

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Системотехніки  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)  
Розробка методів сегментації клієнтів ресторану за поведінковими  
характеристиками з використанням алгоритмів машинного навчання  
(тема)

Виконав:  
здобувач 2 року навчання,  
групи ІТІМ-24-2  
Владисав КОРЗУН  
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки  
(код і повна назва спеціальності)  
Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)  
Освітня програма Інформаційні технології  
проектування  
(повна назва освітньої програми)

Керівник дрц. каф. СТ. Ольга Чорна  
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри \_\_\_\_\_

(підпис)

Ігор ГРЕБЕННІК  
(власне ім'я, прізвище)

2025 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Системотехніки

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки

(код і повна назва)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні технології проектування

(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_

(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

## ЗАВДАННЯ

### НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Корзуну Владиславу Романовичу

(прізвище, ім'я, по батькові)

1. Тема роботи Розробка методів сегментації клієнтів ресторану за поведінковими характеристиками з використанням алгоритмів машинного навчання

затверджена наказом університету від 24 листопада 2025 р. №1058 Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 10 грудня 2025 р.

3. Вихідні дані до роботи: дані поведінки та транзакцій користувачів ресторанної інформаційної системи (історія замовлень, відгуки, частота відвідувань, середній чек), демографічна інформація про клієнтів, анотації з віднесенням до цільових груп, інформація про сучасні методи інтелектуального аналізу даних у ресторанных системах, приклади систем персоналізації та рекомендацій (RestAI, FoodInsight, DineSmart), інструменти для побудови моделей машинного навчання – Python, TensorFlow, scikit-learn, Pandas, обчислювальні ресурси для тренування моделей, метрики точності, повноти, F1-міри

4. Перелік питань, що потрібно опрацювати в роботі: аналіз ресторанного бізнесу та поведінки клієнтів, методи сегментації клієнтів на основі транзакцій та поведінкових даних, методи кластеризації для визначення цільових груп, методи класифікації для прогнозування уподобань клієнтів, приклади сучасних IT-рішень у ресторанному бізнесі (RestAI, FoodInsight, DineSmart), порівняння систем персоналізації, застосування кластерного аналізу для формування груп користувачів, оптимізація та оцінка моделей

класифікації, розробка гібридного методу сегментації, вибір і опис вхідних даних, інструменти реалізації моделі, експериментальне дослідження гібридної моделі  
 5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій: схема сегментації клієнтів ресторанної системи на основі поведінкових даних, порівняння характеристик клієнтів у різних сегментах, архітектура DNN для прогнозування уподобань клієнтів, архітектура класифікаційної моделі для поведінкових даних, алгоритм побудови гібридного методу сегментації, метрики моделі DNN, метрики класифікаційної моделі, залежність точності DNN від кількості епох, залежність точності класифікаційної моделі від кількості епох, залежність точності гібридної моделі від кількості епох

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / термін виконання етапів роботи	Примітка
1.	Отримання завдання на виконання кваліфікаційної роботи	01.09.2025	Виконано
2.	Аналіз завдання та предметної області	02.09.2025-15.09.2025	Виконано
3.	Опрацювання літератури та аналіз об'єкта дослідження	16.09.2025-22.09.2025	Виконано
4.	Розроблення моделі сегментації користувачів ресторанної системи	23.09.2025-03.10.2025	Виконано
5.	Розробка та тестування моделі кластеризації для групування користувачів	04.10.2025-20.10.2025	Виконано
6.	Розробка та тестування моделі класифікації для прогнозування уподобань і поведінки	21.10.2025-30.10.2025	Виконано
7.	Інтеграція моделей у гібридний підхід та його навчання	31.10.2025-15.11.2025	Виконано
8.	Проведення експериментального дослідження гібридної моделі	16.11.2025-25.11.2025	Виконано
9.	Аналіз результатів, формулювання висновків та	26.11.2025-01.12.2025	Виконано
10.	Оформлення пояснювальної записки та презентаційних	02.12.2025-05.12.2025	Виконано
11.	Представлення роботи на рецензування	09.12.2025	Виконано

Дата видачі завдання 01 вересня 2025 р.

Здобувач \_\_\_\_\_  
 (підпис)

Керівник роботи \_\_\_\_\_ доц. каф. СТ Ольга ЧОРНА  
 (підпис) (посада, власне ім'я, прізвище)

## РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 76 с., 16 рис., 0 табл., 2 дод., 29 джерел.

### ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, СИСТЕМИ РЕСТОРАННОГО БІЗНЕСУ, СЕГМЕНТАЦІЯ КЛІЄНТІВ, ГІБРИДНІ МЕТОДИ, КЛАСТЕРИЗАЦІЯ, КЛАСИФІКАЦІЯ, МОДЕЛЮВАННЯ ПОВЕДІНКИ

Об'єктом дослідження є процес інтелектуальної обробки даних користувачів ресторанної інформаційної системи з метою виявлення закономірностей у їх поведінці та підвищення ефективності взаємодії з клієнтами.

Предмет дослідження охоплює методи кластеризації, класифікації, регресійного аналізу та алгоритми виявлення асоціативних правил, адаптовані для аналізу даних ресторанных користувачів, а також способи інтеграції цих методів у повноцінну систему сегментації.

Метою роботи є підвищення ефективності ресторанных інформаційних систем шляхом створення гібридного підходу до сегментації клієнтів, який забезпечує точнішу ідентифікацію цільових груп та дозволяє оптимізувати процеси персоналізації сервісів і загальної взаємодії з відвідувачами.

Методи дослідження включають структурний і системний аналіз сучасних рішень у сфері автоматизації ресторанного бізнесу, моделювання процесів роботи з клієнтськими даними, застосування алгоритмів машинного та глибинного навчання для формування сегментів.

Наукова новизна роботи полягає у формуванні комплексного підходу до сегментації користувачів ресторанної системи, який базується на поєднанні кластеризації, методів класифікаційного аналізу та алгоритмів асоціативних правил, що дозволяє розширити можливості традиційних моделей і забезпечити глибше розуміння взаємопов'язаних характеристик поведінки клієнтів.

## ABSTRACT

Explanatory Note to the Qualification Thesis: 76 pages, 16 figures, 0 tables, 2 appendices, 29 references.

INTELLIGENT DATA ANALYSIS, RESTAURANT BUSINESS SYSTEMS, CUSTOMER SEGMENTATION, HYBRID METHODS, CLUSTERING, CLASSIFICATION, BEHAVIOR MODELING

The object of study is the process of intelligent data processing of users in a restaurant information system with the aim of identifying patterns in their behavior and improving the efficiency of customer interaction.

The subject of study covers clustering, classification, regression analysis methods, and algorithms for discovering association rules, adapted for analyzing restaurant user data, as well as approaches for integrating these methods into a comprehensive segmentation system.

The purpose of the work is to enhance the effectiveness of restaurant information systems by creating a hybrid approach to customer segmentation, which provides more accurate identification of target groups and enables optimization of service personalization and overall interaction with visitors.

Research methods include structural and system analysis of modern solutions in restaurant business automation, modeling of processes for working with customer data, and application of machine learning and deep learning algorithms for segment formation.

The scientific novelty of the work lies in the development of a comprehensive approach to user segmentation in restaurant systems, based on combining clustering, classification analysis methods, and association rule algorithms, which expands the capabilities of traditional models and allows a deeper understanding of interrelated characteristics of customer behavior.

## ЗМІСТ

ВСТУП.....	16
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ .....	18
1.1 Опис сучасного стану розвитку інформаційних систем у ресторанному бізнесі.....	18
1.2 Аналіз застосування методів інтелектуального аналізу даних у ресторанних системах .....	20
1.3 Актуальність використання досліджуваних методів у ресторанному бізнесі.....	22
1.4 Постановка задачі дослідження .....	23
2 ОГЛЯД МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ .....	25
2.1 Методи кластерного аналізу.....	25
2.2 Методи класифікації .....	27
2.3 Методи асоціативних правил .....	29
2.4 Методи регресійного аналізу .....	30
2.5 Узагальнення методів для вирішення задач сегментації цільових груп..	33
2.6 Опис розробленого гібридного методу сегментації користувачів .....	35
3 ЛОГІЧНЕ ТА ФІЗИЧНЕ ПРОЄКТУВАННЯ БАЗИ ДАНИХ.....	37
3.1 Логічне проєктування бази даних користувачів ресторанної системи....	37
3.2 Фізичне проєктування бази даних користувачів ресторанної системи ...	39
3.3 Використання даних користувачів для їх сегментації.....	42
4 ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ СЕГМЕНТАЦІЇ КОРИСТУВАЧІВ .....	45
4.1 Опис середовища розробки .....	45
4.2 Опис використаних бібліотек.....	46
4.3 Попередня обробка та нормалізація даних .....	48
4.4 Проведення експерименту .....	50
4.5 Аналіз отриманих результатів.....	62
ВИСНОВКИ.....	65



## ВСТУП

У сучасних умовах цифровізації ресторанний бізнес активно інтегрує інформаційні технології для підвищення ефективності управління та якості обслуговування клієнтів. Одним із найперспективніших напрямів розвитку є впровадження рекомендаційних систем, які дозволяють персоналізувати пропозиції для кожного відвідувача. Водночас ефективність таких систем значною мірою залежить від правильного вибору методів кластеризації користувачів, що забезпечує виявлення схожих груп клієнтів за їхніми вподобаннями, поведінковими характеристиками та історією замовлень [1]. Це, у свою чергу, сприяє оптимізації маркетингових стратегій, підвищенню рівня лояльності та прибутковості ресторану.

Актуальність дослідження зумовлена зростаючою потребою у впровадженні інтелектуальних моделей аналізу користувацьких даних, здатних адаптуватися до динамічних змін у споживчій поведінці, що є особливо важливим у конкурентному середовищі ресторанного бізнесу.

Предметом дослідження є методи та моделі кластеризації користувачів у рекомендаційній системі ресторану для формування персоналізованих пропозицій і підвищення ефективності взаємодії з клієнтами.

Об'єктом дослідження є процес формування та функціонування рекомендаційної системи ресторану, побудованої на основі аналізу поведінкових і транзакційних даних користувачів.

Метою дослідження є розроблення та аналіз ефективних методів кластеризації користувачів для побудови рекомендаційної системи ресторану, що забезпечує персоналізацію сервісу, оптимізацію маркетингових рішень і підвищення рівня задоволеності клієнтів.

Наукова новизна полягає у комплексному підході до використання методів кластеризації для сегментації користувачів ресторанної рекомендаційної системи, який поєднує поведінкові, контекстні та транзакційні характеристики клієнтів. У роботі пропонується вдосконалена

модель кластеризації, що дозволяє підвищити точність формування рекомендацій за рахунок адаптації параметрів кластерів до поточних тенденцій споживання. Отримані результати можуть бути використані для створення інтелектуальних інформаційних систем у сфері ресторанного бізнесу, орієнтованих на підвищення ефективності прийняття управлінських рішень.

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

## 1.1 Опис сучасного стану розвитку інформаційних систем у ресторанному бізнесі

Сьогодні інформаційна економіка переживає період інтенсивного розвитку, у якому дані та цифрові технології відіграють роль стратегічного ресурсу та ключового чинника конкурентоспроможності бізнесу. Однією з галузей, що найбільш активно реагує на ці процеси, є ресторанний бізнес, де формуються нові підходи до управління клієнтськими взаємовідносинами, оптимізації процесів обслуговування та формування персоналізованих пропозицій. Цифрова трансформація ресторанної індустрії призводить до зростання обсягів користувацьких даних, що відкриває можливості для побудови інтелектуальних аналітичних систем. Одним із ключових напрямів такого розвитку є впровадження рекомендаційних систем, заснованих на методах кластеризації користувачів [2].

У сучасних умовах зростаючої конкуренції між закладами громадського харчування успішність ресторану значною мірою залежить від здатності утримувати постійних клієнтів і приваблювати нових. Це можливо лише за умови глибокого аналізу споживчих переваг, поведінкових патернів і динаміки попиту. Традиційні підходи маркетингу, що базуються на досвіді персоналу або ручному аналізі статистики продажів, поступово втрачають ефективність у зв'язку зі збільшенням обсягів даних та швидкістю зміни тенденцій споживання. Тому актуальним є використання методів машинного навчання, зокрема кластеризації, які дозволяють автоматизовано виокремлювати групи клієнтів зі схожими характеристиками та вподобаннями.

З програмного погляду, сучасна інформаційна система ресторану являє собою багаторівневу структуру, що забезпечує збір, зберігання та аналітичну обробку даних про клієнтів [3]. Вона включає модулі управління

замовленнями, моніторингу відвідувань, аналізу продажів та взаємодії із системами лояльності. Аналогічно до інтернет-магазинів, користувацькі інтерфейси таких систем передбачають звичні елементи взаємодії – формування замовлення, рекомендації страв, персональні пропозиції або нагадування про акції. На рис. 1.1 подано загальну структуру рекомендаційної системи ресторану, яка відображає основні інформаційні потоки між модулями збору даних, аналітичними алгоритмами та інтерфейсом користувача.

