розробці маркетингової стратегії для українських компаній, зокрема в категорії ритейлу.

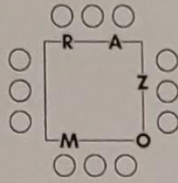
В процесі вирішення задачі, щодо підвищення ефективності маркетингових комунікацій, були застосовані запропоновані методи та моделі машинного навчання, які висвітлені в результатах дослідження. Зокрема, компанією ТОВ «АДАПТИВ МАРКЕТИНГ ІННОВЕЙШЕНЗ» було застосовано запропоновану систему моделей на підґрунті методів машинного навчання для вирішення задачі передбачення часу до наступного замовлення клієнтів з метою оптимізації маркетингових інвестицій та підвищення рівня утримання клієнтів.

Запропоновані у дисертаційному дослідженні Клочко Р.С концептуальні положення щодо моделювання поведінки клієнтів дозволили: підвищити ефективність діяльності підприємства у сфері ритейл; підвищити рівень утримання клієнтів; підвищити ефективність комунікацій з клієнтами через рекламні розсилки, що посприяло зниженню неефективних витрат.

Виконавчий директор



Наталія БОЙКО



Товариство з обмеженою відповідальністю
«РАЗОМ КОМ'ЮНІКЕЙШНЗ»
 03067, м. Київ, вулиця Гарматна, будинок 4
 Код ЄДРПОУ 41172691

№ 18-09/12, від 18.09.2023 р.

Довідка
 Про впровадження результатів дисертаційної роботи
Клочко Ростислава Сергійовича
 на тему:
**«Моделювання поведінки клієнтів
 у системі маркетингового менеджменту»**

На здобуття наукового ступеню доктора філософії (PhD) за спеціальністю 051 –
 «Економіка»

Ключові результати та пропозиції щодо моделювання поведінки клієнтів розроблені та запропоновані у дисертації Клочко Ростислава Сергійовича, що базуються на інструментарії машинного навчання, мають практичну цінність та використовувалися у діяльності ТОВ «РАЗОМ КОМ'ЮНІКЕЙШНЗ».

Зокрема, результати дослідження знайшли своє відображення при проведенні аналізу поведінки українського підприємства та його конкурентів з використанням відкритих даних. Як результат, запропоновані інтелектуальні технології моделювання поведінки бізнес-клієнтів в умовах обмеженості даних дозволили:

- виявити ключові фактори, які впливають на досягнення фінансових цілей підприємством в цільовій категорії;
- розробити для клієнта маркетингову стратегію з підвищення конкурентних переваг, що включає в себе стратегії маркетингового позиціонування та комунікацій;
- спрогнозувати ефективність запропонованих заходів для досягнення поставлених цілей.

Заступник директора
 з аналізу даних



Фаренюк Я. В.