

<https://doi.org/10.31891/2307-5732-2026-365-41>

УДК 004.891.2:519.816

### ПЕЛЕЩАК ІВАН

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-7481-8628>

e-mail: [ivan.r.peleshchak@lpnu.ua](mailto:ivan.r.peleshchak@lpnu.ua)

### ШАРІФОВ АРСЕН

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0000-0193-4073>

e-mail: [arsen.sharifov@proton.me](mailto:arsen.sharifov@proton.me)

### КІСЬ ЯРОСЛАВ

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0003-3421-2725>

e-mail: [yaroslav.p.kis@lpnu.ua](mailto:yaroslav.p.kis@lpnu.ua)

## ПРОГНОЗУВАННЯ КОНВЕРСІЇ В ЦИФРОВОМУ СЕРЕДОВИЩІ НА ОСНОВІ МОДЕЛІ MLP+LSTM З ІНТЕГРАЦІЄЮ ПОКАЗНИКІВ СТІЙКОСТІ ТА ЕВОЛЮЦІЇ ПОВЕДІНКОВИХ СЕГМЕНТІВ

У статті запропоновано підхід до прогнозування конверсії в цифровому середовищі на основі гібридної моделі MLP+LSTM з інтеграцією показників стійкості та еволюції поведінкових сегментів. Сесії користувачів сегментуються за допомогою кластерного аналізу, після чого для кожного сегмента визначаються показники структурної стабільності за індексом CSAI та характеристики еволюційної динаміки за моделлю MONIC. Отримані сегментно-динамічні параметри включаються до розширеного вектора ознак прогностичної моделі. MLP-компонент опрацьовує агреговані session-level ознаки та показники, сформовані на основі CSAI і MONIC, тоді як LSTM-компонент аналізує реальний clickstream користувача як послідовність дій у межах сесії. Такий підхід дає змогу одночасно враховувати інтегральні характеристики сесії, часову структуру взаємодії користувача з цифровим середовищем і властивості сегмента з погляду його стійкості та трансформації. Практичне значення роботи полягає у підвищенні точності прогнозування конверсії та підтримці рішень щодо персоналізації її оптимізації цифрових сервісів.

**Ключові слова:** прогнозування конверсії, MLP, LSTM, поведінкові сегменти, індекс CSAI, модель MONIC

PELESHCHAK IVAN, SHARIFOV ARSEN, KIS YAROSLAV

Lviv Polytechnic National University

## CONVERSION FORECASTING IN A DIGITAL ENVIRONMENT BASED ON AN MLP+LSTM MODEL WITH INTEGRATION OF STABILITY AND EVOLUTION INDICATORS OF BEHAVIORAL SEGMENTS

The article proposes an approach to conversion forecasting in a digital environment based on a hybrid MLP+LSTM model with the integration of behavioral segment stability and evolution indicators. The relevance of the study is determined by the growing need to improve predictive analytics in digital services through the joint consideration of aggregated session characteristics, the temporal logic of user interaction, and the structural properties of behavioral segments to which user sessions belong. Unlike conventional approaches that rely only on tabular session-level features or only on sequential behavioral patterns, the proposed framework combines both types of information and supplements them with segment-dynamic indicators derived from cluster stability and cluster evolution analysis.

At the first stage of the study, user sessions are segmented by means of cluster analysis in order to identify relatively homogeneous behavioral groups within the digital environment. Such segmentation makes it possible to move from isolated session-level analysis to the examination of stable and unstable behavioral subsystems that reflect recurring interaction patterns. At the second stage, the structural stability of the identified segments is assessed using the CSAI index, which provides a quantitative basis for evaluating the internal consistency and resilience of each behavioral cluster. At the third stage, the MONIC model is applied to analyze the evolution of segments across temporal slices of the digital environment. This allows the detection of segment preservation, transition tendencies, merging, splitting, and other forms of structural transformation that may affect conversion behavior.

The indicators obtained from CSAI and MONIC are not treated only as auxiliary analytical results, but are directly integrated into the predictive contour in the form of additional segment-dynamic features. These features extend the descriptive space of the model and allow it to account not only for the current characteristics of an individual session, but also for the stability of the segment to which the session belongs and the potential direction of its structural transformation. In this way, the model incorporates both micro-level behavioral evidence and meso-level segment dynamics.

The predictive module is implemented as a hybrid MLP+LSTM architecture. The MLP component processes aggregated session-level features together with segment-dynamic parameters generated from CSAI and MONIC. The LSTM component processes the real user clickstream as an ordered sequence of actions performed within a session. The fusion of these two branches enables the model to simultaneously capture the integral session profile, the temporal structure of user interaction, and the stability and transformation properties of the corresponding behavioral segment. As a result, the proposed approach creates a more informative basis for forecasting conversion-related outcomes in complex digital systems.

The practical significance of the study lies in the possibility of improving conversion prediction accuracy, identifying unstable or transformation-prone behavioral segments, and supporting decision-making related to personalization, interface optimization, user journey redesign, and adaptive management of digital services. The proposed approach may also serve as a methodological foundation for broader decision support systems focused on behavioral analytics in complex digital environments.

**Keywords:** conversion forecasting, MLP, LSTM, behavioral segments, CSAI index, MONIC model

Стаття надійшла до редакції / Received 17.02.2026

Прийнята до друку / Accepted 16.04.2026

Опубліковано / Published 28.05.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© Пелешак Іван, Шаріфов Арсен, Кісь Ярослав

### Постановка проблеми у загальному вигляді

та її зв'язок із важливими науковими чи практичними завданнями

На тлі швидкого розвитку цифрового середовища управління користувачьким досвідом стало одним із ключових завдань для електронних сервісів. З метою персоналізації взаємодії з користувачами цифрові

платформи активно використовують аналіз цифрових слідів, а також алгоритми машинного навчання для прогнозування конверсії, наміру покупки та відтоку користувачів. Зокрема, у роботі [1] запропоновано моделі реального часу для прогнозування наміру покупки та раннього виявлення ризику залишення сайту, що демонструє високу ефективність індивідуального поведінкового прогнозування.

Поряд із цим, важливим напрямом є некерована кластеризація цифрових слідів, яка дає змогу виявляти приховані поведінкові патерни та формувати сегменти користувачів без попередньої розмітки даних [2]. Такі підходи дозволяють аналізувати поточну структуру аудиторії, виділяти типові й атипові групи користувачів, а також виявляти домінуючі моделі взаємодії з цифровим сервісом [2, 3]. Водночас самі по собі ці моделі переважно орієнтовані на фіксацію поточного стану системи та побудову сегментів на основі історичних даних.

Суттєвою проблемою залишається те, що в умовах змін цифрового продукту, таких як оновлення інтерфейсу, перегляд цінних політик, упровадження нового або модифікація наявного функціоналу, поведінкові сегменти можуть змінюватися, втрачати стійкість або трансформуватися у нові структури. У такому разі статичне представлення кластерів уже не є достатнім для підтримки управлінських рішень. Саме тому питання оцінки стійкості кластерних структур і прогнозування їхньої еволюції в часі набуває особливого значення [4, 5].

Наявні дослідження пропонують окремі інструменти для розв'язання цієї проблеми. Зокрема, роботи з оцінки стабільності кластеризації дають змогу кількісно аналізувати надійність і відтворюваність кластерних структур [4], а перспективна аналітика формує підґрунтя для моделювання сценаріїв типу «what-if» та оцінювання наслідків альтернативних управлінських рішень [6]. Проте ці підходи зазвичай розглядаються ізольовано: метрики стійкості кластерів використовуються переважно як інструмент внутрішньої валідації моделей [3, 6], тоді як сценарний аналіз фокусується на оптимізації цільових показників без належного врахування структурних змін у поведінкових сегментах [6].

Отже, науково-практична проблема полягає не стільки у відсутності окремих методів кластеризації, оцінки стабільності чи сценарного моделювання, скільки у відсутності інтегрованого підходу, який поєднував би результати сегментації цифрових слідів, кількісні показники стійкості кластерів та моделювання сценаріїв управлінських втручань [4, 5, 6]. Така інтеграція є необхідною для переходу від статичного аналізу поведінки користувачів до дослідження можливих сценаріїв еволюції сегментів під впливом змін у цифровому середовищі.

Розв'язання поставленої проблеми створює підґрунтя для більш повного аналізу користувацької поведінки в цифровому середовищі, оскільки дає змогу враховувати не лише належність сесій до певних поведінкових сегментів, а й рівень стійкості цих сегментів, їхню схильність до структурних змін та вплив зазначених характеристик на прогнозування конверсії. Це має важливе значення для розвитку методів інтелектуального аналізу цифрових слідів, оскільки дозволяє поєднати аналіз агрегованих параметрів сесії, реальної послідовності дій користувача та сегментно-динамічних властивостей поведінкової структури аудиторії. Практична цінність такого підходу полягає у підвищенні точності прогнозування цільових дій користувачів у цифрових сервісах, а також у розширенні можливостей персоналізації, оптимізації інтерфейсів та підтримки управлінських рішень у цифровому середовищі.