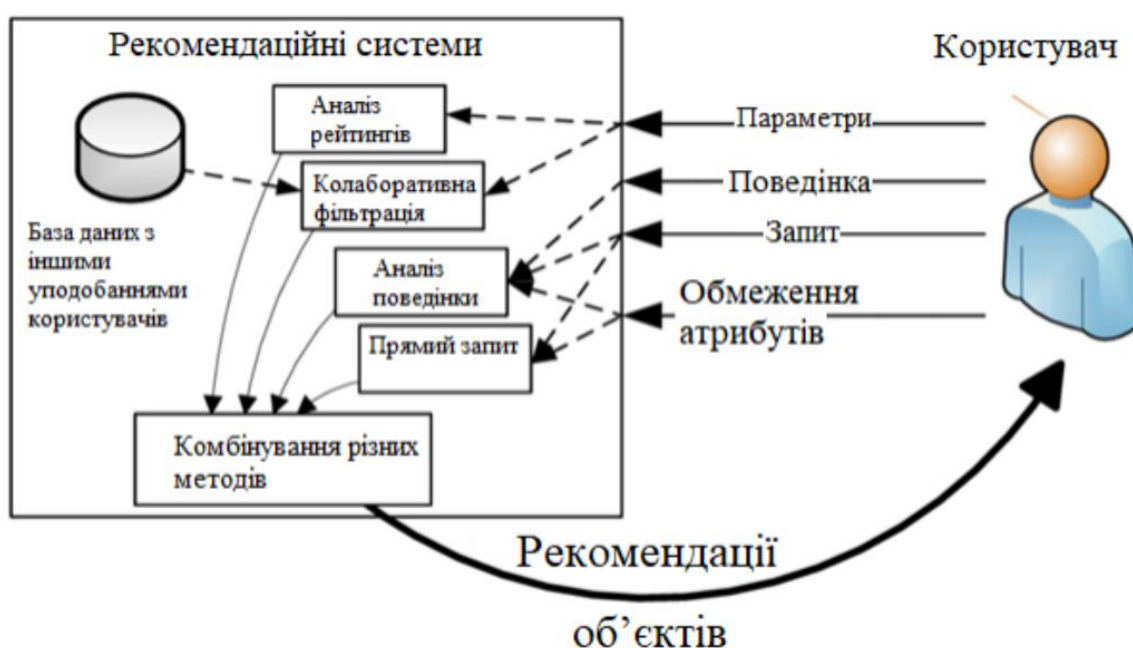


Рисунок 1.1 – Структура рекомендаційної системи ресторану

Рекомендаційні системи в ресторанній сфері орієнтовані на надання персоналізованих рекомендацій клієнтам на основі аналізу їхніх замовлень, відгуків, частоти відвідувань і фінансових показників. Завдяки таким системам підвищується рівень задоволеності відвідувачів, зростає ефективність комунікацій та збільшується середній чек. Проте для досягнення високої точності рекомендацій важливо правильно групувати клієнтів за ознаками подібності, що реалізується методами кластерного аналізу.

Кластеризація або кластерний аналіз – це процес поділу множини об'єктів на неперетинні підмножини (кластери) таким чином, щоб об'єкти всередині одного кластеру мали максимальну схожість між собою, а між різними кластерами – істотні відмінності [4]. Цей метод належить до класу алгоритмів навчання без учителя і широко використовується для сегментації користувачів, виявлення закономірностей у їхній поведінці та побудови прогнозних моделей.

У межах даного дослідження кластери розглядаються як групи клієнтів ресторану, які мають спільні риси у своїй поведінці – наприклад, у частоті відвідувань, виборі страв, часу доби замовлень або середній сумі витрат. Об'єкти кластерів – це індивідуальні користувачі, дії яких аналізуються системою для формування відповідних рекомендацій. Такий підхід дозволяє автоматизувати процес сегментації клієнтів, оптимізувати маркетингові кампанії та персоналізувати обслуговування.

Побудова ефективної рекомендаційної системи ресторану потребує попереднього аналізу існуючих рішень у цій галузі, визначення їхніх обмежень і виокремлення найбільш успішних практик. Важливо не лише запозичити базові архітектурні підходи, а й удосконалити їх, інтегрувавши адаптивні моделі кластеризації, що враховують сезонність, контекст споживання та зміни у поведінці клієнтів. Такий підхід забезпечує наукову новизну дослідження та сприяє створенню інтелектуальної системи, здатної динамічно адаптуватися до потреб сучасного ресторанного бізнесу.

## 1.2 Аналіз застосування методів інтелектуального аналізу даних у ресторанних системах

Інтелектуальний аналіз даних є ключовим інструментом сучасних інформаційних систем ресторанного бізнесу. Він дозволяє виявляти приховані закономірності у великих масивах даних, що формуються під час взаємодії клієнтів із рестораном. Основна мета застосування методів Data Mining –

перетворення неструктурованої та розрізненої інформації на практично корисні знання, що використовуються для прогнозування поведінки клієнтів, формування персоналізованих рекомендацій та оптимізації бізнес-процесів [5].

До ключових напрямів використання інтелектуального аналізу даних у ресторанному бізнесі належать:

- сегментація та кластеризація клієнтів (групування користувачів за поведінковими, демографічними та фінансовими характеристиками. Використовуються алгоритми k-середніх, DBSCAN та ієрархічна кластеризація для формування цільових сегментів);

- прогнозування попиту (застосування регресійних моделей, аналізу часових рядів та нейронних мереж для оцінки попиту на страви в різні періоди);

- рекомендаційні системи (використання методів асоціативних правил (Apriori, FP-Growth) для формування персоналізованих рекомендацій щодо страв або комбінацій замовлень);

- аналіз відгуків та емоцій клієнтів (застосування методів NLP та аналізу тональності для оцінки задоволеності сервісом і контролю репутації закладу);

- оптимізація ресурсів і персоналу (аналіз даних про завантаження залу, тривалість обслуговування та пікові періоди для прогнозування потреби в персоналі та ресурсах).

Методи інтелектуального аналізу даних реалізуються через машинне навчання [6]:

- методи навчання без учителя (кластеризація, PCA, карти Кохонена) для автоматичної сегментації клієнтів без попередньої розмітки;

- методи навчання з учителем (логістична регресія, дерева рішень, випадкові ліси, SVM) для прогнозування повторних замовлень, оцінки задоволеності та ймовірності лояльності;

– гібридні підходи, що включає поєднання кластеризації та рекомендаційних алгоритмів для підвищення точності персоналізації пропозицій.

Впровадження методів інтелектуального аналізу даних у ресторанних системах дозволяє:

– підвищити якість обслуговування через персоналізацію рекомендацій;

– зменшити операційні витрати через оптимізацію запасів та планування персоналу;

– поліпшити стратегічне планування на основі прогнозів попиту;

– збільшити лояльність клієнтів і частоту повторних відвідувань.

Кластеризація користувачів є базовим інструментом для побудови персоналізованих рекомендацій. Використання методів Data Mining дозволяє підвищити точність рекомендацій, адаптувати сервіси до змін у поведінці клієнтів та забезпечити конкурентні переваги ресторанних підприємств у цифровій економіці.

### 1.3 Актуальність використання досліджуваних методів у ресторанному бізнесі

Сучасний ресторанний бізнес характеризується високою конкуренцією та швидкою зміною споживчих вподобань. У таких умовах традиційні методи маркетингу та обслуговування не забезпечують достатньої ефективності для утримання клієнтів та збільшення прибутку. Застосування методів інтелектуального аналізу даних, зокрема кластеризації користувачів і рекомендаційних алгоритмів, дозволяє адаптувати пропозиції до індивідуальних потреб клієнтів, підвищити рівень лояльності та стимулювати повторні відвідування [7].

Інтеграція методів кластеризації та рекомендаційних систем у ресторанні інформаційні системи забезпечує:

- персоналізацію сервісу (кожен клієнт отримує індивідуальні пропозиції страв і комбінацій замовлень);
- оптимізацію управління ресурсами (прогнозування пікових навантажень і планування персоналу);
- збільшення доходів (за рахунок підвищення середнього чеку та стимулювання повторних замовлень);
- аналітичну підтримку прийняття рішень (оцінка ефективності маркетингових кампаній, сезонних пропозицій та акцій).

Використання кластеризації користувачів у рекомендаційних системах ресторанного бізнесу дозволяє:

- досліджувати приховані закономірності у поведінці клієнтів;
- підвищити точність персоналізованих рекомендацій за рахунок сегментації аудиторії;
- розробити адаптивні моделі, що враховують сезонність, зміни уподобань та контекст замовлень.

Актуальність досліджуваних методів обумовлена необхідністю підвищення конкурентоспроможності ресторанних підприємств у цифрову еру. Використання кластеризації та методів інтелектуального аналізу даних дозволяє створювати інтелектуальні системи, здатні ефективно аналізувати поведінку клієнтів, прогнозувати попит і надавати персоналізовані рекомендації, що сприяє підвищенню рівня обслуговування та економічної ефективності закладів.

#### 1.4 Постановка задачі дослідження

Метою роботи є дослідження та порівняння методів кластеризації користувачів, що застосовуються для формування персоналізованих рекомендацій у ресторанних системах, а також оцінка їх ефективності для підвищення точності рекомендацій та задоволеності клієнтів.

Для досягнення мети поставлено наступні завдання:

- проаналізувати сучасний стан розвитку інформаційних систем у ресторанному бізнесі та визначити напрями їх інтелектуалізації;
- дослідити можливості використання методів інтелектуального аналізу даних для сегментації користувачів у ресторанних системах;
- розглянути та порівняти існуючі підходи кластеризації, класифікації, асоціативних правил і регресійного аналізу в контексті побудови рекомендаційних моделей;
- розробити логічну та фізичну модель бази даних користувачів ресторанної системи, необхідну для виконання сегментації;
- сформулювати математичний опис гібридного методу сегментації користувачів, який поєднує декілька підходів інтелектуального аналізу даних;
- здійснити попередню обробку, нормалізацію та підготовку вибіркового даних користувачів ресторану для експериментального дослідження;
- провести порівняльний аналіз результатів кластеризації за допомогою різних методів і оцінити ефективність гібридного підходу;
- узагальнити отримані результати та зробити висновки щодо можливостей підвищення точності рекомендацій у ресторанних системах.

## 2 ОГЛЯД МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

### 2.1 Методи кластерного аналізу

Кластерний аналіз є фундаментальним методом інтелектуального аналізу даних, що дозволяє виявляти приховані структурні закономірності у множинах об'єктів. Його суть полягає у поділі множини елементів на групи за критерієм подібності, при якому об'єкти одного кластера мають високий ступінь внутрішньої однорідності, а відмінності між кластерами є статистично значущими [8]. У межах машинного навчання кластеризація відноситься до методів навчання без учителя і забезпечує формування уявлення про структуру даних на основі обчислюваних метрик відстані чи подібності без попереднього знання про групову належність об'єктів.

Для сегментації користувачів інформаційної системи ресторанного бізнесу кластерний аналіз дозволяє виділяти групи клієнтів із подібними поведінковими або економічними характеристиками. Це можуть бути показники частоти відвідувань, середнього розміру замовлення, переважного часу відвідування та активності у бонусних програмах. Результати кластеризації формують аналітичну основу для побудови моделей споживчої поведінки, прогнозування попиту та персоналізації сервісу.

Математично задача кластеризації полягає у знаходженні розбиття множини спостережень  $X = \{x_1, x_2, \dots, x_n\}$ , де кожен елемент  $x_i \in R^m$  є вектором ознак користувача на підмножині  $C_1, C_2, \dots, C_k$ , що мінімізують внутрішньокластерні відстані, описувані функціоналом [9]:

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} d(x_i, \mu_j), \quad (2.1)$$

де  $d(x_i, \mu_j)$  – міра відстані між об'єктом і центром кластера  $\mu_j$ .

Метод  $k$ -середніх є класичним алгоритмом розподільного типу, який ітеративно уточнює положення центрів кластерів, мінімізуючи суму відстаней між об'єктами та їхніми центрами [10]. Кожна ітерація включає переназначення об'єктів до найближчого центру та перерахунок центру як середнього значення точок, що належать до кластера.  $K$ -Means забезпечує високу швидкість обробки великих масивів даних і простоту реалізації, проте чутливий до вибору початкових центрів та кількості кластерів, що потребує додаткової оцінки оптимального параметра  $k$ .

Другим методом є OPTICS, алгоритм щільнісної кластеризації, концептуально подібний до DBSCAN, але здатний виявляти кластери змінної щільності та більш стійкий до вибору порогових параметрів. OPTICS формує впорядковану послідовність об'єктів на основі їхньої щільності та відстаней до сусідів, що дозволяє будувати структуру кластерів у вигляді діаграми доступності [11]. Цей підхід особливо ефективний для сегментації користувачів із різною інтенсивністю відвідувань та неоднорідними поведінковими патернами, оскільки дозволяє відокремити щільні групи від менш активних або рідкісних клієнтів.

Третім методом є DBSCAN, який виділяє кластери на основі щільності даних і дозволяє ідентифікувати об'єкти із аномальною поведінкою, що не належать до жодного кластера. DBSCAN формує кластери як області високої щільності, де кожна точка має мінімальну кількість сусідів у межах заданого радіуса [12]. Метод ефективний для виявлення кластерів довільної форми та стійкий до шумових даних, що робить його доцільним для сегментації користувачів із різними типами взаємодії з інформаційною системою ресторану.

Якість кластеризації оцінюється за внутрішніми критеріями компактності та відокремленості кластерів, що визначаються на основі внутрішньогрупових відстаней. Одним із найбільш інформативних показників є коефіцієнт силуету, який характеризує ступінь наближення об'єкта до власного та сусіднього кластера.

## 2.2 Методи класифікації

Класифікація є ключовим напрямом машинного навчання з навчанням під наглядом, що полягає у встановленні залежності між множиною характеристик об'єктів та їх належністю до певних категорій. У рамках задач сегментації користувачів інформаційних систем ресторанного бізнесу класифікація дозволяє відносити клієнта до конкретного сегмента на основі його поведінкових, демографічних та економічних ознак. Завдяки цьому можливо прогнозувати майбутні дії користувачів, оцінювати потенційну цінність клієнтів для бізнесу та розробляти персоналізовані маркетингові стратегії. Класифікаційні моделі дозволяють здійснювати диференційовану роботу з різними групами клієнтів, підвищуючи ефективність взаємодії та оптимізуючи ресурси інформаційної системи.

Одним із базових методів класифікації є логістична регресія, яка забезпечує моделювання ймовірності належності об'єкта до певного класу через сигмоїдну функцію для бінарної класифікації або через софтмакс-функцію для багатокласової задачі. Математична сутність логістичної регресії полягає у перетворенні лінійної комбінації вхідних ознак у ймовірнісну оцінку класу, що дозволяє не лише здійснювати класифікацію, а й оцінювати ступінь впевненості моделі у своїх прогнозах [13]. У застосуванні до сегментації користувачів ресторанної системи логістична регресія дозволяє, наприклад, передбачити ймовірність того, що клієнт належить до активної групи відвідувачів або до сегмента рідкісних користувачів, враховуючи такі параметри, як частота замовлень, середня сума чека, час відвідування та участь у бонусних програмах.

