

**Київський національний торговельно-економічний
університет**

Кафедра кібернетики та системного аналізу

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

**«Моделювання процесів формування споживчих
пріоритетів в електронній торгівлі»**

Студента 2 курсу, 1м групи,

спеціальності
051 «Економіка»

освітньої програми
«Економічна кібернетика»

Науковий керівник
доктор фізико-математичних наук,
професор

Гарант освітньої програми
доктор фізико-математичних наук,
професор

Барильченка
Василя
Васильовича

Пурський Олег
Іванович

Гамалій
Володимир
Федорович

Київ 2018

Київський національний торговельно-економічний університет

Факультет обліку, аудиту та інформаційних систем
Кафедра кібернетики та системного аналізу
Спеціальність 051 «Економіка»
Спеціалізація «Економічна кібернетика»

Зав. кафедри _____ **Затверджую**
Роскладка А. А.
«05» грудня 2017р.

**Завдання
на випускн кваліфікаційну роботу (проект) студента**

Барильченка Василя Васильовича

(прізвище, ім'я, по батькові)

1. Тема випускної кваліфікаційної роботи (проекту):

«Моделювання процесів формування споживчих пріоритетів в електронній торгівлі»

Затверджена наказом ректора від «29» листопада 2017 р. № 4058

2. Строк здачі студентом закінченої роботи 15 листопада 2018 року

3. Цільова установка та вихідні дані до роботи

Мета роботи: є розробка моделей, методів та програмно-інструментальних засобів визначення споживчих уподобань і інформаційної бази підприємства електронної торгівлі з використанням інтелектуального аналізу даних.

Об'єктом дослідження є процеси формування споживчих пріоритетів в електронній торгівлі.

Предметом дослідження є методи, моделі та інформаційна технологія визначення споживчих пріоритетів в електронній торгівлі.

4. Перелік графічного матеріалу:

5. Консультанти по роботі (проекту) із зазначенням розділів, які консультують:

Розділ	Консультант (прізвище, ініціали)	Підпис, дата	
		Завдання видав	Завдання прийняв
1	Пурський О.І.	05.12.2017 р.	05.12.2017 р.
2	Пурський О.І.	05.12.2017 р.	05.12.2017 р.
3	Пурський О.І.	05.12.2017 р.	05.12.2017 р.

6. Зміст випускної кваліфікаційної роботи (проекту) (перелік питань за кожним розділом):

ВСТУП

РОЗДІЛ 1. СУЧАСНИЙ СТАН І ПЕРСПЕКТИВИ РОЗВИТКУ СИСТЕМ ЕЛЕКТРОННОЇ ТОРГІВЛІ

1.1. Електронна торгівля: проблематика, сутність і поняття

1.2. Теоретичні методи в дослідженні механізмів функціонування систем електронної торгівлі

1.3. Специфіка електронної торгівлі в Україні: аналіз і перспективи розвитку

Висновки до розділу 1

РОЗДІЛ 2. МЕТОДИКА ВИЗНАЧЕННЯ СПОЖИВЧИХ ПРІОРИТЕТІВ В ЕЛЕКТРОННІЙ ТОРГІВЛІ НА ОСНОВІ ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

- 2.1. Інтелектуальний аналіз даних як частина методики визначення споживчих пріоритетів в електронній торгівлі
- 2.2. Збір, підготовка і аналіз інформації про споживачів та їх активності на ринку електронної торгівлі
- 2.3. Методи кластеризації споживачів ринку електронної торгівлі
- 2.4. Обґрунтування вибору методів перевірки і візуалізації даних кластеризації
- 2.5. Визначення методу розпізнавання даних кластеризації споживачів електронної торгівлі

Висновки до розділу 2

РОЗДІЛ 3. ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЗНАЧЕННЯ СПОЖИВАЧИХ ПРІОРИТЕТІВ В ЕЛЕКТРОННІЙ ТОРГІВЛІ

- 3.1. Специфіка програмно-апаратних засобів та методів для розробки інформаційної технології
- 3.2. Інформаційно-логічна модель процесу визначення споживчих пріоритетів
- 3.3. Побудова інформаційної бази системи і програмна реалізація

Висновки до розділу 3

ВИСНОВКИ

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

ДОДАТКИ

Додаток А

Додаток Б

Додаток В

7. Календарний план виконання роботи (проекту)

№ пор.	Назва етапів випускної кваліфікаційної роботи	Строк виконання етапів роботи (проекту)	
		За планом	Фактично
1	2	3	4
1	Вибір теми випускної кваліфікаційної роботи	01.10.2017	01.10.2017
2	Розробка та затвердження завдання на випускну кваліфікаційну роботу	05.12.2017	05.12.2017
3	Вступ	01.04.2018	
4	Розділ 1. Сучасний стан і перспективи розвитку систем електронної торгівлі	01.05.2018	
5	Розділ 2. Методика визначення споживчих пріоритетів в електронній торгівлі на основі застосування методів інтелектуального аналізу даних	20.06.2018	
6	Підготовка статті у збірник наукових статей магістрів	15.09.2018	
7	Розділ 3. Інформаційна технологія визначення споживачих пріоритетів в електронній торгівлі	01.10.2018	
8	Висновки	01.11.2018	
10	Здача випускної кваліфікаційної роботи	15.11.2018	

	на кафедрі науковому керівнику		
11	Попередній захист кваліфікаційної роботи	22.11.2018	
12	Виправлення зауважень, рецензування випускної кваліфікаційної роботи	25.11.2018	
13	Представлення готової випускної кваліфікаційної роботи на кафедрі	28.11.2018	
14	Публічний захист кваліфікаційної роботи	За розкладом роботи ЕК	

8. Дата видачі завдання «05» грудня 2017 р.