Наукова новизна нашого дослідження полягає у розробленні моделі прогнозування конверсії в цифровому середовищі на основі архітектури MLP+LSTM з інтеграцією показників стійкості та еволюції поведінкових сегментів. На відміну від наявних підходів, запропонована модель поєднує в єдиному прогностичному контурі агреговані session-level ознаки, реальні clickstream-послідовності дій користувача, показники структурної стійкості сегментів за індексом CSAI та характеристики їхньої еволюційної динаміки, визначені за моделлю MONIC. Такий підхід дозволяє враховувати не лише поточні характеристики окремої сесії, а й властивості поведінкового сегмента, до якого вона належить, з погляду його стабільності, змінності та потенційних трансформацій, що підвищує інформативність моделі прогнозування конверсії.

### **Аналіз досліджень та публікацій**

Аналіз літератури надає змогу виділити кілька основних напрямів, у межах яких досліджується поведінка користувачів у цифрових середовищах. До них належать моделі прогнозування намірів у реальному часі, методи некерованої кластеризації цифрових слідів, підходи до оцінки стабільності кластерних структур, а також прескриптивні моделі сценарного аналізу та зміни концепцій.

У роботі [1] запропоновано дворівневу архітектуру прогнозування купівельних намірів та відтоку користувачів на основі даних з цифрових слідів. Перший модуль використовує агреговані характеристики сесій, як-от тип відвідувача, тривалість перегляду сторінки, наближення до особливих дат у поєднанні з класичним класифікаторами, зокрема з багатощаровим перцептроном, випадковими лісами тощо. Другий модуль базується на рекурентній нейронній мережі з довгою короткостроковою пам'яттю, яка аналізує послідовність дій користувача та формує оцінку ймовірного дострокового залишення сайту в режимі реального часу. Сукупно ці моделі демонструють високий рівень точності при прогнозуванні поведінки окремого користувача.

У дослідженні [2] розглянуто не менш важливий напрям – некерована кластеризація цифрових слідів, яка дозволяє виявляти приховані поведінкові патерни без попередньої розмітки даних. Запропоновано підхід, у якому на основі послідовностей дій користувачів будують граф подібності: вузли відповідають користувачам, ребра – мірі схожості їхніх цифрових слідів, а кластери у цьому графі інтерпретуються як групи з подібними моделями поведінки. Для виділення ієрархії таких груп застосовується ієрархічна кластеризація за ітеративним «відсіканням» домінуючих ознак, а на наступних рівнях дрібніші підгрупи з відмінними ключовими

характеристиками. Продемонстровано, що такий метод дає змогу виявити як типові сегменти аудиторії, так і аномальні, зокрема «сплячих» користувачів та облікові записи з підозрілою активністю.

Питання стабільності кластерних структур та оцінки надійності кластеризації було розглянуто у роботі [4], в якій було запропоновано підхід до оцінки стійкості кластерів та кластеризації в цілому за допомогою ресумпшювання даних та порівняння отриманих розбиттів за допомогою особливих індексів стабільності. Такий підхід формує теоретичну базу для аналізу структурної стійкості сегментів.

Окреме дослідження [7] присвячене саме еволюції користувацьких сегментів. Автори аналізують, наскільки стабільними є кластери, побудовані за історією взаємодії з сервісом і показують, що значна частина сегментів трансформується, або ж взагалі зникає вже протягом кількох періодів спостережень. Це спостереження чітко підкреслює, що кластери, які були побудовані на основі певної статичної вибірки даних, не можна розглядати як незмінні «типи» користувачів і нам необхідно враховувати їх динаміку та зміну середовища.

Подальший розвиток цього напрямку простежується у роботах, де сегментація розглядається не ізольовано, а в поєднанні з прогнозуванням поведінки користувачів і підтримкою управлінських рішень. Зокрема, у дослідженні [8] запропоновано інтегровану архітектуру, яка поєднує прогнозування відтоку з подальшою сегментацією клієнтів. Такий підхід є важливим з погляду прикладної аналітики, оскільки дозволяє не лише класифікувати користувачів за схожими характеристиками, а й інтерпретувати сегменти через призму ризику втрати клієнта та вибору цільових стратегій утримання. У контексті даної роботи це підтверджує доцільність переходу від суто описової кластеризації до інтегрованого аналітичного контуру, у якому сегменти стають основою для подальших управлінських дій.

Окремий інтерес становлять дослідження, спрямовані на підвищення якості сегментації у складних багатовимірних просторах ознак. Так, автори роботи [9] запропонували розширений варіант регуляризованого K-means для сегментації клієнтів у високовимірних даних із корельованими змінними. Цінність цієї роботи полягає в тому, що вона фокусується не лише на самому факті побудови кластерів, а й на їхній інтерпретованості та стійкості в умовах надлишкових і взаємопов'язаних ознак. Для даного дослідження це є важливим, оскільки цифрові сліди користувачів також часто мають багатовимірний і корельований характер, а отже якість побудованої сегментації безпосередньо впливає на надійність подальшого аналізу стійкості та еволюції сегментів.

Ще один близький до тематики даної статті напрям представлено у роботі [10], де сегментація користувачів поєднується з аналізом їхнього проходження крізь етапи AIDA-моделі в середовищі Wi-Fi-реклами. Автори застосовують кластеризацію не тільки для виявлення однорідних груп користувачів, а і для опису їхнього просування вздовж ланцюга взаємодії — від уваги до цільової дії. Це важливо, оскільки демонструє зміщення фокусу сучасних досліджень від статичних профілів до поведінкових траєкторій, що концептуально узгоджується з потребою врахування динаміки сегментів і підтримує доцільність використання моделей, орієнтованих на еволюцію кластерів у часі.

Суттєвий внесок у розвиток динамічної сегментації зроблено в роботах Elsevier, де поведінка користувачів аналізується як часовий процес. Зокрема, автори роботи [11] запропонували методологію аналізу поведінки клієнтів на основі кластеризації часових рядів, що дозволяє виявляти домінуючі патерни змін у часі замість одноразового поділу клієнтської бази на статичні групи. Такий підхід є принципово важливим для даної роботи, оскільки підкреслює, що поведінковий сегмент слід розглядати не як фіксовану категорію, а як змінну структуру, чутливу до часового контексту та зовнішніх впливів.

Подібну логіку розвивають і новіші дослідження, де динаміка сегментів розглядається вже не лише на рівні опису, а і на рівні оптимізації. У роботі [12] запропоновано підхід до сегментації поведінки клієнтів на основі часових рядів із генетичною оптимізацією ваг ознак і подальшою спектральною кластеризацією. Автори прямо підкреслюють, що традиційні статичні методи сегментації часто не враховують часову мінливість поведінки, через що поступаються динамічним моделям у точності та прикладній корисності. Для даної статті ця робота є показовою, оскільки вона підтверджує доцільність переходу від одноразового виділення сегментів до аналітики, яка відстежує їх трансформацію у часі та допускає адаптацію під управлінські цілі.

Нарешті, окремої уваги заслуговують роботи, присвячені безпосередньо аналітиці еволюції кластерів і моделюванню сценаріїв типу «what-if». У публікації [13] представлено фреймворк Cluster Evolution Analytics і програмний пакет clusEvol, орієнтовані на дослідження того, як змінюються об'єкти та їхні кластери в часових зрізах, а також на побудову аналітики для сценарного аналізу. Важливість цієї роботи для даного дослідження полягає в тому, що вона методологічно наближається до ідеї переходу від статичної кластеризації до аналізу траєкторій кластерних змін та оцінювання наслідків альтернативних втручань. Саме це підсилює аргументацію на користь інтеграції показників стійкості кластерів, моделей еволюції сегментів і систем підтримки прийняття рішень в єдиному дослідницькому контурі.

Подальше розширення підходів до сегментації простежується у роботах, орієнтованих безпосередньо на середовище електронної комерції. Зокрема, у дослідженні [14] запропоновано модель LRFS для сегментації онлайн-покупців на основі поведінкових характеристик сесій. Важливість цієї роботи полягає в тому, що автори адаптують сегментацію саме до специфіки електронної торгівлі, де суттєве значення мають не лише транзакційні характеристики, а й особливості веб-поведінки користувача в межах сесії. Для даного дослідження це є релевантним, оскільки підтверджує доцільність використання цифрових слідів онлайн-взаємодії як основи для побудови поведінкових сегментів у динамічному цифровому середовищі.

Інший напрям розвитку представлений у роботі [15], де для сегментації використано багатоповедінкову

RFM-модель у поєднанні з удосконаленим алгоритмом самоорганізованих карт Кохонена. На відміну від класичного RFM-підходу, автори намагаються врахувати ширший спектр взаємодій клієнта з системою, що підвищує аналітичну чутливість моделі до різних типів користувацької активності. У контексті даної статті це є важливим, оскільки демонструє можливість переходу від спрощених моделей сегментації до багатофакторного опису поведінки, який краще узгоджується з природою цифрових слідів і створює підґрунтя для подальшої оцінки стійкості та змінності сегментів.

Суттєве значення для розвитку динамічної сегментації мають дослідження, у яких поведінка користувачів моделюється як послідовність часових станів. Так, у роботі [16] запропонували аналітичну рамку, що поєднує модель RFM із методами кластеризації часових рядів для динамічної сегментації. Автори прямо виходять із того, що одноразове віднесення клієнта до певного кластера не відображає його реальної поведінкової еволюції. Для даної роботи це особливо важливо, оскільки такий підхід концептуально наближається до ідеї аналізу не тільки наявної структури сегментів, а й змін цієї структури в часі, що є одним із ключових елементів побудови систем підтримки прийняття рішень у цифровому середовищі.