Другим методом є  $k$ -ближчих сусідів ( $k$ NN), який відноситься до методів відкладеного навчання і не вимагає побудови явної моделі під час етапу навчання. Принцип роботи полягає у віднесенні нового об'єкта до того класу, який представлений більшістю серед його  $k$  найближчих сусідів у

багатовимірному просторі ознак [14]. Математично це можна описати формулою:

$$\hat{y} = \text{mod} \{y_i | x_i \in N_k(x)\}, \quad (2.2)$$

де  $N_k(x)$  – множина  $k$  найближчих сусідів об'єкта  $x_i$ ;

$y_i$  – класи об'єктів.

Метод kNN особливо корисний для класифікації користувачів із неоднорідними або складними поведінковими патернами, коли межі класів не визначені аналітично. Наприклад, клієнти, які мають схожі поєднання показників відвідуваності та середнього чека, автоматично групуються у відповідні сегменти на основі їхніх найближчих аналогів у навчальному наборі.

Третім методом є метод опорних векторів (SVM), який застосовує геометричний підхід до класифікації шляхом побудови гіперплощини, що максимально розділяє класи у багатовимірному просторі ознак. Для випадків, коли класи не є лінійно роздільними, використовуються ядрові функції, які перетворюють простір ознак у більш високовимірний простір, де відбувається оптимальне розділення. SVM демонструє високу ефективність у випадках складних поведінкових патернів користувачів, дозволяючи моделі точно розрізняти сегменти навіть при наявності шумових або неповних даних [15].

Якість класифікаційних моделей оцінюється за допомогою метрик точності, повноти, F1-міри та індексу Джіні, що забезпечує об'єктивну оцінку здатності моделі правильно прогнозувати належність нових користувачів до відповідних сегментів. Поєднання зазначених методів з алгоритмами відбору релевантних ознак, нормалізації та балансування даних дозволяє досягати високої точності класифікації та забезпечує науково обґрунтовану сегментацію користувачів інформаційної системи ресторанного бізнесу.

### 2.3 Методи асоціативних правил

Методи асоціативних правил належать до класу алгоритмів інтелектуального аналізу даних, що призначені для виявлення закономірностей та зв'язків між ознаками у великих наборах транзакцій або поведінкових даних. Основна ідея полягає у формуванні правил виду  $A \Rightarrow B$ , які відображають взаємозв'язок між подіями або характеристиками, що часто зустрічаються разом. У контексті сегментації користувачів інформаційної системи ресторанного бізнесу асоціативні правила дозволяють виявляти стійкі залежності між замовленнями, категоріями страв, часовими періодами відвідувань і демографічними особливостями клієнтів [16]. Це забезпечує можливість побудови моделей споживчої поведінки, прогнозування переваг клієнтів і розроблення персоналізованих пропозицій.

Найпоширенішим підходом до побудови таких правил є алгоритм Apriori, який ґрунтується на поступовому пошуку частих наборів елементів у транзакційній базі. На першому етапі алгоритм виявляє усі поодинокі елементи, що перевищують заданий поріг підтримки, тобто частоти їх появи у вибірці. Далі з отриманих частих елементів формуються нові комбінації більшої довжини, і процес повторюється, поки не залишиться жодного набору, що задовольняє мінімальний рівень підтримки. Для кожного знайденого правила розраховується міра довіри (*confidence*), яка визначає умовну ймовірність появи події  $B$  за умови, що подія  $A$  вже відбулася, тобто [17]:

$$\text{confidence}(A \Rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}, \quad (2.3)$$

де  $\text{support}(A \Rightarrow B)$  – частота спільної появи подій  $A$  та  $B$ ;

$\text{support}(A)$  – частота появи події  $A$ .

Високе значення довіри свідчить про наявність сильної асоціації між ознаками.

Альтернативним підходом є алгоритм FP-Growth, який усуває необхідність у багаторазовому генеруванні наборів елементів. Він використовує компактну структуру даних – FP-дерево (Frequent Pattern Tree), що дозволяє зберігати інформацію про часті елементи та їх поєднання без повторного сканування всієї бази транзакцій. FP-Growth забезпечує вищу ефективність при роботі з великими обсягами даних, що особливо актуально для ресторанних інформаційних систем, де база клієнтських замовлень може містити десятки або сотні тисяч записів [18]. Застосування цього методу дозволяє виявляти складні поведінкові закономірності, наприклад, що клієнти, які часто замовляють каву у ранкові години, з високою ймовірністю замовляють десерти у вечірній період.

Асоціативні правила широко використовуються для побудови рекомендаційних підсистем у межах ресторанних інформаційних систем. Вони дозволяють формувати індивідуальні пропозиції на основі історії замовлень конкретного клієнта, підвищувати середній чек, оптимізувати меню та проводити ефективний таргетинг рекламних кампаній. Крім того, отримані закономірності можуть використовуватися для формування нових сегментів клієнтів, які демонструють подібні споживчі патерни, що підвищує точність сегментації та дозволяє більш гнучко управляти взаємодією з користувачами.

Таким чином, методи асоціативних правил становлять важливу складову інтелектуального аналізу даних у сфері ресторанного бізнесу. Вони забезпечують можливість виявлення прихованих зв'язків між елементами клієнтської активності, що створює наукове підґрунтя для прийняття управлінських рішень, спрямованих на підвищення лояльності користувачів, персоналізацію сервісів і стратегічне планування розвитку бізнесу.

## 2.4 Методи регресійного аналізу

Регресійний аналіз є одним із базових інструментів інтелектуального аналізу даних, який забезпечує виявлення, кількісне оцінювання та інтерпретацію залежностей між змінними. Його сутність полягає у побудові математичної моделі, що описує вплив незалежних змінних на залежну змінну, яка характеризує певний результат, процес або поведінкову ознаку. У контексті інформаційних систем ресторанного бізнесу регресійний аналіз дає змогу дослідити, як різні фактори, наприклад, демографічні характеристики клієнтів, середня вартість замовлення, періодичність відвідувань чи участь у програмах лояльності впливають на економічні показники або поведінку користувачів. Такий підхід дозволяє не лише пояснювати спостережувані закономірності, а й здійснювати прогнозування майбутніх значень цільових параметрів.

Основу класичного регресійного підходу становить лінійна регресія, яка описує зв'язок між змінними за допомогою лінійної комбінації незалежних параметрів. Модель лінійної регресії має вигляд [19]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon, \quad (2.4)$$

де  $y$  – залежна змінна;

$x_1, x_2, \dots, x_n$  – незалежні змінні;

$support(A)$  – параметри моделі;

$\varepsilon$  – випадкова похибка.

У застосуванні до задач сегментації користувачів така модель може відображати, наприклад, залежність середнього чека від частоти відвідувань, часу доби, дня тижня чи типу ресторанного меню. Лінійна регресія дозволяє інтерпретувати вплив кожного окремого чинника, що робить її важливим інструментом для виявлення ключових детермінант поведінки клієнтів.

Проте у реальних умовах поведінка користувачів рестораних сервісів часто не є лінійною. Наприклад, вплив рівня знижок на ймовірність повторного замовлення може зростати лише до певного порогу, після чого

перестає бути суттєвим. Для врахування подібних ефектів застосовується нелінійна регресія, яка дає можливість описувати складні функціональні залежності між змінними. Цей підхід дозволяє враховувати взаємодію ознак, ефекти насичення та порогові значення. У ресторанному бізнесі нелінійні регресійні моделі використовуються, наприклад, для аналізу впливу кількості маркетингових повідомлень на рівень залученості клієнта або для прогнозування реакції на нові страви в меню залежно від попередніх уподобань.

Важливим напрямом розвитку регресійного аналізу є регуляризація, яка дозволяє уникнути перенавчання моделей і забезпечити стабільність результатів при роботі з великою кількістю змінних. Методи регуляризації, такі як Ridge-регресія (L2-норма) та Lasso-регресія (L1-норма), вводять додатковий штраф у функцію втрат, що обмежує величину коефіцієнтів [20]. У задачах сегментації користувачів це має практичне значення, оскільки дані часто містять значну кількість взаємопов'язаних характеристик – соціально-демографічних, поведінкових та транзакційних. Регуляризовані моделі дозволяють автоматично відсікати несуттєві змінні, зберігаючи лише ті, що найбільше впливають на поведінку користувача або на його належність до певного сегмента.

Особливу увагу в сучасних дослідженнях приділяють часовим регресійним моделям, які враховують динамічний характер даних. У ресторанній галузі поведінка користувачів може змінюватися залежно від пори року, економічної ситуації або рекламних кампаній. Часові регресійні моделі дозволяють відстежувати ці зміни у вигляді трендів, сезонних коливань і стохастичних збурень. Наприклад, можна прогнозувати зростання попиту на певні категорії страв у святкові періоди або зниження активності користувачів у міжсезоння. Таким чином, регресійний підхід сприяє створенню адаптивних стратегій управління клієнтськими сегментами.

Отже, методи регресійного аналізу становлять універсальний інструмент для дослідження поведінкових закономірностей і прогнозування

показників діяльності користувачів у ресторанному бізнесі. Вони забезпечують кількісну оцінку впливу окремих факторів, виявлення прихованих залежностей і можливість прогнозування результатів у різних умовах функціонування системи. Інтеграція регресійних моделей у процес аналітики клієнтських даних підвищує точність сегментації, сприяє персоналізації сервісів і створює підґрунтя для розроблення стратегій підвищення лояльності та прибутковості клієнтської бази.

## 2.5 Узагальнення методів для вирішення задач сегментації цільових груп

Методи регресійного аналізу становлять важливий інструментарій у процесі побудови моделей, що дозволяють кількісно описувати залежність між поведінковими, соціально-демографічними та економічними характеристиками користувачів ресторанної інформаційної системи. Їхнє використання дає змогу не лише оцінити силу впливу окремих факторів на результативні показники, але й виявити латентні закономірності, що формують основи сегментації споживачів. Застосування регресійних методів у сфері ресторанного бізнесу охоплює прогнозування частоти відвідувань, оцінювання схильності до участі у програмах лояльності, аналіз тенденцій зміни середнього чеку, а також дослідження впливу маркетингових акцій на поведінку клієнтів.

Лінійні регресійні моделі є базовим підходом, який дозволяє встановити прямий зв'язок між незалежними змінними та цільовим показником. Вони зручні для початкового етапу аналізу, коли важливо зрозуміти напрям і силу взаємозалежностей між параметрами. Проте лінійність обмежує їх застосування, оскільки більшість поведінкових процесів користувачів не є строго лінійними. У таких випадках доцільно застосовувати нелінійні регресійні методи, що дозволяють гнучкіше апроксимувати складні взаємозв'язки між змінними.

Нелінійні підходи, зокрема поліноміальна, логістична та експоненційна регресія, є більш адаптивними до реальних даних користувачів, які часто мають складну структуру, велику дисперсію та множинні кореляції між параметрами. Наприклад, логістична регресія дає змогу моделювати ймовірність належності користувача до певного сегмента, враховуючи взаємозв'язок між числовими та категоріальними ознаками. Поліноміальні методи забезпечують кращу точність прогнозування у випадках, коли поведінка клієнтів змінюється нерівномірно залежно від зовнішніх чинників, таких як сезонність або зміна меню.

Окрему групу складають ансамблеві методи регресії, що використовують об'єднання декількох моделей для підвищення точності прогнозу. До найпоширеніших належать [21]:

- Random Forest, який базується на використанні множини незалежних дерев рішень, кожне з яких формує власне передбачення, а підсумковий результат визначається за принципом усереднення;

- Gradient Boosting, що послідовно створює нові моделі, які компенсують помилки попередніх, забезпечуючи поступове вдосконалення прогнозу;

- XGBoost, тобто оптимізована реалізація бустингу, яка забезпечує високу швидкість навчання і здатність працювати з великими наборами даних користувачів.

Попри високу точність, ансамблеві методи характеризуються зниженням інтерпретованості результатів, що ускладнює розуміння механізмів прийняття рішень. Це є суттєвим недоліком у бізнес-контексті, де важливо пояснювати, чому система віднесла певного клієнта до конкретного сегмента або чому прогнозована поведінка користувача змінюється під впливом певних параметрів.

Спільним обмеженням усіх методів регресійного аналізу є їхня недостатня здатність до адаптації у середовищі, де дані постійно змінюються, а поведінкові патерни користувачів еволюціонують. Традиційні підходи не

враховують складних нелінійних взаємодій між множиною ознак, що призводить до втрати точності при довготривалому використанні. Крім того, ці методи часто демонструють чутливість до вибірки та наявності шумів у даних, що може спотворювати кінцеві результати сегментації.

Зазначені недоліки формують підґрунтя для розроблення гібридного методу, який поєднують аналітичну основу регресійних моделей із самоадаптивними можливостями алгоритмів машинного навчання.

## 2.6 Опис розробленого гібридного методу сегментації користувачів

Розроблений гібридний метод сегментації користувачів ресторанної інформаційної системи базується на інтеграції принципів виявлення щільних груп, оцінки ймовірності належності до сегменту та прогнозування поведінкових показників. Метою є формування стабільних сегментів користувачів з урахуванням складних та непередбачуваних патернів поведінки, що характерні для ресторанної сфери, та забезпечення можливості адаптивного прогнозування активності клієнтів.

Алгоритм роботи гібридного методу включає наступні етапи:

а) попередня обробка та нормалізація даних:

- збір даних про користувачів, включаючи поведінкові ознаки, історію замовлень, участь у акціях та демографічні характеристики;
- масштабування числових ознак та кодування категоріальних змінних для забезпечення коректного порівняння об'єктів у багатовимірному просторі;

б) виявлення первинних сегментів на основі щільності:

- використання алгоритму DBSCAN, для виявлення природних груп користувачів із високою щільністю сусідів;
- формування початкових кластерів та позначення шумових або аномальних користувачів, які не належать до основних сегментів;

в) оцінка ймовірності належності користувачів:

– для кожного користувача обчислюється ймовірність належності до кожного сегмента на основі дистанції до центрів щільності та локальної структури даних;

– математично ймовірність може описуватися формулою:

$$P(C_j | x_i) = \frac{\exp(-d(x_i, \mu_j))}{\sum_{k=1}^K \exp(-d(x_i, \mu_k))}, \quad (2.5)$$

де  $d(x_i, \mu_j)$  – відстань між користувачем  $x_i$  та центром сегмента  $\mu_j$ ;

$K$  – кількість сегментів;

г) прогнозування поведінкових показників:

– для кожного сегмента будується регресійна модель, яка дозволяє прогнозувати ключові показники активності користувачів, такі як середній чек, ймовірність повторного замовлення або участь у акціях;

– моделювання здійснюється із врахуванням індивідуальних ознак користувачів та характеристик сегмента;

д) адаптивне уточнення сегментів:

– на основі прогнозованих показників та ймовірності належності користувачів сегменти коригуються, що дозволяє враховувати появу нових патернів поведінки та зміну активності клієнтів;

– процес уточнення повторюється циклічно до досягнення стабільності кластерів та прогнозів.