Ще один корисний напрям представлено у роботі [17], де запропоновано евристичний метод кластеризації GPHS для задач сегментації клієнтів. Цінність цієї праці полягає в тому, що вона фокусується на ситуаціях, коли клієнтські дані є нечіткими, неоднорідними або складними для поділу традиційними методами. Для даного дослідження це важливо з методологічної точки зору, оскільки цифрові сліди користувачів також часто характеризуються високою варіативністю, шумом і неоднозначністю меж між поведінковими групами. Отже, ця робота підсилює тезу про те, що на етапі побудови сегментів доцільно враховувати не лише класичні алгоритми, а й більш адаптивні методи кластеризації.

Важливе прикладне продовження теми наведено у роботі [18], де прогнозування онлайн-покупки поєднано з аналізом характеристик користувача та патернів перегляду сторінок, а також із кластеризацією на рівні сесій. Ця праця показує, що сегментація може бути не лише інструментом опису аудиторії, а й допоміжним елементом побудови предиктивних моделей у цифровій комерції. Для даної статті це має особливу цінність, оскільки підтверджує продуктивність інтегрованого підходу, в якому сегментація, прогнозування поведінки та аналіз цифрових слідів не розглядаються окремо, а формують спільний аналітичний контур для підтримки рішень.

Проведений аналіз показав, що сучасні дослідження у сфері цифрової аналітики переважно розвивають окремі напрями, пов'язані з кластеризацією користувацької поведінки, оцінюванням стійкості кластерної структури, аналізом часової еволюції сегментів та прогнозуванням індивідуальних цільових дій. Водночас недостатньо опрацьованим залишається підхід, за якого результати сегментації, показники стійкості поведінкових сегментів і характеристики їхньої трансформаційної динаміки інтегруються безпосередньо у модель прогнозування конверсії. Незважаючи на наявність значної кількості праць, присвячених кластеризації цифрових слідів, оцінці стабільності кластерів, динамічній сегментації та нейромережевому аналізу session-level і clickstream-даних, зазначені підходи здебільшого використовуються ізольовано. Це зумовлює методологічний розрив між сегментацією аудиторії, оцінкою стійкості поведінкових підсистем, аналізом їхньої еволюції та побудовою прогностичних моделей. Подолання такого розриву потребує створення інтегрованого підходу, у межах якого агреговані ознаки сесії, реальні послідовності користувацьких дій, індекс CSAI та показники MONIC спільно формують інформативний простір моделі прогнозування конверсії.

### Формулювання цілей статті

Метою статті є розроблення моделі прогнозування конверсії в цифровому середовищі на основі архітектури MLP+LSTM з інтеграцією показників стійкості та еволюції поведінкових сегментів. Запропонований підхід передбачає поєднання агрегованих характеристик користувацьких сесій, реальних clickstream-послідовностей дій користувачів, результатів сегментації, показників стійкості сегментів за індексом CSAI та характеристик їхньої еволюційної динаміки, отриманих за моделлю MONIC, у межах єдиного прогностичного контуру.

Об'єктом дослідження є поведінкові сегменти користувачів цифрового середовища, сформовані на основі цифрових слідів сесійної взаємодії та clickstream-даних.

Предметом дослідження є методи кластеризації цифрових слідів, оцінювання стійкості кластерної структури, аналізу еволюції поведінкових сегментів, формування сегментно-динамічних ознак та побудови моделі прогнозування конверсії на основі архітектури MLP+LSTM.

Для досягнення поставленої мети у статті розв'язуються такі завдання:

- сформувати простір агрегованих поведінкових ознак на основі цифрових слідів користувачів;
- виконати сегментацію користувацьких сесій методами кластерного аналізу;
- оцінити стійкість виявлених сегментів за допомогою індексу CSAI;
- сформувати сегментно-динамічні ознаки на основі результатів CSAI та MONIC;
- інтегрувати агреговані session-level ознаки, clickstream-послідовності та сегментно-динамічні характеристики в модель прогнозування конверсії на основі архітектури MLP+LSTM;
- виконати апробацію запропонованого підходу для прогнозування конверсії на основі набору даних Online Shoppers Purchasing Intention Dataset.

### Виклад основного матеріалу

Розроблена модель прогнозування конверсії реалізує інтегрований підхід до аналізу користувацької поведінки в цифровому середовищі. У межах цього підходу дані про користувацькі сесії перетворюються на кілька взаємопов'язаних типів подання: агреговані поведінкові ознаки, реальні clickstream-послідовності дій та сегментно-динамічні характеристики, сформовані на основі оцінки стійкості й еволюції поведінкових сегментів. Запропонований підхід поєднує аналіз внутрішньої стійкості кластерної структури з аналізом трансформацій сегментів у часі, що дає змогу враховувати як стабільність поведінкових груп, так і можливі переходи між ними при побудові моделі прогнозування конверсії.

Для моделювання та тестування використано набір даних «Online Shoppers Purchasing Intention» [19], який містить 12 330 сесій, кожна з яких представлена вектором  $X = \{n_i, d_i, g_i, c_i\}$ , де:

- $n_i$  і  $d_i$  – це навігаційні метрики, які показують кількість відвіданих сторінок адміністративного, інформаційного, товарного характерів та тривалість перебування на них.
- $g_i$  – це індикатор залученості «Google Analytics», який є показником відмов, виходу та цінності сторінки.
- $c_i$  – це контекстуальні змінні, тобто технічні характеристики пристрою: операційна система, регіон та часова близькість до особливих дат.

На етапі попередньої обробки даних вихідний набір було завантажено без виявлених пропусків значень, тому додаткове заповнення пропусків не виконувалося. Для подальшого аналізу відібрано лише числові поведінкові ознаки, а саме кількість та тривалість перегляду адміністративних, інформаційних і товарних сторінок, показник виходу, цінність сторінки та близькість сеансу до особливих дат. Після відбору ознак усі безперервні змінні було приведено до єдиного масштабу за допомогою стандартизації, що забезпечило коректність подальшого кластерного аналізу та оцінювання стійкості поведінкових сегментів.

У дослідженні використовувалися агреговані поведінкові ознаки сесії, сформовані за аналогією з підходом [1], у якому для аналізу наміру покупки застосовуються session-level характеристики взаємодії користувача з вебресурсом. До таких ознак належали: тривалість перегляду адміністративних сторінок, тривалість перегляду інформаційних сторінок, тривалість перегляду товарних сторінок, показник відмов, показник виходів, цінність сторінки та наближеність до особливого дня. Зазначені характеристики використовувалися як база для сегментації користувацьких сесій, оцінювання стійкості виявлених кластерів та подальшого формування сегментно-динамічних параметрів.

Для побудови прогностичного модуля дані подавалися у двох взаємодоповнювальних формах. Перша форма являла собою агрегований вектор session-level ознак, який використовувався в MLP-компоненті моделі. Друга форма була представлена реальними clickstream-послідовностями дій користувача в межах сесії, що використовувалися в LSTM-компоненті для моделювання часової структури поведінки. Крім того, на основі результатів кластеризації, індексу CSAI та моделі MONIC формували додатковий набір сегментно-динамічних ознак, який розширював вхідний простір прогностичної моделі та дозволяв враховувати не лише індивідуальні характеристики сесії, а й властивості поведінкового сегмента з погляду його стійкості та еволюції.

Для побудови поведінкових сегментів використано алгоритм  $k$ -means, який добре поєднується з числовою природою ознак сесій та стандартизованим масштабом даних. Алгоритм шукає таке розбиття множини сесій  $X = \{x_1, x_2, \dots, x_n\}$  на  $k$  кластерів  $S_1, S_2, \dots, S_k$ , яке мінімізує сумарну квадратичну відстань до центрів:

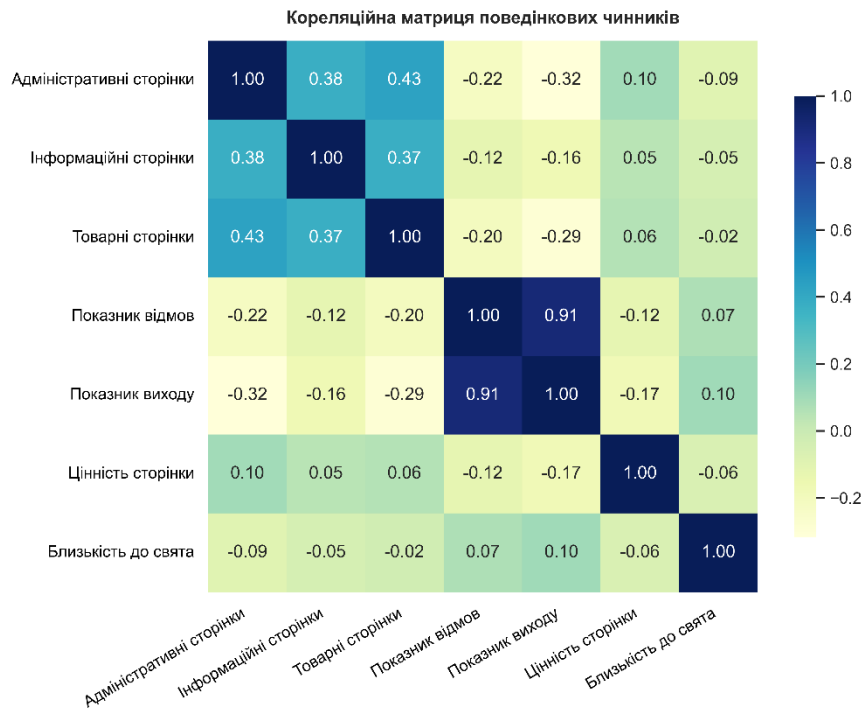
$$J(S_1, \dots, S_k) = \sum_{j=1}^k \sum_{x_i \in S_j} \|x_i - \mu_j\|^2, \quad (1)$$

де  $\mu_j$  інтерпретується як центр кластера  $S_j$ . Центроїди обчислюються як середні вектори об'єктів у кластері, відстань між сесіями вимірюється за евклідовою метрикою, а початкові центроїди ініціалізовано методом  $k$ -means++, що зменшує ризик потрапляння у локальні мінімуми.

Кількість кластерів  $k$  визначено на основі методу «лікоть» та силуетного індексу: послідовне навчання моделей з  $k = 2, \dots, 8$  показало згасання приросту поясненої дисперсії після  $k = 4$ , тоді як силуетний показник досягає локального максимуму саме для чотирикластерного розбиття. Додаткову перевірку якості та стійкості отриманих сегментів виконано за допомогою індексу CSAI, який надалі використовується як основний показник надійності кластерної структури.

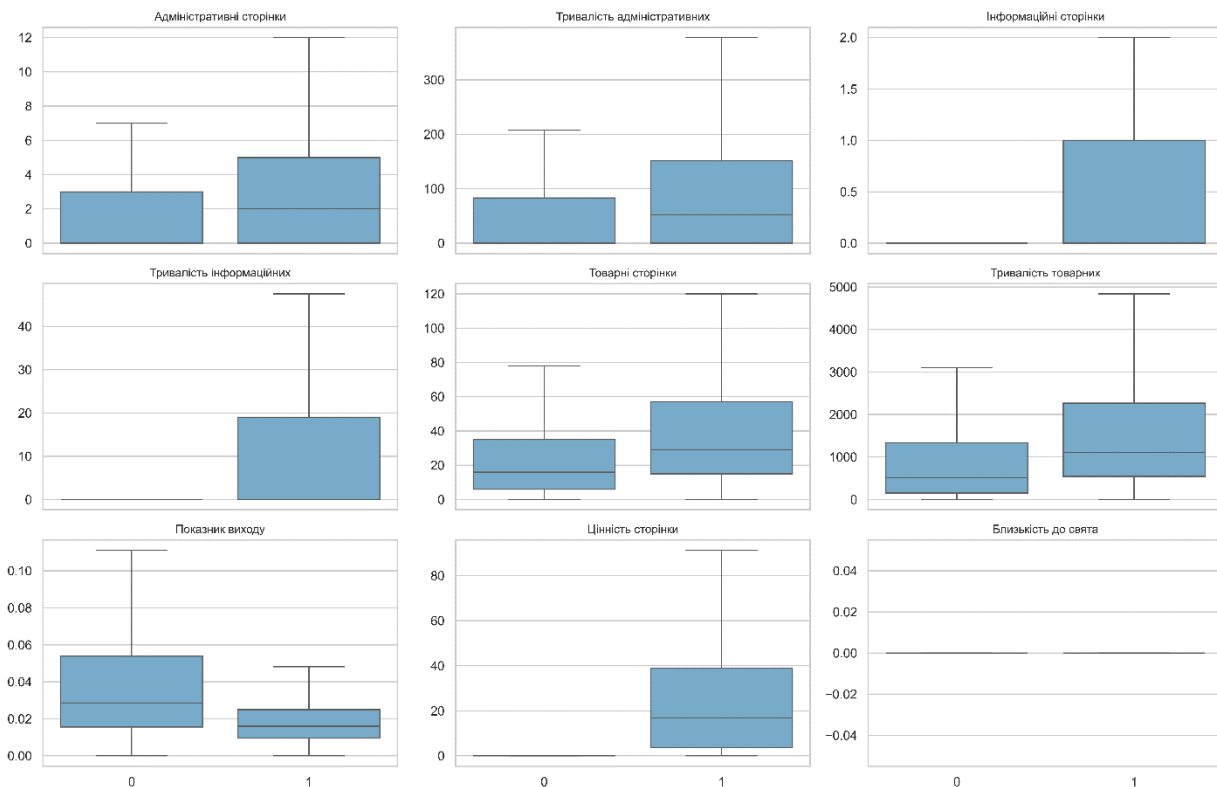
Кореляційний аналіз (Рис. 1) виявив критично високу залежність між показниками відмов та показниками виходу ( $r = 0.91$ ), що свідчить про їх спільну природу як індикаторів проблем у користувацькому досвіді. З огляду на це показник відмов у подальшому аналізі було вилучено, а показник виходу використано як узагальнений індикатор раннього залишення сайту.

Порівняння розподілів поведінкових ознак для сесій з покупкою та без неї (Рис. 2) засвідчує, що конверсійні сесії мають вищі значення більшості кількісних показників взаємодії з сайтом. Водночас найбільш виражений зсув спостерігається для показника цінності сторінки, що підтверджує його ключову роль як індикатора ймовірності транзакції та узгоджується з результатами попередніх досліджень.



**Рис. 1. Кореляційна матриця поведінкових чинників**

**Числові поведінкові ознаки залежно від конверсії**



**Рис. 2. Числові поведінкові ознаки залежно від конверсії**

Для підтвердження виявлених ядер активності в систему впроваджено індекс CSAI (Clustering Stability Assessment Index), який оцінює стабільність ознак через механізм предискретизації даних. Даний індекс розраховується за формулою:

$$CSAI(X_i, A) = \frac{1}{m} \sum_{i=1}^m \left[ \frac{1}{w} \cdot \sum_{j=1}^n \left( \frac{1}{s} \sum_{k=1}^s (\hat{x}_{ijk} - \hat{v}_{ijk})^2 \right) \right], \quad (2)$$

де  $m$  – кількість партій,  $n$  – кількість сегментів, а  $(\hat{x} - \hat{v})^2$  – середньоквадратична відстань між

структурами ознак тренувальних та тестових даних.

Спираючись на підходи до інтерпретації середнього Жаккарівського індексу для аналізу стабільності кластерів, запропоновані зокрема у роботі Хеннінга [20], де високі значення подібності між розбиттями трактується як ознака надійності кластерної структури, а також з урахуванням рекомендацій авторів індексу CSAI, у даній роботі сегменти з  $CSAI \geq 0,85$  розглядаються як стійкі, тоді як нижчі значення інтерпретуються як прояв підвищеної чутливості до зміни вибірки.

Стабільність поведінкових сегментів за індексом CSAI

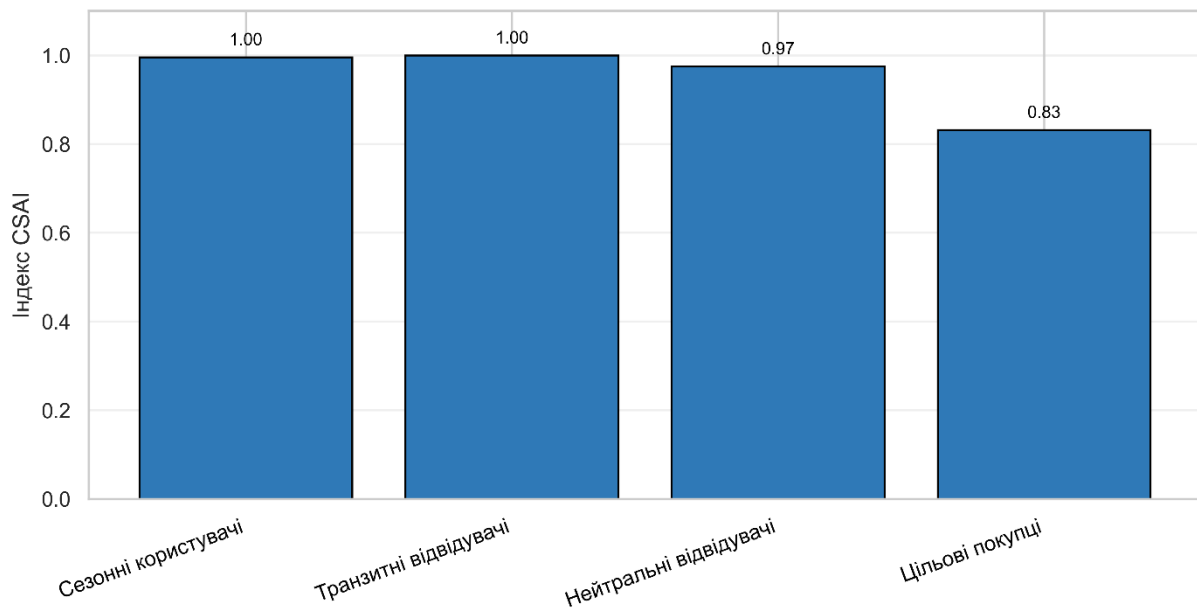


Рис. 3. Стабільність поведінкових сегментів за індексом CSAI

Аналіз розрахованих значень індексу CSAI (рис. 3) показав, що сегменти «Сезонні користувачі», «Транзитні відвідувачі» та «Нейтральні відвідувачі» характеризуються високою структурною стійкістю (CSAI від 0.97 до 1.00) і формують стійке ядро системи. Натомість сегмент «Цільові покупці» має дещо нижче значення  $CSAI = 0.83$ , яке лише наближається до прийнятого порога 0.85, що свідчить про підвищену чутливість цього сегмента до змін вибірки та можливих управлінських втручань. Саме тому подальший аналіз сценаріїв зосереджено на моделюванні міграцій користувачів між стійким ядром і цільовим сегментом. Прескриптивний компонент системи реалізує сценарії типу «якщо-то» за допомогою керованих змін поведінкових ознак, після чого динаміка еволюційних переходів між базовим станом  $t_0$  та стимульованим станом  $t_1$  оцінюється за допомогою моделі MONIC (MONItoring Clusters), яка відстежує еволюцію кластерів у часі на основі міри їх перекриття.