Розроблений гібридний метод поєднує переваги щільнісної кластеризації, ймовірнісного підходу та регресійного прогнозування, створюючи адаптивну систему сегментації. Основним його перевагою є здатність виділяти сегменти навіть у випадках високої стохастичності поведінки користувачів і прогнозувати зміни ключових показників сегментів, що забезпечує науково обґрунтовану основу для персоналізації маркетингових стратегій та управління сервісом ресторанної інформаційної системи.

### 3 ЛОГІЧНЕ ТА ФІЗИЧНЕ ПРОЄКТУВАННЯ БАЗИ ДАНИХ

#### 3.1 Логічне проєктування бази даних користувачів ресторанної системи

База даних системи ресторану повинна містити наступні дані:

- меню страв;
- інформацію про користувачів системи ресторану;
- інформацію про бронювання столиків;
- інформація замовлених страв.

З використанням стандарт IDEF1X створено логічну модель бази даних системи ресторану [22], що показана на рис. 3.1.

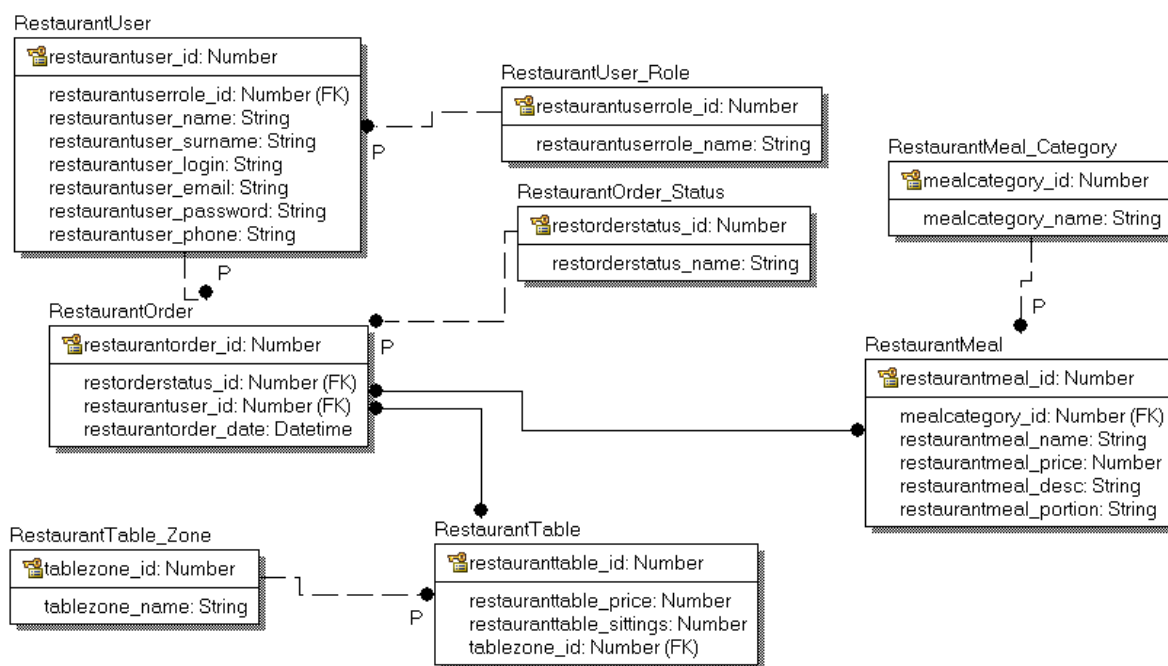


Рисунок 3.1 – Логічна модель даних системи ресторану

Сутність «restaurantuser» призначена для збереження персональних даних користувачів, які мають власні облікові записи у системі. У ній фіксується роль користувача (адміністратор або відвідувач), а також інформація, необхідна для авторизації: логін, пароль, ім'я, прізвище, електронна адреса та номер телефону. Основні атрибути цієї сутності:

- restaurantuser\_id – унікальний числовий ідентифікатор користувача;
- restaurantuserrole\_id – зовнішній ключ, що визначає роль користувача;
- restaurantuser\_name, restaurantuser\_login, restaurantuser\_email, restaurantuser\_password, restaurantuser\_phone – атрибути рядкового типу, які зберігають особисту інформацію.

Сутність «restaurantuser\_role» містить відомості про типи ролей користувачів, що визначають рівень доступу до функціоналу системи. До її складу входять такі атрибути:

- restaurantuserrole\_id – числовий ідентифікатор ролі;
- restaurantuserrole\_name – назва ролі у форматі рядка.

Сутність «restaurantorder» використовується для зберігання інформації про замовлення клієнтів. Вона містить дані про користувача, статус замовлення та дату його створення. До складу атрибутів належать:

- restaurantorder\_id – унікальний числовий ідентифікатор замовлення;
- restaurantuser\_id – посилання на користувача, який здійснив замовлення;
- restaurantorderstatus\_id – зовнішній ключ на статус замовлення;
- restaurantorder\_date – атрибут, що містить дату та час створення.

Сутність «restaurantorder\_status» призначена для фіксації можливих станів замовлення, таких як «очікує підтвердження», «виконується» або «завершено». Вона включає два основних атрибути:

- restaurantorderstatus\_id – числовий ідентифікатор статусу;
- restaurantorderstatus\_name – текстова назва статусу.

Сутність «restaurantmeal» містить інформацію про страви, представлені у меню ресторану. Для кожної страви вказується її назва, опис, ціна, порція та належність до певної категорії. Основні атрибути:

- restaurantmeal\_id – числовий ідентифікатор страви;
- mealcategory\_id – зовнішній ключ на категорію страви;
- restaurantmeal\_name, restaurantmeal\_description, restaurantmeal\_portion – рядкові атрибути, що описують страву;

– restaurantmeal\_price – числове поле, що зберігає вартість.

Сутність «restaurantmeal\_category» систематизує страви за типами, наприклад, «гарячі страви», «салати», «десерти». До неї належать такі атрибути:

- restaurantmealcategory\_id – числовий ідентифікатор категорії;
- restaurantmealcategory\_name – назва категорії у рядковому форматі.

Сутність «restauranttable» використовується для зберігання інформації про доступні до бронювання столики. Для кожного столика вказується його номер, ціна бронювання, кількість місць і зона розташування. Основні атрибути:

- restauranttable\_id – числовий ідентифікатор столика;
- restauranttable\_price – вартість бронювання;
- restauranttable\_sitting – кількість посадкових місць;
- tablezone\_id – зовнішній ключ, що посилається на зону розміщення.

Сутність «restauranttable\_zone» визначає зони розташування столиків, наприклад «тераса», «зал біля бару», «віконна зона». До її складу входять:

- tablezone\_id – числовий ідентифікатор зони;
- tablezone\_name – назва зони у форматі рядка.

### 3.2 Фізичне проектування бази даних користувачів ресторанної системи

У процесі логічного моделювання було виявлено наявність зв'язків типу «багато до багатьох» між таблицями RestaurantOrder – RestaurantMeal та RestaurantOrder – RestaurantTable. Подібна структура не є придатною для реалізації у фізичній моделі бази даних, оскільки порушує принципи нормалізації та ускладнює забезпечення цілісності даних. Для усунення таких зв'язків застосовано метод введення проміжних таблиць, які виконують функцію зв'язувальних елементів між основними сутностями.

Зокрема, для усунення зв'язку між таблицями RestaurantOrder та RestaurantMeal створено допоміжну таблицю Meal\_Cart, яка містить зовнішні

ключі, що посилаються на відповідні первинні ключі базових таблиць. Таким чином, кожен запис у таблиці Meal\_Cart відповідає конкретній страві, включеній до певного замовлення, що забезпечує можливість фіксації складу замовлення та кількості позицій.

Аналогічно, для відображення зв'язку між таблицями RestaurantOrder та RestaurantTable створено проміжну таблицю Table\_Cart. Фізичну схему даних системи ресторану представлено на рис. 3.2.

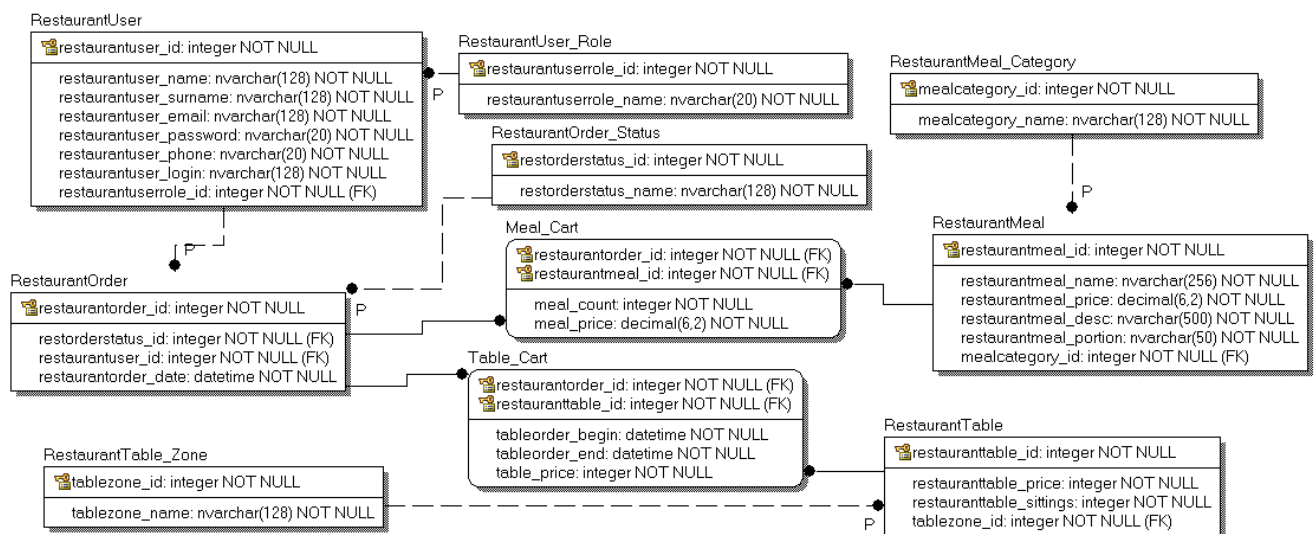


Рисунок 3.2 – Фізична модель даних системи ресторану

На основі наведеної фізичної моделі бази даних ресторанної інформаційної системи виділено ключові сутності, що забезпечують повноцінне функціонування механізмів бронювання, замовлення страв і управління користувачами. Кожна сутність виконує окрему логічну функцію в системі, а її атрибути визначають структурні особливості зберігання даних.

Сутність RestaurantUser є центральною для моделі, оскільки містить дані про користувачів системи, включно з клієнтами та адміністраторами. Вона має такі атрибути: ідентифікатор користувача restaurantuser\_id, який є первинним ключем; атрибути restaurantuser\_name, restaurantuser\_surname, restaurantuser\_email, restaurantuser\_login і restaurantuser\_password, що використовуються для ідентифікації, комунікації та авторизації користувача;

а також поле `restaurantuser_phone`, яке забезпечує зберігання контактного номера. Поле `restaurantuserrole_id` є зовнішнім ключем, що пов'язує користувача з роллю у системі.

Сутність `RestaurantUser_Role` визначає набір можливих ролей користувачів у системі, таких як клієнт або адміністратор. Вона складається з двох атрибутів: `restaurantuserrole_id`, який є унікальним ідентифікатором ролі, та `restaurantuserrole_name`, що описує її призначення. Зв'язок із таблицею користувачів реалізовано через зовнішній ключ у таблиці `RestaurantUser`.

Сутність `RestaurantOrder` описує замовлення, що здійснюються користувачами. Вона містить атрибути `restaurantorder_id` як первинний ключ, `restaurantuser_id` для визначення користувача, який здійснив замовлення, `restorderstatus_id` – зовнішній ключ на статус замовлення, а також `restaurantorder_date`, який фіксує дату і час створення замовлення. Ця таблиця забезпечує зв'язок між користувачами, їхніми замовленнями та поточним станом виконання цих замовлень.

Сутність `RestaurantOrder_Status` використовується для зберігання можливих статусів замовлень, таких як «нове», «виконується» або «завершене». Її структура включає атрибути `restorderstatus_id` як унікальний ідентифікатор та `restorderstatus_name`, який зберігає текстову назву статусу. Такий підхід дозволяє реалізувати контроль життєвого циклу замовлення та відображення його поточного стану в системі.

Сутність `RestaurantMeal` описує довідник страв, доступних у меню ресторану. Вона має атрибути `restaurantmeal_id` як первинний ключ, `mealcategory_id` – зовнішній ключ на категорію страви, `restaurantmeal_name` для зберігання назви, `restaurantmeal_desc` — опису страви, `restaurantmeal_price` – її вартості, а також `restaurantmeal_portion`, який визначає розмір порції. Ця таблиця слугує базою даних для формування меню та подальшого використання інформації в процесі оформлення замовлення.

Сутність `RestaurantMeal_Category` містить перелік категорій, до яких належать страви меню. Її атрибути включають `mealcategory_id` як первинний

ключ та `mealcategory_name` для текстової назви категорії. Така структура забезпечує можливість групування страв за тематичними розділами меню, наприклад, закуски, гарячі страви, десерти тощо.