$$overlap(X, Y) = \frac{|objects(X) \cap objects(Y)|}{|objects(X)|} \geq \tau, \quad (3)$$

де  $\tau$  – поріг виживання сегмента, який встановлений на рівні 0.7, оскільки саме це значення узгоджується з рекомендаціями щодо стійкості ієрархічних та розділових структур у складних наборах даних.

У рамках даної роботи MONIC застосовується для порівняння двох станів системи. Базовий стан відповідає кластеризації сесій на вихідних стандартизованих даних, тоді як стимульований стан – кластеризації після модельного управлінського втручання, у межах якого поведінкові ознаки частини користувачів наближуються до профілю цільового сегмента. Для кожної сесії фіксуються номери кластерів у базовому та стимульованому станах, і на цій основі обчислюється міра перекриття між сегментами.

Події збереження сегмента, поглинання, розщеплення сегмента та об'єднання сегментів інтерпретуються відповідно як збереження сегмента за умови достатнього перекриття, приєднання значної частини менш цінних сегментів до стійкого ядра, поділ одного сегмента на декілька з відмінними поведінковими профілями та злиття кількох подібних сегментів в один. У наведеному прикладі базовий та стимульований стани є синтетично змодельованими й відображають прогнозовану реакцію сегментів на сценарій оптимізації UX, а не результати спостережень у реальних часових зрізах.

Сценарій втручання не передбачає прямого керування окремими показниками (наприклад, показником виходу чи цінністю сторінки), оскільки ці величини є результатом комплексних змін у сервісі. У моделі вони коригуються як узагальнені проксі-метрики ефекту від управлінських дій (спрощення навігації, оптимізація вмісту сторінок, зміни в промоакціях тощо), що в реальній системі мають зменшувати ймовірність раннього виходу та підвищувати корисність перегляду. У межах даної роботи розглядається один агрегований сценарій такого типу, в якому припускається, що реалізований комплекс заходів приводить до покращення поведінкових характеристик для частини користувачів, тоді як інші параметри сесій залишаються незмінними.

Міграція поведінкових сегментів за сценарієм оптимізації UX

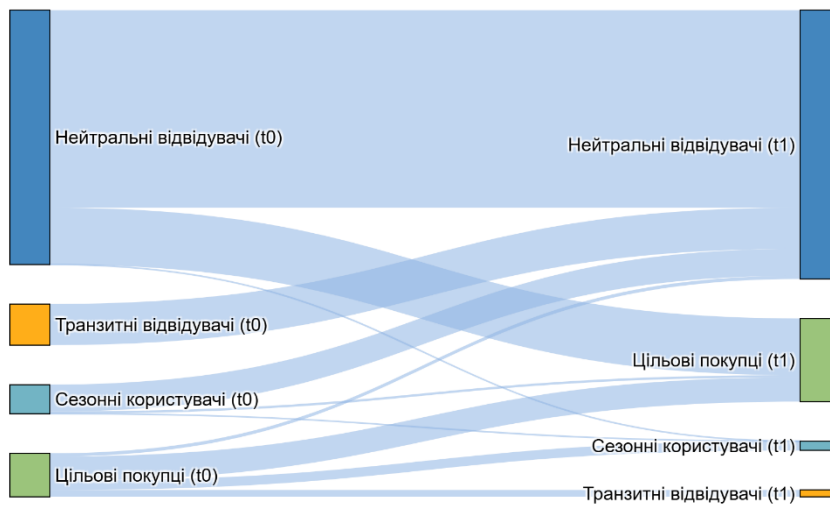


Рис. 4. Модельний приклад міграції поведінкових сегментів за сценарієм оптимізації UX

Результати візуалізації прогностичних потоків (Рис. 4) демонструють зміну розподілу користувачів між поведінковими сегментами під впливом сценарію оптимізації UX. Частина «Транзитних відвідувачів» та «Нейтральних відвідувачів» мігрує до сегментів «Цільові покупці» та «Сезонні користувачі», що відображає перехід від менш цінних або невизначених моделей поведінки до більш цільових сценаріїв взаємодії з сервісом. Це забезпечує прогнозоване посилення сегментів із вищим потенціалом конверсії та підтверджує, що кероване наближення поведінкових профілів до цільового сегмента може бути використане як інструмент підтримки управлінських рішень.

Додатковий аналіз структури сегментів у базовому та стимульованому станах показав, що частка аудиторії, яка належить до стійкого ядра (сегменти з високим CSAI), зберігається на високому рівні, тоді як у цільовому сегменті спостерігається приріст частки сесій. Така динаміка узгоджується з візуалізацією міграцій на Рис. 4 і свідчить про те, що запропонований сценарій сприяє перерозподілу користувачів з нейтральних і транзитних груп до більш цінних сегментів із вищою ймовірністю транзакції.

Структура поведінкових ядер у базовому та стимульованому станах

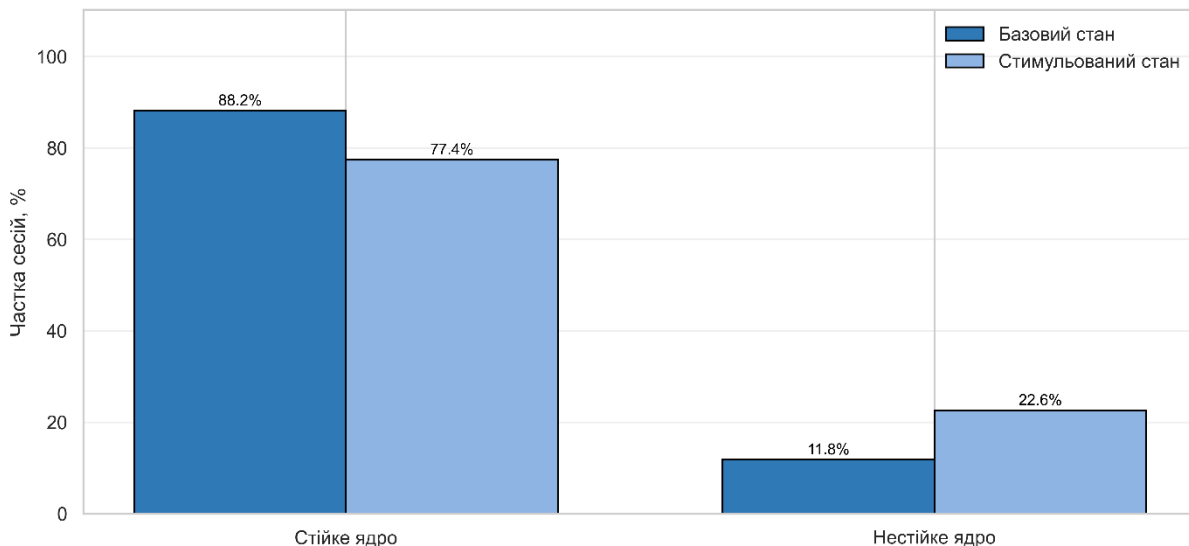


Рис. 5. Зміна частки стійкого ядра та сегменту «Цільові покупці» між базовим і стимульованим станами

Архітектура прогностичного модуля (Рис. 6) побудована на основі гібридного поєднання багатощарового перцептрона (MLP) та мережі довготривалої короткочасної пам'яті (LSTM), що дає змогу одночасно опрацьовувати агреговані характеристики користувацької сесії, сегментно-динамічні параметри та реальні послідовності дій користувача в цифровому середовищі. Така архітектура орієнтована на поєднання табличного та послідовнісного подання даних у межах єдиного прогностичного контуру.

Для цього вхідні дані формувалися у двох взаємодоповнювальних каналах подання. Перший канал являв собою розширений стандартизований вектор ознак сесії, який включав агреговані session-level характеристики взаємодії користувача з сайтом, зокрема тривалість перегляду адміністративних сторінок, тривалість перегляду

інформаційних сторінок, тривалість перегляду товарних сторінок, показник виходів, цінність сторінки та наближеність до особливого дня. Крім того, до цього вектора включалися сегментно-динамічні характеристики, сформовані на основі результатів кластеризації, індексу CSAI та моделі MONIC. Вони відображали належність сесії до певного поведінкового сегмента, рівень його стійкості, а також особливості еволюційної динаміки сегмента у часових зрізах цифрового середовища.