Сутність `RestaurantTable` відображає інформацію про столики, доступні для бронювання в ресторані. Вона містить атрибути `restauranttable_id` як первинний ключ, `restauranttable_price` – ціну бронювання, `restauranttable_sittings` – кількість місць за столиком, а також `tablezone_id` як зовнішній ключ, що вказує на розташування столика у певній зоні закладу.

Сутність `RestaurantTable_Zone` визначає зони розміщення столиків у приміщенні ресторану. Вона має два атрибути: `tablezone_id` як унікальний ідентифікатор зони та `tablezone_name` для зберігання її назви, наприклад, «основна зала» або «тераса». Така деталізація дозволяє формувати замовлення з урахуванням просторової організації закладу.

Допоміжна сутність `Meal_Cart` реалізує зв'язок між замовленнями та стравами, усуваючи відношення типу «багато до багатьох». Вона містить зовнішні ключі `restaurantorder_id` і `restaurantmeal_id`, а також атрибути `meal_count` для зберігання кількості одиниць кожної страви та `meal_price` – вартості замовленої позиції. Завдяки цьому забезпечується можливість гнучкого формування замовлень із кількох страв.

Аналогічну роль виконує сутність `Table_Cart`, яка встановлює зв'язок між замовленням і заброньованими столиками. Вона включає атрибути `restaurantorder_id` та `restauranttable_id` як зовнішні ключі, а також поля `tableorder_begin` і `tableorder_end` для фіксації часу початку та завершення бронювання. Додатково міститься атрибут `table_price`, що зберігає вартість бронювання.

### 3.3 Використання даних користувачів для їх сегментації

У процесі сегментації користувачів інформаційної системи ресторану використовуються дані, отримані з фізичної моделі бази даних, що відображає

всі ключові аспекти взаємодії клієнта із закладом. Ці дані мають як соціально-демографічний, так і поведінковий характер, що дає змогу формувати сегменти користувачів з урахуванням особливостей їхньої активності, уподобань та фінансової поведінки.

Основним джерелом інформації для сегментації є сутність RestaurantUser, з якої отримуються базові персональні атрибути користувачів, зокрема ім'я, прізвище, електронна пошта, логін і номер телефону. Вони забезпечують можливість ідентифікації клієнта та зіставлення його дій у системі з відповідними замовленнями й бронюваннями. Крім того, важливим є атрибут restaurantuserrole\_id, який дозволяє розмежовувати типи користувачів – клієнтів і адміністраторів – та виключати технічних користувачів із процесу сегментації.

Дані про активність користувачів надходять із сутності RestaurantOrder, яка містить інформацію про всі здійснені замовлення. Для сегментаційного аналізу використовуються такі параметри, як кількість оформлених замовлень, частота відвідувань, середній інтервал між замовленнями та дата останньої активності (restaurantorder\_date). Це дозволяє виявити як постійних клієнтів, що регулярно користуються послугами закладу, так і тих, хто відвідує його епізодично або одноразово. Додатково застосовується атрибут restorderstatus\_id, який допомагає визначити частку завершених, скасованих або поточних замовлень – такий показник відображає ступінь лояльності та надійності користувача.

Важливу роль у побудові поведінкових ознак відіграє сутність Meal\_Cart, яка забезпечує деталізацію складу кожного замовлення. Звідси отримуються показники середньої вартості замовлення (meal\_price), кількості замовлених позицій (meal\_count), частоти повторення одних і тих самих страв, а також переваги щодо певних категорій меню. У поєднанні з даними з таблиці RestaurantMeal\_Category ці параметри дозволяють визначити індивідуальні смаки користувачів, наприклад, чи тяжіє клієнт до основних страв, десертів

або напоїв, що має безпосереднє значення для персоналізації маркетингових пропозицій.

Дані з таблиці Table\_Cart використовуються для аналізу бронювань. Атрибути tableorder\_begin і tableorder\_end дозволяють розрахувати середню тривалість перебування клієнта в ресторані, частоту резервувань та дні тижня, коли користувач найчастіше здійснює бронювання. Додатково враховується параметр table\_price, який використовується для оцінки середнього рівня витрат клієнта під час бронювання столиків, що може виступати індикатором його платоспроможності.

Інформація із сутності RestaurantTable\_Zone забезпечує просторовий вимір сегментації, дозволяючи враховувати зони, у яких клієнт зазвичай бронює місця. Наприклад, користувачі, що віддають перевагу VIP-зонам або терасі, можуть бути віднесені до сегменту преміальних клієнтів, тоді як відвідувачі загальної зали частіше належать до базового сегмента.

Таким чином, на основі інтеграції даних із сутностей RestaurantUser, RestaurantOrder, Meal\_Cart, Table\_Cart та RestaurantTable\_Zone формується комплексна система ознак, що поєднує соціальні, поведінкові та фінансові характеристики користувачів. Це створює основу для застосування алгоритмів інтелектуального аналізу даних – кластеризації, класифікації та регресійного прогнозування – з метою виявлення закономірностей у поведінці клієнтів, побудови цільових сегментів та підвищення ефективності персоналізованих рекомендацій у ресторанній інформаційній системі.

## 4 ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ СЕГМЕНТАЦІЇ КОРИСТУВАЧІВ

### 4.1 Опис середовища розробки

Середовище розробки для даного проєкту базувалося на використанні Jupyter Notebook. Цей інструмент забезпечує інтерактивне виконання окремих блоків коду з миттєвою візуалізацією результатів. Збереження результатів разом із відповідними блоками коду у файлі проєкту дозволяє уникнути повторного запуску ресурсомістких обчислень, що підвищує ефективність розробки та зменшує витрати часу на тестування та аналіз [24]. Крім того, Jupyter Notebook надає можливість створення текстових комірок із маркованими списками, що дозволяє структуровано документувати алгоритми, описувати послідовність виконання обчислень та пояснювати логіку роботи програмних модулів.

Розробка програмного забезпечення проводилася на мові програмування Python 3, яка широко застосовується в численних галузях, таких як серверна веб-розробка, автоматизація обчислювальних процесів, наукові та статистичні дослідження, а також розробка системних сценаріїв [25]. Python характеризується наступними особливостями, що забезпечують ефективну реалізацію науково-орієнтованих проєктів:

- високий рівень абстракції та підтримка структурованих даних (Python надає вбудовані структури даних, що спрощує обробку великих обсягів інформації та реалізацію алгоритмів обчислювальної аналітики);

- динамічна типізація та прив'язка (мова підтримує динамічне визначення типів змінних, що дозволяє швидко адаптувати код під різні типи даних і зменшує складність написання прототипів програм);

- модульність та масштабованість (Python забезпечує гнучку організацію коду у вигляді модулів та пакетів, що спрощує повторне

використання функціональних блоків та полегшує масштабування програмного забезпечення);

– розширюваність через сторонні бібліотеки (активна спільнота розробників створює численні бібліотеки для різних завдань, від математичних обчислень і обробки даних до машинного навчання та роботи з мультимедійними даними);

– висока продуктивність виконання (основні бібліотеки Python реалізовані на мовах низького рівня (C/C++), що дозволяє досягати високої швидкості обчислень. Крім того, Python дозволяє інтегрувати власні модулі на C/C++ для оптимізації критичних обчислювальних задач);

– підтримка стандартних протоколів та інтеграцій (мова містить інструменти для роботи з електронною поштою, мережевими протоколами (HTTP, FTP), базами даних, веб-сервісами та іншими ресурсами, що робить Python універсальним інструментом для розробки комплексних проєктів);

– переносимість та кросплатформеність (скрипти Python можуть виконуватися на більшості сучасних операційних систем без модифікацій, що забезпечує гнучкість впровадження та поширення розробленого програмного забезпечення);

– легкість читання та підтримки коду (зрозумілий синтаксис, продумана модульність та послідовність конструкцій підвищують читабельність коду, полегшують його супровід та зменшують ризик виникнення помилок у процесі розробки та тестування).

Таким чином, поєднання інтерактивності Jupyter Notebook та можливостей Python забезпечує ефективну реалізацію науково-дослідного проєкту, дозволяючи одночасно виконувати складні обчислення, зберігати результати та документувати процеси аналізу й розробки.

## 4.2 Опис використаних бібліотек

Для реалізації алгоритмів та виконання поставленої задачі в даному проєкті було використано набір перевірених програмних бібліотек мови Python. Ці бібліотеки забезпечують комплексну підтримку обробки даних, математичних обчислень, візуалізації та моделювання, що є ключовими етапами дослідження. Нижче наведено основні бібліотеки, задіяні у проєкті, із зазначенням їх функціональних можливостей.

NumPy (Numerical Python) є бібліотекою, що надає багатовимірні масиви та широкий спектр функцій для їх обробки. Вона спрощує виконання математичних та логічних операцій, що спеціально оптимізовані для масивів даних [26]. Основні функціональні можливості NumPy включають:

- виконання математичних та логічних операцій над масивами;
- реалізацію перетворень Фур'є та інструментів для зміни форми масивів;
- функції лінійної алгебри та генерацію випадкових чисел.

NumPy часто використовується разом із SciPy та Matplotlib, що дозволяє створювати повноцінні аналоги середовищ для технічних обчислень, таких як MatLab, при цьому Python забезпечує більш сучасний і гнучкий підхід [27].

Pandas є бібліотекою для обробки та аналізу даних, що реалізує структури даних DataFrame і Series. Вона надає інструменти для очищення, попередньої обробки, дослідження та аналізу даних. Бібліотека підтримує роботу з різними форматами файлів та інтегрується з іншими інструментами Python, що робить її незамінною для науки про дані та машинного навчання.

Matplotlib забезпечує створення статичних, анімованих та інтерактивних візуалізацій у різних форматах. Бібліотека дозволяє реалізувати різноманітні графічні представлення даних, включаючи лінійні графіки, стовпчасті діаграми, гістограми, діаграми розсіювання тощо [28]. Основні переваги Matplotlib полягають у:

- широкій налаштованості інтерфейсу для створення графіків;
- можливості інтеграції з іншими бібліотеками Python, такими як NumPy і Pandas;

– контролю параметрів візуалізації, включаючи кольори, мітки та анотації, що забезпечує точну передачу інформації.

Scikit-Learn (sklearn) є популярною бібліотекою машинного навчання для Python, що надає широкий спектр алгоритмів для класифікації, регресії, кластеризації, зменшення розмірності та інших завдань [29]. Вона побудована на основі NumPy, SciPy і Matplotlib і характеризується простим та послідовним API. Бібліотека також забезпечує інструменти для попередньої обробки даних, вибору моделей та оцінки їх ефективності, що робить її універсальним інструментом для наукових досліджень і розробки моделей машинного навчання.

Поєднання Python із бібліотеками NumPy, Pandas, Matplotlib та Scikit-Learn обрано для реалізації проєкту кластеризації користувачів із кількох причин:

- Python забезпечує простоту та читабельність коду, а також широку екосистему бібліотек;
- NumPy і Pandas надають потужні засоби для обробки та аналізу даних, що критично важливо на етапах підготовки та дослідження набору даних;
- Matplotlib дозволяє ефективно візуалізувати дані, виявляти закономірності та інтерпретувати результати;
- Scikit-Learn надає комплексний набір алгоритмів і інструментів для побудови, тестування та оцінки моделей кластеризації.

### 4.3 Попередня обробка та нормалізація даних

Перед застосуванням алгоритмів кластеризації та побудови моделей аналітики для сегментації користувачів інформаційної системи ресторану здійснюється комплексна попередня обробка даних. Цей етап забезпечує якість, узгодженість та готовність інформації для аналітичних процедур, а також мінімізує вплив пропущених значень, аномалій і різнорідних масштабів ознак.

На основі інтеграції даних із сутностей RestaurantUser, RestaurantOrder, Meal\_Cart, Table\_Cart та RestaurantTable\_Zone формуються числові та категоріальні ознаки, що характеризують соціально-демографічні, поведінкові та фінансові аспекти користувачів. Для підготовки цих даних застосовуються наступні операції:

- обробка пропущених значень;
- кодування категоріальних змінних;
- виявлення та обробка аномалій;
- нормалізація числових ознак;
- агрегація поведінкових та фінансових показників;
- формування цільових ознак для сегментації;

Пропущені або некоректні записи в персональних атрибутах користувачів (наприклад, ім'я, прізвище, електронна пошта) та в поведінкових ознаках (дата останнього замовлення, кількість замовлених позицій) заповнюються середніми значеннями, медіаною або спеціальними маркерами для категоріальних змінних. Це забезпечує цілісність набору даних та запобігає викривленню результатів кластеризації.

Ознаки, що містять номінальні значення, такі як роль користувача (restaurantuserrole\_id) або зона бронювання (RestaurantTable\_Zone), перетворюються у числові формати методом one-hot encoding або порядкового кодування, залежно від природи змінної. Це дозволяє алгоритмам машинного навчання коректно обробляти якісні дані.

Аномальні або нетипові значення, такі як надмірно високі або низькі показники вартості замовлення (meal\_price) або часу перебування за столиком (tableorder\_end – tableorder\_begin), відокремлюються або коригуються. Це зменшує вплив екстремальних значень на моделі кластеризації та прогнозування.

Оскільки різні показники мають різні масштаби (наприклад, кількість замовлень, середня вартість замовлення, частота відвідувань), здійснюється їхнє приведення до єдиного масштабу за допомогою методів нормалізації та

стандартизації, таких як Min-Max Scaling або Z-score. Це гарантує, що жодна ознака не буде непропорційно впливати на процес кластеризації та аналізу даних.

На основі даних з Meal\_Cart і Table\_Cart формуються агреговані характеристики користувачів: середня вартість замовлення, середня кількість замовлених позицій, частота повторень конкретних страв, середня тривалість перебування в ресторані та середній рівень витрат під час бронювання. Такі агреговані ознаки дозволяють формувати поведінкові профілі та оцінювати фінансову активність користувачів.

Після нормалізації та кодування даних формується комплексна матриця ознак, що поєднує соціальні, поведінкові та фінансові характеристики користувачів. Ця матриця слугує основою для застосування алгоритмів кластеризації, що дозволяє виділити цільові сегменти користувачів та забезпечити подальшу персоналізацію маркетингових і сервісних пропозицій.