Другий канал подання був представлений реальним clickstream користувача, тобто впорядкованою послідовністю дій у межах сесії. Таке подання відображало часову структуру взаємодії користувача з цифровим середовищем і дозволяло враховувати послідовність переходів між сторінками, типи дій, інтенсивність взаємодії та інші поведінкові характеристики, доступні в межах журналу сесії. На відміну від реконструйованого блокового подання, використання реального clickstream забезпечує більш коректне застосування LSTM-компонента для аналізу часових залежностей у поведінці користувача.

У запропонованій архітектурі MLP-компонент виконує функцію аналізу інтегрального профілю сесії та сегментно-динамічних властивостей поведінкового середовища, тоді як LSTM-компонент моделює часову логіку взаємодії користувача в межах конкретної сесії. Поєднання цих двох модулів дозволяє одночасно враховувати індивідуальні параметри сесії, поведінкову послідовність дій користувача та характеристики сегмента, до якого належить сесія, з поглядом його стійкості та потенційних структурних змін.

Інтеграція обох модулів у межах єдиного прогностичного контуру забезпечує комплексне оцінювання ймовірності конверсії. Якщо сегментно-динамічні характеристики, отримані на основі CSAI та MONIC, відображають структурний стан поведінкових сегментів і тенденції їх змін, то clickstream-компонент дає змогу врахувати безпосередню часову траєкторію взаємодії користувача з цифровим сервісом. У результаті модель формує прогноз конверсії з урахуванням як макрорівня поведінкової структури аудиторії, так і мікрорівня окремої сесії.

MLP-гілка отримувала на вході стандартизований розширений вектор ознак сесії та складалася з двох повнозв'язних шарів на 64 і 32 нейрони відповідно з функцією активації ReLU. LSTM-гілка отримувала на вході реальну clickstream-послідовність дій користувача, для обробки якої використовувався один LSTM-шар на 32 приховані одиниці. Виходи MLP та LSTM-гілок об'єднувалися на етапі конкатенації, після чого передавалися до фінального повнозв'язного шару з sigmoid-активацією, який формував оцінку ймовірності конверсії користувачької сесії.

### Hybrid MLP + LSTM Architecture for Conversion Forecasting

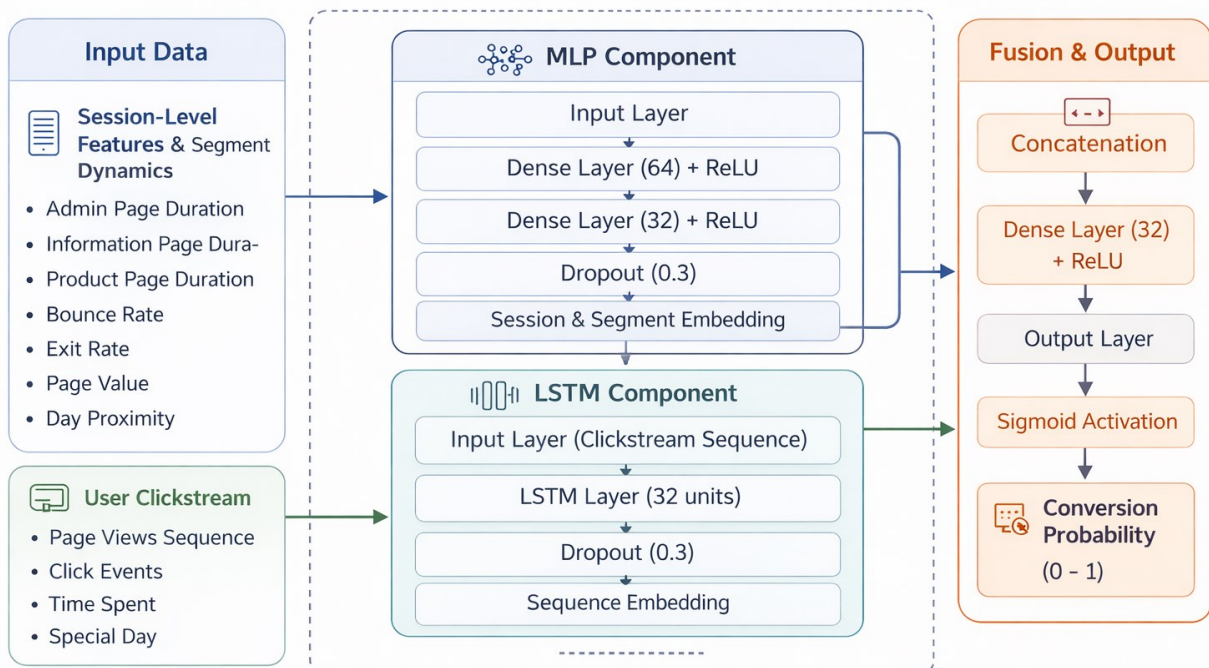


Рис. 6. Архітектура запропонованої моделі нейромережі MLP+LSTM

Навчання моделі прогнозування конверсії в цифровому середовищі на основі архітектури MLP+LSTM з інтеграцією показників стійкості та еволюції поведінкових сегментів виконувалося на тренувальній вибірці з використанням алгоритму Adam з початковою швидкістю навчання 0.001. Кількість епох становила 50, розмір пакета — 64. Для запобігання перенавчання застосовувався механізм ранньої зупинки (early stopping) з контролем значення функції втрат на валідаційній вибірці та параметром patience = 5. У разі відсутності покращення впродовж п'яти послідовних епох навчання припинялося, а ваги найкращої конфігурації моделі автоматично відновлювалися. Для забезпечення коректності оцінювання дані було поділено на тренувальну, валідаційну та тестову вибірки у співвідношенні 70:15:15. Тренувальна вибірка використовувалася для оновлення

ваг моделі, валідаційна — для контролю процесу навчання та вибору найкращої конфігурації, а тестова — виключно для підсумкової незалежної оцінки узагальнювальної здатності моделі.

Як функцію втрат використовували бінарну крос-ентропію, що відповідає бінарній природі задачі прогнозування конверсії. Якість навчання оцінювали на валідаційній вибірці за метрикою ROC-AUC, а фінальну модель обирали як конфігурацію з найкращим значенням ROC-AUC на валідаційних даних. Після завершення навчання її ефективність перевіряли на тестовій вибірці за показниками Accuracy, Precision, Recall, F1-score та ROC-AUC.

Обрана конфігурація архітектури є компромісом між виразністю моделі та ризиком перенавчання в умовах одночасного використання агрегованих session-level ознак, сегментно-динамічних характеристик і реальних clickstream-послідовностей. Зокрема, застосування двох прихованих шарів у MLP-компоненті забезпечує виявлення нелінійних взаємозв'язків між агрегованими ознаками сесії та параметрами, сформованими на основі CSAI і MONIC, тоді як один LSTM-шар дозволяє моделювати часові залежності в реальній послідовності дій користувача в межах сесії без надмірного ускладнення архітектури.

Для перевірки ефективності запропонованого нейромережевого підходу результати моделі MLP+LSTM з інтеграцією показників стійкості та еволюції поведінкових сегментів було порівняно з кількома базовими моделями (Табл. 1), що широко застосовуються для прогнозування конверсії користувачів у цифровому середовищі.

Таблиця 1

Порівняння базових моделей та запропонованої моделі

Модель	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.84	0.79	0.72	0.75	0.88
Decision Tree	0.83	0.76	0.70	0.73	0.85
Random Forest	0.90	0.84	0.80	0.82	0.93
MLP+LSTM	0.90	0.85	0.82	0.83	0.94
<b>MLP+LSTM (CSAI+MONIC)</b>	<b>0.94</b>	<b>0.86</b>	<b>0.84</b>	<b>0.85</b>	<b>0.95</b>

У межах запропонованого підходу показники CSAI та MONIC інтегруються в прогностичну модель як додаткові сегментно-динамічні ознаки, що розширюють простір вхідних параметрів. Індекс CSAI використовується для кількісного оцінювання рівня стійкості поведінкового сегмента, до якого належить сесія, тоді як модель MONIC дозволяє визначати характеристики еволюційної динаміки сегментів, зокрема ознаки збереження, переходів, злиття або розщеплення кластерів у часових зрізах цифрового середовища. Отримані параметри включаються до розширеного вектора ознак, який подається на вхід MLP-компонента разом з агрегованими session-level характеристиками. Таким чином, конфігурація MLP+LSTM (CSAI+MONIC) у Табл. 1 відрізняється від базового варіанта MLP+LSTM не лише використанням агрегованих ознак сесії та реального clickstream, а й включенням до моделі параметрів, що відображають стійкість і трансформаційну динаміку поведінкових сегментів.

Для більш детального аналізу структури помилок моделі було побудовано матрицю заплутаності (Рис. 7), яка дозволяє оцінити співвідношення правильних і хибних класифікацій та виявити особливості розпізнавання конверсійних і неконверсійних сесій.