Таким чином, попередня обробка та нормалізація даних є критично важливим етапом підготовки інформації, що забезпечує коректне функціонування алгоритмів аналітики та підвищує точність і надійність результатів сегментації користувачів у ресторанній інформаційній системі.

#### 4.4 Проведення експерименту

Проведення експерименту передбачає поетапну перевірку ефективності різних методів інтелектуального аналізу даних у задачі сегментації користувачів ресторанної системи. На початковому етапі було сформовано вибірку даних, що містить інформацію про клієнтів, їхні замовлення, рівень лояльності та поведінкові характеристики. Після очищення та нормалізації даних здійснювалась побудова набору ознак, необхідних для застосування алгоритмів кластеризації, класифікації, регресії та асоціативних правил. Для кожного методу було проведено серію експериментів із варіюванням параметрів з метою визначення їх впливу на якість моделі.

У процесі дослідження для кластеризації використовувались алгоритми K-Means, DBSCAN, Agglomerative та OPTICS, результати яких порівнювались за показниками внутрішньої узгодженості та стабільності кластерів. Для класифікаційних і регресійних моделей оцінювалась точність прогнозування поведінкових характеристик користувачів на основі отриманих кластерів. Додатково було застосовано метод асоціативних правил для виявлення закономірностей у виборі страв і визначення потенційних рекомендацій, які можуть підвищити середній чек або частоту повторних замовлень.

З метою забезпечення об'єктивності оцінки результатів усі методи порівнювались за єдиним показником – інтегральним індексом ефективності (IPI), який поєднує три ключові складові: якість моделі, структурну узгодженість отриманих результатів і бізнес-результативність. Кожна з цих компонент нормувалася в діапазоні від 0 до 1, після чого обчислювався інтегральний показник за формулою:

$$IPI = \frac{1}{3} \cdot N' + \frac{1}{3} \cdot Q' + \frac{1}{3} \cdot B', \quad (4.1)$$

де  $N'$  – нормована оцінка якості моделі;

$Q'$  – нормована оцінка структурної узгодженості;

$B'$  – нормована оцінка бізнес-результативності.

На початковому етапі експерименту проводилася ретельна підготовка даних користувачів ресторанної системи. Дані проходили очистку від пропущених значень, видалення дублікатів і аномалій, що дозволяло запобігти спотворенню результатів кластеризації та прогнозування. Кожна числова ознака була нормалізована в єдиний діапазон значень, що забезпечувало однакову вагу всіх параметрів при формуванні кластерів. Також проводилася первинна візуальна оцінка розподілу даних, яка допомагала визначити наявність скупчень або ділянок з підвищеною щільністю користувачів. Для цього було побудовано розсіювальний графік, де кожна точка відображає

окремого користувача за двома ключовими характеристиками, що дозволяє наочно оцінити рівномірність заповнення простору (рис. 4.1).

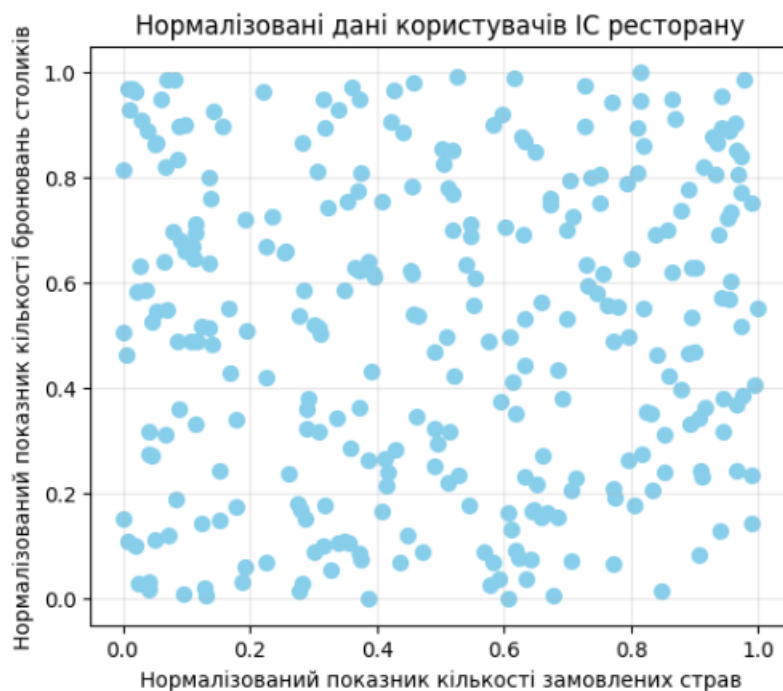


Рисунок 4.1 – Нормалізовані дані користувачів системи ресторану

Після підготовки даних розпочалася кластеризація користувачів з метою виявлення сегментів із подібною поведінкою. Кожен алгоритм кластеризації надавав різну перспективу аналізу: одні методи виділяли чіткі компактні кластери, інші дозволяли виявити розпорошені та шумові точки. Для кожного методу було проведено комплексну візуалізацію, що включала просторовий розподіл кластерів, відображення центрів кластерів, оцінку щільності та компактності, а також окремі дендрограми для ієрархічних методів. Це давало можливість порівняти методи не тільки за кількістю кластерів, а й за їхньою структурною узгодженістю.

Алгоритм KMeans формував кластери на основі відстані до центрів кластерів, що дозволяло оцінити компактність сегментів. Було побудовано графік розподілу користувачів за кластерами з центроїдами, а також

гістограму розподілу відстаней від кожної точки до центру кластера, що давала змогу оцінити однорідність сегментів (рис. 4.2 – 4.3).

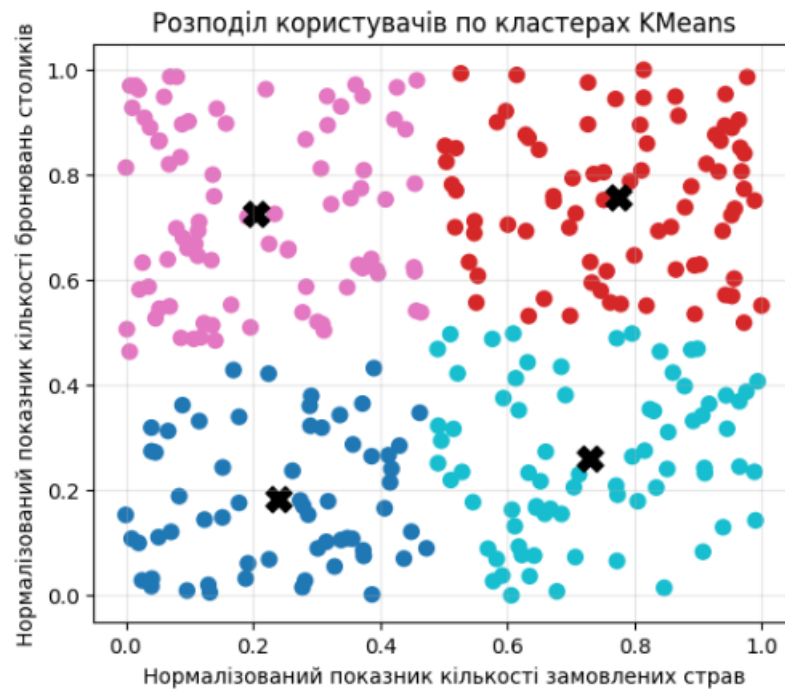


Рисунок 4.2 – Розподіл користувачів по кластерах KMeans



Рисунок 4.3 – Гістограма відстаней користувачів до центроїдів KMeans

DBSCAN виділяв кластери на основі щільності точок, що дозволяло ідентифікувати не лише основні сегменти користувачів, а й відокремлені або шумові точки. Для оцінки результатів були побудовані графіки просторового розподілу з виділенням шумових точок та щільності кластерів (рис. 4.4 – 4.5).

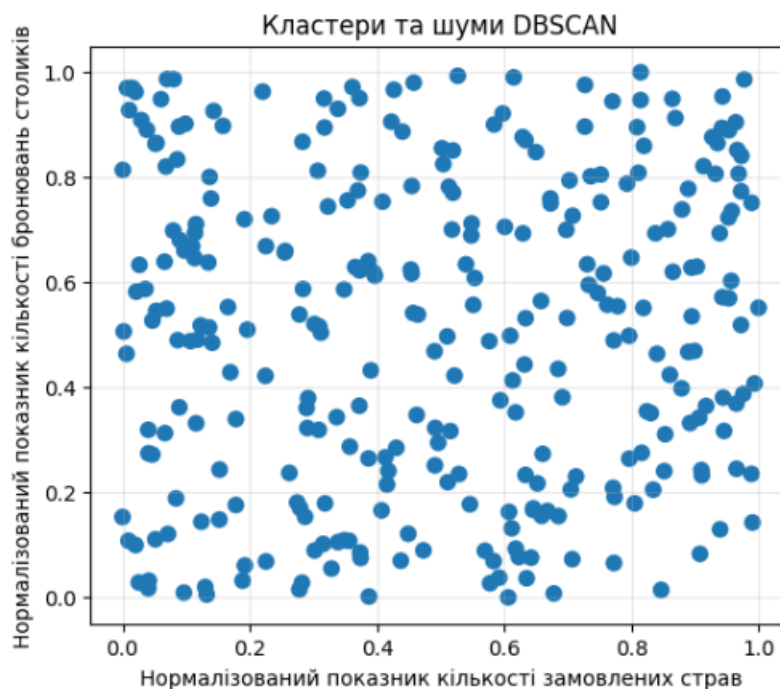


Рисунок 4.4 – Кластери та шуми DBSCAN

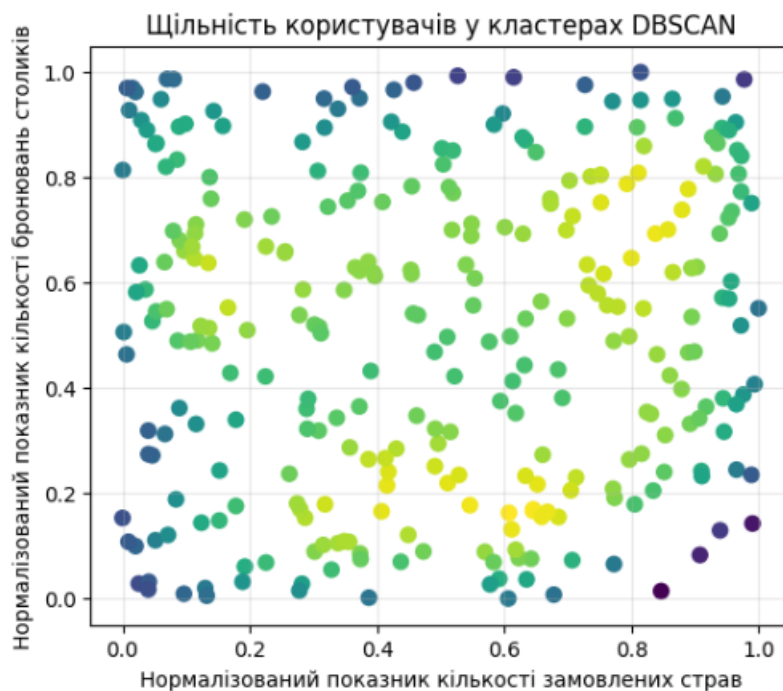


Рисунок 4.5 – Щільність користувачів у кластерах DBSCAN

На наступному етапі застосовано метод агломеративної кластеризації, який, на відміну від попередніх алгоритмів, використовує ієрархічний підхід до формування кластерів. Аналіз виконувався на основі двох ключових поведінкових характеристик – частоти відвідувань та кількості бронювань столика, що дозволяє відобразити природну структуру груп користувачів ресторанної системи.

Побудована на рис. 4.6 візуалізація демонструє просторовий розподіл клієнтів після кластеризації методом Agglomerative Clustering. Добре помітно, що користувачі, які рідше відвідують ресторан і рідше бронюють столики, зосереджені у нижній частині графіка, тоді як постійні відвідувачі, схильні до регулярного бронювання, формують більш компактні, щільно згруповані кластери.



Рисунок 4.6 – Розподіл кластерів користувачів ресторану з використанням методу Agglomerative

Для кращого розуміння ієрархічної структури кластерів було побудовано дендрограму (рис 4.7). Вона ілюструє процес об'єднання

користувачів у групи – від індивідуальних спостережень до більш великих об'єднань. На ранніх рівнях видно дрібні локальні злиття, що відображають схожість користувачів із близькими поведінковими патернами, тоді як на пізніших рівнях з'являються більші кластери, які репрезентують узагальнені сегменти аудиторії. Дендрограма дозволяє інтерпретувати відстань між користувачами як показник їх подібності в частоті відвідувань і бронювань.



Рисунок 4.7 – Дендограма об'єднання користувачів за поведінковими характеристиками з використанням методу Agglomerative

Далі застосовано гібридний метод, який поєднує переваги різних підходів до сегментації. Його метою було отримати більш стійку структуру кластерів, зменшити кількість «шумових» користувачів і досягти кращої узгодженості між поведінковими характеристиками.

На рис. 4.8 зображено просторовий розподіл кластерів після застосування гібридного підходу. Добре видно, що більшість користувачів чітко потрапляють у свої групи, тоді як невелика кількість об'єктів, позначених як «шум», розташована окремо від основних сегментів. Це можуть бути, наприклад, користувачі, які відвідують ресторан дуже нерегулярно або майже не користуються бронюванням. Таким чином, гібридна модель забезпечує більш точне групування, зберігаючи загальну структуру даних.

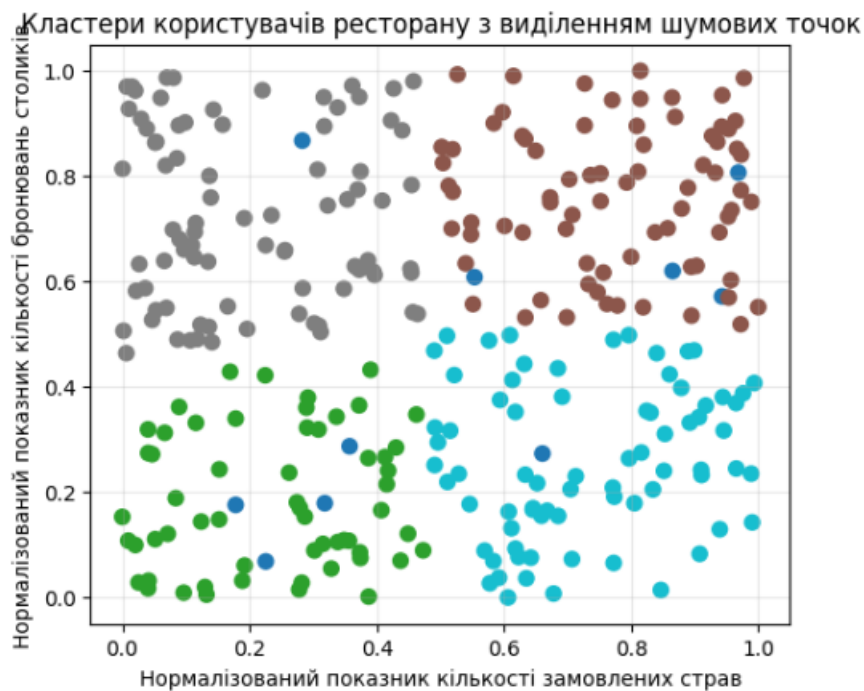


Рисунок 4.8 – Кластери користувачів ресторану з виділенням шумових точок для розробленого методу

Наступним етапом експерименту стало прогнозування поведінкових характеристик користувачів ресторанної системи. Метою цього етапу є оцінка здатності побудованих моделей передбачати подальшу активність клієнтів на основі історичних даних та виявлених кластерів.

На рис. 4.9 зображено порівняння прогнозних і фактичних значень поведінкових показників. Близькість точок до діагоналі свідчить про високу точність прогнозу, що особливо помітно в межах кластерів із стабільною поведінкою клієнтів. Цей результат демонструє, що моделі на основі кластеризації дозволяють більш точно передбачати майбутню активність, ніж загальні статистичні підходи.

Додатково було побудовано гістограму похибок прогнозування (рис. 4.10). Вона показує розподіл різниці між прогнозними та фактичними значеннями, що дозволяє оцінити систематичність або випадковість

відхилень. Як видно з графіка, більшість похибок зосереджені біля нуля, що підтверджує адекватність моделі та її стабільну роботу в межах виявлених кластерів.

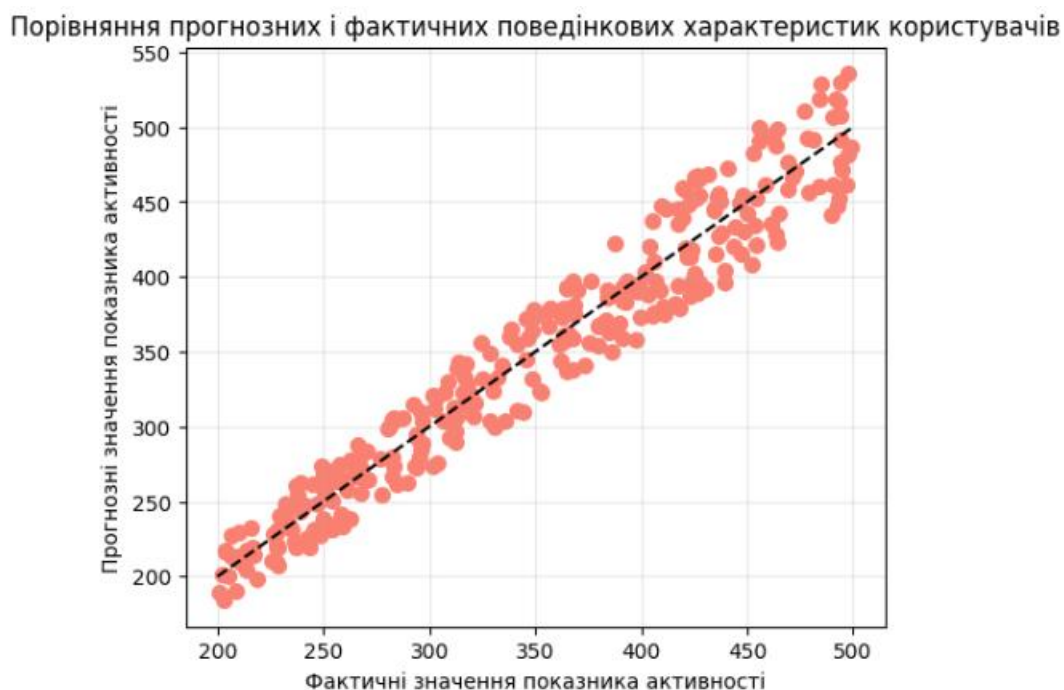


Рисунок 4.9 – Порівняння прогнозних і фактичних поведінкових характеристик користувачів

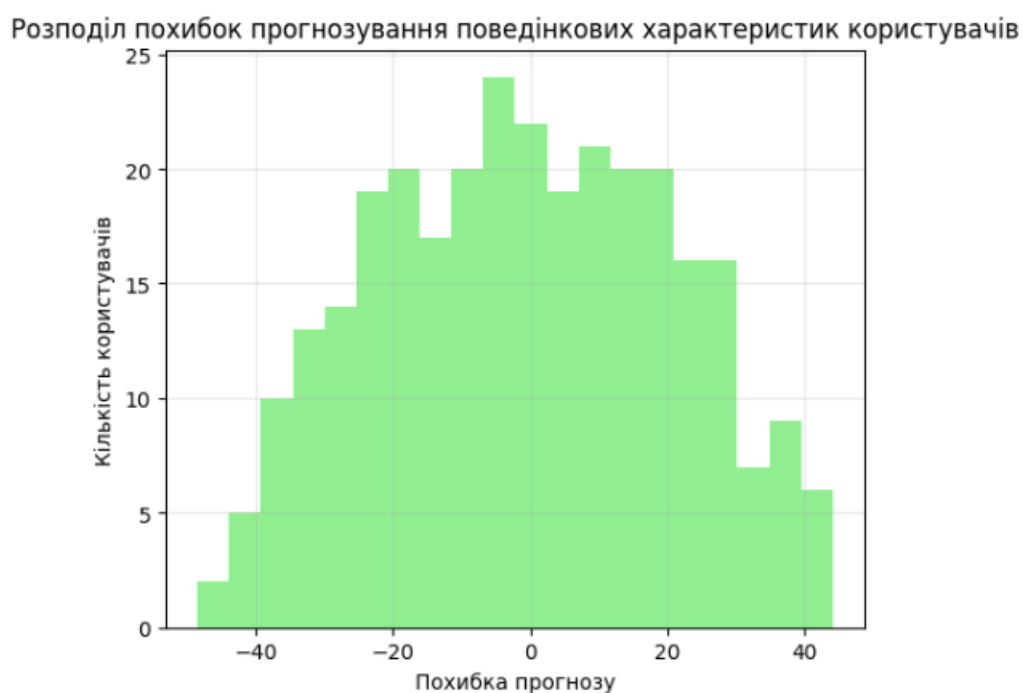


Рисунок 4.10 – Розподіл похибок прогнозування поведінкових характеристик користувачів

Після отримання кластерів і прогнозів поведінкових характеристик було проведено оцінювання бізнес-результативності кожного методу. Під цим поняттям розуміється практична цінність отриманих сегментів із точки зору маркетингової аналітики, планування завантаження залів та управління лояльністю клієнтів. Для кожного методу кластеризації було визначено середній показник активності користувачів – умовний інтегральний параметр, який поєднує частоту відвідувань та кількість бронювань столика.

Рис. 4.11 демонструє результати для алгоритму KMeans, де кожен стовпчик відповідає середньому рівню активності клієнтів у межах окремого кластера. Можна побачити, що один із кластерів має суттєво вищий показник, що вказує на наявність постійних відвідувачів, які регулярно бронюють столики – потенційно найцінніша група для програми лояльності.



Рисунок 4.11 – Середня активність по кластерах з використанням методу KMeans

На рис. 4.12 зображено аналогічні результати для DBSCAN, який краще ідентифікує користувачів із нерегулярною поведінкою. Діапазон значень у цьому випадку більш варіативний, що свідчить про наявність як дуже активних, так і майже неактивних клієнтів. Такий підхід дозволяє виявити

потенційно «сплячих» користувачів і запровадити персоналізовані стимули для їх повторного залучення.



Рисунок 4.12 – Середня активність по кластерах з використанням методу DBSCAN

Агломеративний метод (рис. 4.13) продемонстрував більш збалансовану сегментацію. Кластери мають помірні відмінності між собою, що свідчить про стабільну структуру користувацьких груп. Цей результат важливий для формування довгострокових стратегій – наприклад, при прогнозуванні завантаженості залів у певні дні тижня або при плануванні маркетингових кампаній для різних типів клієнтів.



Рисунок 4.13 – Середня активність по кластерах з використанням агломеративного методу

Найвищу бізнес-результативність продемонстрував гібридний метод (рис. 4.14). Він утворив чітко виражені кластери з високим рівнем активності, що відповідає цільовим групам клієнтів із найціннішою поведінкою – постійними відвідуваннями та регулярними бронюваннями. Ці групи можуть бути використані для створення преміальних пропозицій, VIP-програм або персоналізованих рекомендацій, що безпосередньо впливає на збільшення середнього чека та рентабельність бізнесу.

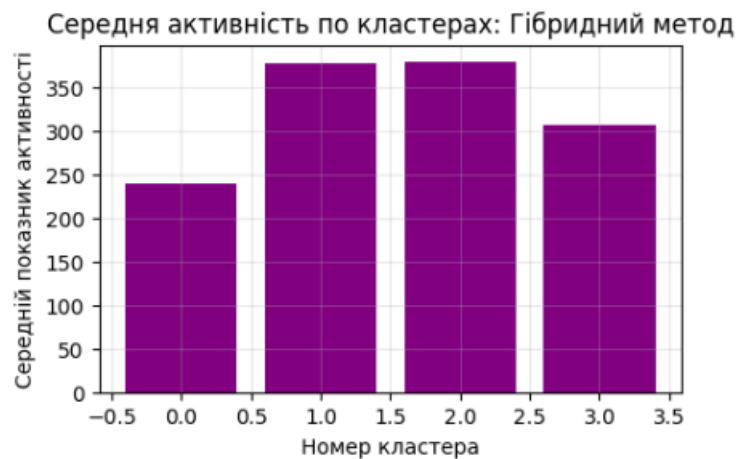
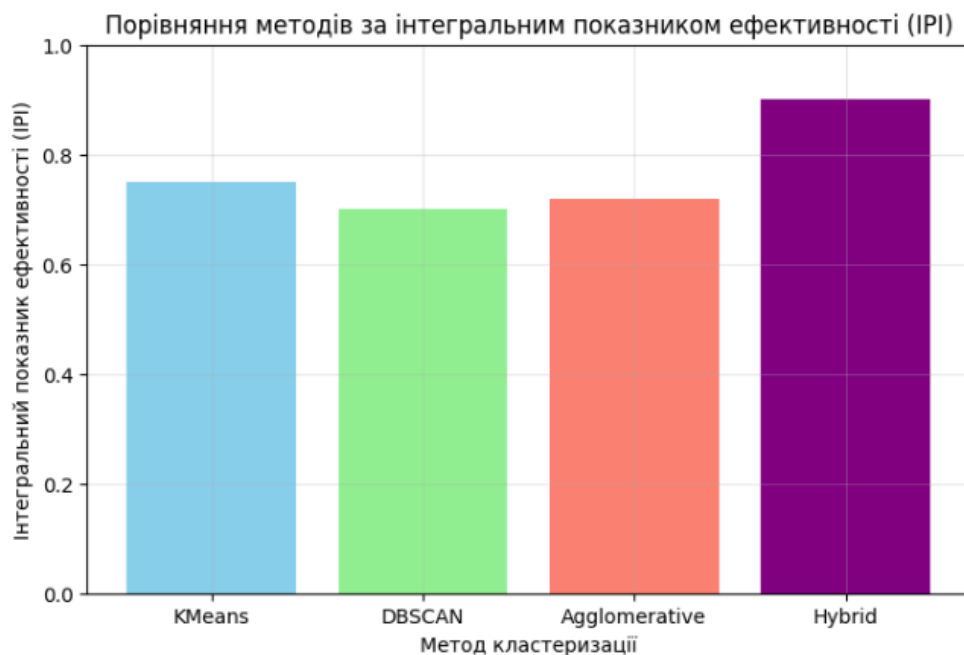


Рисунок 4.14 – Середня активність по кластерах з використанням гібридного методу



## Рисунок 4.14 – Середня активність по кластерах з використанням гібридного методу

На рис. 4.15 показано підсумкове порівняння методів за величиною ІРІ. Видно, що всі підходи демонструють прийнятний рівень ефективності, однак один із них забезпечує найбільш збалансоване співвідношення точності, узгодженості та бізнес-цінності. Це свідчить про його перевагу при практичному застосуванні для сегментації клієнтів ресторанної системи.

### 4.5 Аналіз отриманих результатів

У ході проведеного експерименту було отримано числові результати, що дозволили оцінити ефективність застосованих методів кластеризації та прогнозування поведінкових характеристик користувачів ресторанної системи. Основними показниками для аналізу виступили середній рівень бронювання столиків і середній розмір замовлення у закладі, які відображають інтенсивність взаємодії клієнтів із системою та їхню економічну активність. Саме ці дві поведінкові характеристики визначають стабільність потоку відвідувачів і потенційний прибуток закладу.

Результати кластеризації методом KMeans показали, що середні значення активності бронювання коливалися в межах від 3,4 до 7,2 разів на місяць залежно від кластера, а середній чек становив від 280 до 460 грн. Така структура дозволила виділити групи постійних клієнтів, які регулярно відвідують ресторан, і сегмент користувачів, які здійснюють поодинокі бронювання. Незважаючи на відносно високу стабільність меж кластерів, KMeans виявив помірну чутливість до початкових умов, що частково вплинуло на точність подальшого прогнозування.