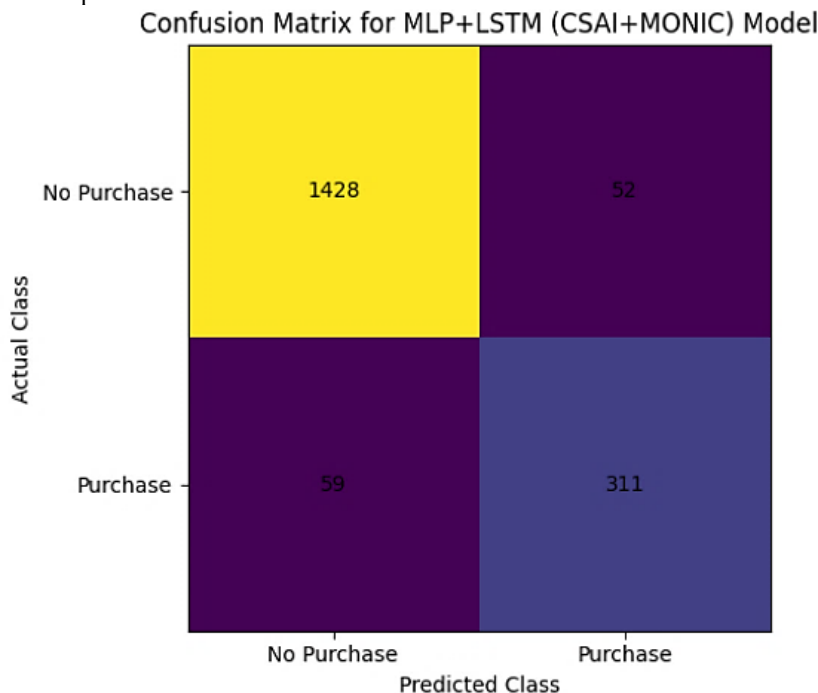


Рис. 7. Матриця заплутаності для моделі MLP+LSTM (CSAI+MONIC)

Аналіз матриці заплутаності показує, що запропонована модель успішно ідентифікує більшість конверсійних сесій. Зокрема, 311 випадків покупки було правильно передбачено моделлю, тоді як лише 59 конверсійних сесій було класифіковано як неконверсійні. Частка хибно-позитивних прогнозів залишається відносно низькою (52 випадків), що свідчить про достатньо збалансовану роботу моделі між точністю та повнотою класифікації.

Запропонований підхід до прогнозування конверсії має низку обмежень, зумовлених особливостями формування ознакового простору, сегментації поведінкових сесій та інтеграції сегментно-динамічних характеристик у прогностичну модель. По-перше, значення показників CSAI та MONIC залежать від якості кластеризації, вибору простору поведінкових ознак і прийнятої кількості кластерів, тому отримані оцінки стійкості та еволюції сегментів можуть змінюватися за альтернативних конфігурацій кластерного аналізу. По-друге, аналіз еволюції сегментів за моделлю MONIC суттєво залежить від способу формування часових зрізів цифрового середовища, а також від повноти й репрезентативності даних, що використовуються для відстеження структурних змін поведінкових груп. По-третє, ефективність LSTM-компонента визначається якістю подання реального clickstream, зокрема повнотою журналів дій користувача, способом кодування подій, довжиною послідовностей і процедурою їх уніфікації. Крім того, результати моделі можуть залежати від способу масштабування ознак, параметрів навчання та обраної архітектурної конфігурації, що потребує додаткової перевірки стійкості висновків за альтернативних налаштувань. У зв'язку з цим подальша емпірична верифікація підходу доцільна на різних типах цифрових сервісів, з використанням інших конфігурацій сегментації, часових зрізів і способів подання clickstream-даних.

### Подяка

Дослідження було проведено за грантової підтримки Національного фонду досліджень України "Методи аналізу та оптимізації мультимодальних даних для моделей глибокого навчання у військовій сфері", project registration No 2025.07/0017 від 24/12/2025.

### Висновки з даного дослідження

#### і перспективи подальших розвідок у даному напрямі

У статті розроблено модель прогнозування конверсії в цифровому середовищі на основі архітектури MLP+LSTM з інтеграцією показників стійкості та еволюції поведінкових сегментів. Запропонований підхід поєднує кластеризацію користувацьких сесій, оцінювання стійкості поведінкових сегментів за індексом CSAI, аналіз їхньої еволюційної динаміки за моделлю MONIC та прогностичне моделювання ймовірності конверсії в межах єдиного аналітичного контуру. Це дозволило розширити традиційні підходи до прогнозування цільових дій користувачів за рахунок урахування не лише поточних характеристик окремої сесії, а й властивостей сегмента, до якого вона належить, з погляду його стійкості та структурних змін.

У ході дослідження обґрунтовано доцільність використання кластерного аналізу для виявлення неоднорідних поведінкових патернів користувачів у цифровому середовищі. На основі агрегованих поведінкових ознак сесій було сформовано сегменти, що відображають різні типи взаємодії користувачів із цифровим сервісом. Подальше застосування індексу CSAI дало змогу кількісно оцінити рівень стійкості виявлених сегментів, а використання моделі MONIC — визначити характеристики їхньої еволюційної динаміки в часових зрізах цифрового середовища. Це підтвердило, що поведінкові сегменти доцільно розглядати не як статичні класи, а як динамічні поведінкові підсистеми, здатні змінюватися під впливом умов взаємодії користувача з цифровим середовищем.

Окремим результатом дослідження стало розроблення прогностичного модуля на основі гібридної архітектури MLP+LSTM. У запропонованій моделі MLP-компонент опрацьовував агреговані session-level ознаки та сегментно-динамічні характеристики, сформовані на основі результатів кластеризації, індексу CSAI і моделі MONIC, тоді як LSTM-компонент аналізував реальні clickstream-послідовності дій користувача в межах сесії. Поєднання цих двох каналів подання даних дало змогу врахувати одночасно інтегральний профіль сесії, часову структуру поведінки користувача та властивості сегмента, до якого належить сесія. Порівняння з базовими моделями показало, що запропонований гібридний підхід забезпечує вищу якість прогнозування конверсії, а включення показників стійкості та еволюції поведінкових сегментів підвищує інформативність прогностичної моделі.

Практичне значення отриманих результатів полягає в тому, що запропонована модель дозволяє виявляти сегменти користувачів із різним рівнем поведінкової стійкості та конверсійного потенціалу, аналізувати їхню еволюційну динаміку та використовувати ці характеристики для підвищення точності прогнозування цільових дій у цифрових сервісах. Такий підхід розширює можливості персоналізації, оптимізації інтерфейсів, адаптації цифрових сценаріїв взаємодії та підтримки рішень щодо управління поведінкою аудиторії. Таким чином, показники стійкості та еволюції сегментів використовуються не лише як засіб аналітичної інтерпретації кластерної структури, а і як безпосередні інформаційні компоненти моделі прогнозування конверсії.

Перспективи подальших досліджень пов'язані насамперед із розширенням емпіричної бази та подальшою верифікацією запропонованого підходу на різних типах цифрових середовищ. Доцільним є проведення порівняльних експериментів для різних конфігурацій кластеризації, способів формування сегментно-динамічних ознак, параметрів часових зрізів для аналізу еволюції сегментів, а також стратегій подання та кодування реальних clickstream-послідовностей. Перспективним є також включення до моделі додаткових

контекстних характеристик, зокрема джерела трафіку, типу пристрою, часових параметрів активності та маркетингових факторів, що може підвищити точність оцінювання стійкості поведінкових сегментів і прогнозування конверсії.

Подальший розвиток роботи доцільно пов'язати з удосконаленням механізмів формування та інтеграції сегментно-динамічних характеристик у прогностичний контур, а також із перевіркою запропонованої моделі на інших типах цифрових сервісів, зокрема в освітніх платформах, фінансових системах, медичних інформаційних середовищах та багатоканальних клієнтських екосистемах.