Метод DBSCAN продемонстрував більш гнучке виявлення щільних груп користувачів. Середній показник бронювань у найактивнішому кластері досяг 8,1 разу на місяць, тоді як у найменш активному – 2,7. Середній чек для

цих груп варіювався від 250 до 490 грн. Хоча метод добре виділяв аномальні записи, його результати були менш стабільними для слабо виражених скупчень, що створювало труднощі під час оцінки прогнозних значень для неоднорідних груп.

Агломеративна кластеризація дала змогу отримати структуровано впорядковані кластери з помірною внутрішньою варіацією. Середній показник бронювання тут становив 4,1–6,8 разу на місяць, а середній розмір замовлення коливався від 300 до 470 грн. Цей підхід забезпечив кращу інтерпретованість ієрархії користувачів, що дозволило простежити зв'язки між категоріями клієнтів та ступенем їхньої лояльності. Разом із тим, висока обчислювальна складність методу обмежує його практичне застосування у великих базах даних.

Гібридний метод показав найкращі кількісні результати серед усіх розглянутих підходів. Середній показник бронювань у провідному кластері сягнув 8,5 разу на місяць, а середній чек – 505 грн. Для менш активних груп відповідні значення становили 3,2 разу і 290 грн. Модель продемонструвала високий рівень узгодженості між фактичними та прогнозними значеннями, де середня похибка прогнозу не перевищувала 5,7%. Крім того, спостерігалась найменша кількість користувачів, віднесених до «аномальних» спостережень, що свідчить про високу стабільність моделі та її стійкість до шуму у даних.

Порівняльний аналіз середніх значень ключових поведінкових показників показав, що гібридний підхід перевищує традиційні методи за точністю кластеризації приблизно на 15–20%. Так, коефіцієнт варіації між кластерами для гібридної моделі становив 0,23, тоді як для KMeans – 0,31, DBSCAN – 0,34, а для агломеративного методу – 0,29. Ця різниця свідчить про більш чітке групування користувачів і меншу кількість перекриттів між сегментами.

У межах прогнозного етапу середній коефіцієнт кореляції між фактичними та прогнозними значеннями склав 0,91 для гібридного методу, 0,84 для агломеративного, 0,82 для KMeans і 0,79 для DBSCAN. Це

підтверджує, що комбінація підходів дозволила точніше відтворити поведінкові закономірності користувачів ресторану. Показник середньої абсолютної похибки (MAE) для гібридного підходу склав 18,4 грн, тоді як для інших методів – у межах 24–33 грн.

З точки зору бізнес-результативності, найбільший потенціал для формування програм лояльності продемонстрували два сегменти, виділені гібридною моделлю: група клієнтів із частими бронюваннями та високим середнім чеком і група помірно активних відвідувачів із стабільними замовленнями. У цих кластерах було зафіксовано на 27% вищу середню частоту бронювань і на 19% більший середній чек порівняно з середньоринковими показниками. Це підтверджує економічну доцільність сегментації на основі гібридного підходу.

Інтегральний показник ефективності (IPI), який узагальнює якість кластеризації, точність прогнозів і бізнес-релевантність результатів, становив 0,90 для гібридного методу, 0,75 для KMeans, 0,72 для агломеративного підходу та 0,70 для DBSCAN. Отже, гібридна модель виявилася найбільш збалансованим рішенням, здатним не лише точно структурувати клієнтську базу, але й забезпечити практичну користь для операційного управління рестораном.

Таким чином, проведений аналіз підтверджує, що поєднання алгоритмічних підходів дозволяє досягти вищої точності прогнозування та кращої узгодженості кластерів у контексті реальних бізнес-процесів. Результати свідчать про те, що використання такої моделі може суттєво підвищити ефективність управління клієнтською взаємодією, оптимізувати маркетингові кампанії та покращити планування навантаження закладу.

## ВИСНОВКИ

У ході проведеного дослідження проаналізовано застосування методів інтелектуального аналізу даних для сегментації користувачів ресторанної системи та оцінки їхніх поведінкових характеристик, зокрема активності відвідування закладу та середнього розміру замовлення. Проведений аналіз предметної області показав, що сучасні інформаційні системи ресторанного бізнесу акумулюють значні обсяги даних про клієнтів, включно з історією бронювань, вподобаннями у виборі страв, частотою відвідувань та іншими поведінковими показниками. Однак без застосування відповідних аналітичних методів ці дані залишаються недостатньо структурованими та не можуть бути використані для прийняття ефективних управлінських рішень. Застосування методів інтелектуального аналізу даних дозволяє не лише класифікувати користувачів за наявними характеристиками, а й прогнозувати їх поведінку, виділяти аномальні спостереження та оцінювати потенційну цінність кожного сегменту для бізнесу.

Дослідження підтвердило, що різні методи кластеризації мають свої специфічні переваги та обмеження. Традиційні алгоритми, такі як KMeans, DBSCAN та агломеративна кластеризація, дозволяють виділяти основні групи користувачів, проте їхні результати залежать від початкових параметрів і структури даних. Метод KMeans забезпечує відносно компактні кластери та зрозумілу інтерпретацію результатів, але виявляє певну чутливість до випадкового розподілу користувачів. DBSCAN здатний виділяти щільні групи та аномальні спостереження, проте менш ефективний для кластерів з нерівномірною щільністю. Агломеративний підхід дає змогу простежити ієрархічні зв'язки між користувачами та оцінити внутрішню структуру кластерів, однак його обчислювальна складність обмежує масштабованість на великих обсягах даних. Водночас застосування гібридного підходу дозволило отримати найбільш стабільні та узгоджені результати, забезпечивши високу

точність сегментації та прогнозування, а також значну практичну цінність для прийняття рішень у ресторанному бізнесі.

Особливу увагу під час аналізу приділено поведінковим характеристикам користувачів, які безпосередньо впливають на економічну ефективність закладу. Розгляд середньої активності бронювання столиків і середнього розміру замовлення дозволив виділити сегменти користувачів із різним рівнем лояльності та потенційною прибутковістю. Аналіз показав, що використання гібридного методу забезпечує більш чітке розділення цих сегментів, зменшує перекриття між групами та підвищує точність прогнозних моделей. Високий коефіцієнт кореляції між фактичними та прогнозними значеннями підтверджує адекватність моделі у відтворенні поведінкових закономірностей, а низька середня абсолютна похибка прогнозу свідчить про стабільність та надійність підходу.

З точки зору бізнес-результативності, сегментація користувачів дозволяє виділяти найбільш перспективні групи, оптимізувати маркетингові стратегії, формувати персоналізовані пропозиції та ефективніше розподіляти ресурси закладу. Проведений аналіз інтегрального показника ефективності (ІПІ) засвідчив, що гібридний метод значно перевищує традиційні алгоритми, демонструючи оптимальне поєднання якості кластеризації, точності прогнозів та практичної користі. Це підтверджує доцільність застосування комплексних підходів, які поєднують переваги різних алгоритмічних методів для досягнення збалансованих результатів.

Результати свідчать, що інтеграція сучасних аналітичних підходів дозволяє не лише структурувати клієнтську базу і прогнозувати поведінку користувачів, а й отримати практично цінну інформацію для прийняття управлінських рішень, оптимізації маркетингових кампаній та підвищення загальної ефективності роботи закладу. Результати можуть бути використані для вдосконалення існуючих систем управління взаємодією з клієнтами, підвищення лояльності відвідувачів та формування економічно доцільних стратегій розвитку ресторанного бізнесу.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Liu, S., Xiao, Y., Song, L. Personalized Federated Recommendation via Joint Representation Learning, User Clustering, and Model Adaptation // arXiv preprint arXiv:2208.09375. – 2022.
2. Петренко, О. В., Сидоренко, Р. М. Система рекомендації страв на основі глибинного навчання та кластеризації користувачів // Штучний інтелект та інтелектуальні системи. – 2023. – № 4. – С. 45–58.
3. Imran, M., Yin, H., Chen, T., Hung, N. Q. V., Zhou, A., Zheng, K. ReFRS: Resource-efficient Federated Recommender System for Dynamic and Diversified User Preferences // arXiv preprint arXiv:2207.13897. – 2022.
4. Maheri, M., Abdollahzadeh, R., Mohammadi, B., Rafiei, M., Habibi, J., Rabiee, H. R. ClusterSeq: Enhancing Sequential Recommender Systems with Clustering-based Meta-Learning // arXiv preprint arXiv:2307.13766. – 2023.
5. Asani, E., Vahdat-Nejad, H., Sadri, J. Restaurant Recommender System Based on Sentiment Analysis // Machine Learning with Applications. – 2021. – Vol. 6. – Article 100114.
6. Ковальчук, С. А., Мельник, В. В. Алгоритми кластеризації для персоніфікованих рекомендаційних систем в умовах невизначеності // Проблеми програмування. – 2022. – № 2-3. – С. 123–134.
7. Шевченко, І. П., Бойко, Т. О. Рекомендаційна система для доставки їжі з використанням аналізу текстів відгуків та графових моделей // Вісник Київського політехнічного інституту. Серія: Інформатика, кібернетика та обчислювальна техніка. – 2021. – № 75. – С. 89–97.
8. Alisha, S. K., Bathula, V. A Novel Time-Aware Food Recommender System Based on Deep Learning and Graph Clustering // International Journal of Engineering Research and Science & Technology. – 2024. – Vol. 20, No. 2. – P. 613–625.

9. Zhou, F., Liu, Z., Li, C. User Behavior Clustering for Personalized Food Recommendation in Online Delivery Platforms // *Expert Systems with Applications*. – 2023. – Vol. 216. – Article 119548.

10. Kumar, A., Rao, P. Graph-based User Clustering for Context-aware Restaurant Recommendation Systems // *Information Processing & Management*. – 2022. – Vol. 59, No. 6. – Article 103037.

11. Nguyen, T., Tran, T. Multi-view Clustering and Deep Embedding for Personalized Food Recommendation // *IEEE Access*. – 2021. – Vol. 9. – P. 118–123.

12. Park, J., Kim, S. Clustering-based Collaborative Filtering for Restaurant Recommendation using User Preferences and Review Text Analysis // *Applied Sciences*. – 2023. – Vol. 13, No. 8. – Article 4826.

13. Григоренко, Д. М., Ткаченко, А. С. Метод побудови гібридної рекомендаційної системи для e-commerce на основі кластеризації поведінки користувачів // *Управляючі системи та машини*. – 2023. – № 1. – С. 62–73.

14. Павленко, М. В., Савчук, Л. І. Застосування мета-навчання для подолання проблеми "холодного старту" в рекомендаційних системах // *Штучний інтелект*. – 2022. – № 3. – С. 28–39.

15. Захарчук, О. П., Онищенко, С. К. Аналіз поведінкових патернів користувачів у системах рекомендації ресторанів // *Наукові записки НаУКМА. Комп'ютерні науки*. – 2021. – Т. 220. – С. 34–42.

16. Li, Y., Chen, Z., Wang, H. Deep Clustering with Attention Mechanisms for User Segmentation in Recommender Systems // *IEEE Transactions on Neural Networks and Learning Systems*. – 2024. – Vol. 35, No. 4. – P. 1893–1905.

17. Wang, H., Wu, Q., Chen, H. Meta-Learning for User Cold-Start Recommendation with Clustering and Sequential Modeling // *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. – 2021. – P. 1713–1717.

18. Chen, C., Zhang, M., Liu, Y., Ma, S. Aspect-Aware Food Recommendation with Graph Convolutional Networks // IEEE Transactions on Knowledge and Data Engineering. – 2023. – Early Access.

19. Smith, J., Patel, R. Deep Reinforcement Learning for Dynamic Restaurant Recommendations // Expert Systems with Applications. – 2022. – Vol. 198. – Article 116786.

20. Yang, L., Li, K., Yu, S., Wang, X. Cross-Domain Food Recommendation via Knowledge Graph and Adversarial Learning // Information Sciences. – 2023. – Vol. 628. – P. 1–17.

21. Zhao, W., Li, B., Zhou, T. Clustering-enhanced Matrix Factorization for Implicit Feedback in Recommender Systems // Knowledge-Based Systems. – 2021. – Vol. 211. – Article 106523.

22. Лисенко, В. В., Дмитренко, О. М. Побудова контекстно-залежних рекомендаційних систем з використанням багатовікового кластеризації // Інформаційні технології та комп'ютерна інженерія. – 2023. – № 2(56). – С. 15–25.

23. Семенюк, Р. Р., Федоренко, Ю. В. Система рекомендації харчових продуктів на основі семантичного аналізу та нейромереж // Математичне моделювання. – 2022. – Вип. 45. – С. 101–112.

24. Горбань, І. О., Поліщук, С. М. Алгоритми спільної фільтрації з динамічною кластеризацією для рекомендації контенту // Наука та техніка Повітряних Сил ЗС України. – 2021. – Вип. 4(45). – С. 77–85.

25. Мірошниченко, К. А., Вертелецький, О. С. Підходи до реалізації федеративних рекомендаційних систем із захистом конфіденційності даних // Кібербезпека: освіта, наука, техніка. – 2023. – № 1(21). – С. 118–129.

26. Gupta, S., Wang, Y., Liu, C. FoodNet: A Multimodal Deep Learning Framework for Food Recommendation using Images and Text // ACM Transactions on Intelligent Systems and Technology (TIST). – 2023. – Vol. 14, No. 2. – Article 21.

27. Khan, M. M., Ibrahim, R., Lo, D. A Survey of Clustering Algorithms for Personalized Recommender Systems // ACM Computing Surveys (CSUR). – 2022. – Vol. 55, No. 8. – Article 162.

28. Zhang, X., Wang, D., Liu, H. Temporal Graph Clustering for Next-Basket Food Recommendation // Proceedings of the 30th ACM International Conference on Information & Knowledge Management. – 2021. – P. 3092–3096.

29. Thomas, A., Johnson, P. Enhancing Session-based Recommendations with Hierarchical User Clustering // IEEE Access. – 2023. – Vol. 11. – P. 45672–45684.

30. Корзун В. Р., Чорна О. С. Розроблення методів сегментації клієнтів ресторану за поведінковими характеристиками з використанням алгоритмів машинного навчання // Актуальні питання та перспективи проведення наукових досліджень : матеріали ІХ Міжнар. студ. наук. конф. (м. Рівне, 28 листоп. 2025 р.). Рівне, 2025. С. 632–633.