### Література

1. Sakar C. O., Polat S. O., Katircioglu M., Kocakoglu M. Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks // *Neural Computing and Applications*. 2019. Vol. 31, No. 10. P. 6893–6908. DOI: 10.1007/s00521-018-3523-0.
2. Wang G., Zhang X., Tang S., Gong S. Unsupervised Clickstream Clustering for User Behavior Analysis // *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016. P. 225–236. DOI: 10.1145/2858036.2858107.
3. Kwan I. S. Y., Fong J., Wong H. K. An e-customer behavior model with online analytical mining for internet marketing planning // *Decision Support Systems*. 2005. Vol. 41, No. 1. P. 189–204. DOI: 10.1016/j.dss.2004.11.012.
4. Tarekegn A. N., Tessem B., Rabbi F. A New Cluster Validation Index Based on Stability Analysis (CSAI) // *Pattern Recognition Applications and Methods. ICPRAM 2025*. 2025. Vol. 1. P. 377–384. DOI: 10.5220/0013309100003905.
5. Spiliopoulou M., Ntoutsis E., Theodoridis Y., Schult R. MONIC: Modeling and monitoring cluster transitions // *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. 2006. P. 706–711. DOI: 10.1145/1150402.1150491.
6. Lepenioti K., Bousdekis A., Apostolou D., Mentzas G. Prescriptive analytics: Literature review and research challenges // *International Journal of Information Management*. 2020. Vol. 50. P. 57–70. DOI: 10.1016/j.ijinfomgt.2019.04.003.
7. Spiliopoulou M., Ntoutsis E., Theodoridis Y., Schult R. MONIC and Followups on Modeling and Monitoring Cluster Transitions // *Lecture Notes in Computer Science*. 2013. Vol. 8190. P. 622–626. DOI: 10.1007/978-3-642-40994-3\_41.
8. Wu S., Yau W.-C., Ong T.-S., Chong S.-C. Integrated Churn Prediction and Customer Segmentation Framework for Telco Business // *IEEE Access*. 2021. Vol. 9. P. 62118–62136. DOI: 10.1109/ACCESS.2021.3073776.
9. Zhao H.-H., Luo X.-C., Ma R., Lu X. An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation with Correlated Variables // *IEEE Access*. 2021. Vol. 9. P. 48405–48412. DOI: 10.1109/ACCESS.2021.3067499.
10. Wong S. Y., Ong L. Y., Leow M. C. AIDA-Based Customer Segmentation With User Journey Analysis for Wi-Fi Advertising System // *IEEE Access*. 2024. Vol. 12. P. 111468–111480. DOI: 10.1109/ACCESS.2024.3424833.
11. Abbasimehr H., Shabani M. A New Methodology for Customer Behavior Analysis Using Time Series Clustering: A Case Study on a Bank's Customers // *Kybernetes*. 2021. Vol. 50, No. 2. P. 221–242. DOI: 10.1108/K-09-2018-0506.
12. Hamidi H., Haghi B. An Approach Based on Data Mining and Genetic Algorithm to Optimizing Time Series Clustering for Efficient Segmentation of Customer Behavior // *Computers in Human Behavior Reports*. 2024. Vol. 16. Art. 100520. DOI: 10.1016/j.chbr.2024.100520.
13. Morales-Oñate V., Morales-Oñate B. clusEvol: An R Package for Cluster Evolution Analytics // *SoftwareX*. 2024. Vol. 28. Art. 101921. DOI: 10.1016/j.softx.2024.101921.
14. Khan R. H., Dofadar D. F., Alam M. G. R., Siraj M., Hassan M. R., Hassan M. M. LRFS: Online Shoppers' Behavior-Based Efficient Customer Segmentation Model // *IEEE Access*. 2024. Vol. 12. P. 96462–96480. DOI: 10.1109/ACCESS.2024.3420221.
15. Liao J., Jantan A., Ruan Y., Zhou C. Multi-Behavior RFM Model Based on Improved SOM Neural Network Algorithm for Customer Segmentation // *IEEE Access*. 2022. Vol. 10. P. 122501–122512. DOI: 10.1109/ACCESS.2022.3223361.
16. Abbasimehr H., Bahrini A. An Analytical Framework Based on the Recency, Frequency, and Monetary Model and Time Series Clustering Techniques for Dynamic Segmentation // *Expert Systems with Applications*. 2022. Vol. 192. Art. 116373. DOI: 10.1016/j.eswa.2021.116373.
17. Sun Z.-H., Zuo T.-Y., Liang D., Ming X., Chen Z., Qiu S. GPHC: A Heuristic Clustering Method to Customer Segmentation // *Applied Soft Computing*. 2021. Vol. 111. Art. 107677. DOI: 10.1016/j.asoc.2021.107677.
18. Kim S., Shin W., Kim H.-W. Predicting Online Customer Purchase: The Integration of Customer Characteristics and Browsing Patterns // *Decision Support Systems*. 2024. Vol. 177. Art. 114105. DOI: 10.1016/j.dss.2023.114105.
19. Online Shoppers Purchasing Intention Dataset [Electronic resource] // *UCI Machine Learning Repository*. – 2019. – Mode of access: <https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset> (date of access: 03.12.2026).

20. Hennig C. Cluster-wise assessment of cluster stability // *Computational Statistics & Data Analysis*. – 2007. – Vol. 52, No. 1. – P. 258–271. – DOI: 10.1016/j.csda.2006.11.025.

### References

1. Sakar C. O., Polat S. O., Katircioglu M., Kocakoglu M. Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks // *Neural Computing and Applications*. 2019. Vol. 31, No. 10. P. 6893–6908. DOI: 10.1007/s00521-018-3523-0.
2. Wang G., Zhang X., Tang S., Gong S. Unsupervised Clickstream Clustering for User Behavior Analysis // *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016. P. 225–236. DOI: 10.1145/2858036.2858107.
3. Kwan I. S. Y., Fong J., Wong H. K. An e-customer behavior model with online analytical mining for internet marketing planning // *Decision Support Systems*. 2005. Vol. 41, No. 1. P. 189–204. DOI: 10.1016/j.dss.2004.11.012.
4. Tarekegn A. N., Tessem B., Rabbi F. A New Cluster Validation Index Based on Stability Analysis (CSAI) // *Pattern Recognition Applications and Methods. ICPRAM 2025*. 2025. Vol. 1. P. 377–384. DOI: 10.5220/0013309100003905.
5. Spiliopoulou M., Ntoutsis E., Theodoridis Y., Schult R. MONIC: Modeling and monitoring cluster transitions // *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. 2006. P. 706–711. DOI: 10.1145/1150402.1150491.
6. Lepenioti K., Bousdekis A., Apostolou D., Mentzas G. Prescriptive analytics: Literature review and research challenges // *International Journal of Information Management*. 2020. Vol. 50. P. 57–70. DOI: 10.1016/j.ijinfomgt.2019.04.003.
7. Spiliopoulou M., Ntoutsis E., Theodoridis Y., Schult R. MONIC and Followups on Modeling and Monitoring Cluster Transitions // *Lecture Notes in Computer Science*. 2013. Vol. 8190. P. 622–626. DOI: 10.1007/978-3-642-40994-3\_41.
8. Wu S., Yau W.-C., Ong T.-S., Chong S.-C. Integrated Churn Prediction and Customer Segmentation Framework for Telco Business // *IEEE Access*. 2021. Vol. 9. P. 62118–62136. DOI: 10.1109/ACCESS.2021.3073776.
9. Zhao H.-H., Luo X.-C., Ma R., Lu X. An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation with Correlated Variables // *IEEE Access*. 2021. Vol. 9. P. 48405–48412. DOI: 10.1109/ACCESS.2021.3067499.
10. Wong S. Y., Ong L. Y., Leow M. C. AIDA-Based Customer Segmentation With User Journey Analysis for Wi-Fi Advertising System // *IEEE Access*. 2024. Vol. 12. P. 111468–111480. DOI: 10.1109/ACCESS.2024.3424833.
11. Abbasimehr H., Shabani M. A New Methodology for Customer Behavior Analysis Using Time Series Clustering: A Case Study on a Bank's Customers // *Kybernetes*. 2021. Vol. 50, No. 2. P. 221–242. DOI: 10.1108/K-09-2018-0506.
12. Hamidi H., Haghi B. An Approach Based on Data Mining and Genetic Algorithm to Optimizing Time Series Clustering for Efficient Segmentation of Customer Behavior // *Computers in Human Behavior Reports*. 2024. Vol. 16. Art. 100520. DOI: 10.1016/j.chbr.2024.100520.
13. Morales-Oñate V., Morales-Oñate B. clusEvol: An R Package for Cluster Evolution Analytics // *SoftwareX*. 2024. Vol. 28. Art. 101921. DOI: 10.1016/j.softx.2024.101921.
14. Khan R. H., Dofadar D. F., Alam M. G. R., Siraj M., Hassan M. R., Hassan M. M. LRFS: Online Shoppers' Behavior-Based Efficient Customer Segmentation Model // *IEEE Access*. 2024. Vol. 12. P. 96462–96480. DOI: 10.1109/ACCESS.2024.3420221.
15. Liao J., Jantan A., Ruan Y., Zhou C. Multi-Behavior RFM Model Based on Improved SOM Neural Network Algorithm for Customer Segmentation // *IEEE Access*. 2022. Vol. 10. P. 122501–122512. DOI: 10.1109/ACCESS.2022.3223361.
16. Abbasimehr H., Bahrini A. An Analytical Framework Based on the Recency, Frequency, and Monetary Model and Time Series Clustering Techniques for Dynamic Segmentation // *Expert Systems with Applications*. 2022. Vol. 192. Art. 116373. DOI: 10.1016/j.eswa.2021.116373.
17. Sun Z.-H., Zuo T.-Y., Liang D., Ming X., Chen Z., Qiu S. GPHC: A Heuristic Clustering Method to Customer Segmentation // *Applied Soft Computing*. 2021. Vol. 111. Art. 107677. DOI: 10.1016/j.asoc.2021.107677.
18. Kim S., Shin W., Kim H.-W. Predicting Online Customer Purchase: The Integration of Customer Characteristics and Browsing Patterns // *Decision Support Systems*. 2024. Vol. 177. Art. 114105. DOI: 10.1016/j.dss.2023.114105.
19. Online Shoppers Purchasing Intention Dataset [Electronic resource] // *UCI Machine Learning Repository*. – 2019. – Mode of access: <https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset> (date of access: 03.12.2026).
20. Hennig C. Cluster-wise assessment of cluster stability // *Computational Statistics & Data Analysis*. – 2007. – Vol. 52, No. 1. – P. 258–271. – DOI: 10.1016/j.csda.2006.11.025.