9. Керівник випускного кваліфікаційної роботи (проекту)

Пурський О.І.

(прізвище, ініціали, підпис)

10. Гарант освітньої програми

Гамалій В. Ф.

(прізвище, ініціали, підпис)

11. Завдання прийняв до виконання студент-дипломник

Барильченко В.В.

(прізвище, ініціали, підпис)

12. Відгук керівника випускної кваліфікаційної роботи (проекту)

Керівник випускної кваліфікаційної роботи (проекту)

(підпис, дата)

13. Висновок про випускню кваліфікаційну роботу (проект)

Випускна кваліфікаційна робота (проект) студента _____
(прізвище, ініціали)
може бути допущена до захисту в екзаменаційній комісії.

Гарант освітньої програми _____ Гамалій В. Ф.
(підпис, прізвище, ініціали)

Завідувач кафедри _____ Роскладка А.А.
(підпис, прізвище, ініціали)

« _____ » 2018 р.

Анотація

У випускній кваліфікаційній роботі представлено результати теоретичних і прикладних досліджень, що полягають у розробці інформаційної технології визначення споживчих пріоритетів в електронній торгівлі. Проаналізовано типи та моделі інформаційних систем визначення споживчих пріоритетів, визначено методи та алгоритми побудови інформаційних систем на основі машинного навчання, розглянуто можливості використання алгоритмів машинного навчання для побудови різних типів інформаційних систем, досліджено інструментальні засоби для розробки інформаційних системи в електронній комерції.

Ключові слова: інформаційна технологія визначення споживчих пріоритетів, електронна торгівля, машинне навчання, алгоритми машинного навчання, методи машинного навчання.

Anotation

The graduation qualification work presents the results of theoretical and applied research, consisting in the development of information technology for determining consumer priorities in electronic trade. The types and models of information systems for determining consumer priorities are analyzed, methods and algorithms of building systems on the basis of machine learning are analyzed, the possibilities of using machine learning algorithms for construction of different types of information systems are considered, tools for developing the information system in electronic commerce are explored.

Key words: information technology for determining consumer priorities, electronic trade, machine learning, machine learning algorithms, machine learning methods.

ЗМІСТ

ВСТУП.....	11
РОЗДІЛ 1. СУЧАСНИЙ СТАН І ПЕРСПЕКТИВИ РОЗВИТКУ СИСТЕМ ЕЛЕКТРОННОЇ ТОРГІВЛІ.....	16
1.1. Електронна торгівля: проблематика, сутність і поняття.....	16
1.2. Теоретичні методи в дослідженні механізмів функціонування систем електронної торгівлі.....	29
1.3. Специфіка електронної торгівлі в Україні: аналіз і перспективи розвитку.....	33
Висновки до розділу 1.....	43
РОЗДІЛ 2. МЕТОДИКА ВИЗНАЧЕННЯ СПОЖИВЧИХ ПРІОРИТЕТІВ В ЕЛЕКТРОННІЙ ТОРГІВЛІ НА ОСНОВІ ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.....	45
2.1. Інтелектуальний аналіз даних як частина методики визначення споживчих пріоритетів в електронній торгівлі.....	45
2.2. Збір, підготовка і аналіз інформації про споживачів та їх активності на ринку електронної торгівлі.....	50
2.3. Методи кластеризації споживачів ринку електронної торгівлі.....	57
2.4. Обґрунтування вибору методів перевірки і візуалізації даних кластеризації.....	71
2.5. Визначення методу розпізнавання даних кластеризації споживачів електронної торгівлі.....	79
Висновки до розділу 2.....	91
РОЗДІЛ 3. ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЗНАЧЕННЯ СПОЖИВАЧИХ ПРІОРИТЕТІВ В ЕЛЕКТРОННІЙ ТОРГІВЛІ.....	93
3.1. Специфіка програмно-апаратних засобів та методів для розробки інформаційної технології.....	93

3.2. Інформаційно-логічна модель процесу визначення споживчих пріоритетів.....	96
3.3. Побудова інформаційної бази системи і програмна реалізація.....	98
Висновки до розділу 3.....	110
ВИСНОВКИ.....	111
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	113
ДОДАТКИ.....	121
Додаток А.....	121
Додаток Б.....	124
Додаток В.....	129

ВСТУП

Актуальність теми дослідження. Сучасні світові економічні умови, глобалізація економіки, прискорення процесів розвитку ринків, інформаційних технологій, соціально-політичні фактори вимагають від торговельних підприємств нових підходів до формування споживчого попиту та пропозиції, виробленню адекватних методичних рішень та інструментальних засобів в сфері управління клієнтською базою і уподобаннями клієнтів, особливо це стосується таких нових форм ведення торговельної діяльності, як електронна торгівля. Однією з найбільш важливих умов забезпечення конкурентоспроможності підприємств електронної торгівлі є побудова ефективної концепції взаємин з клієнтами, яка здатна забезпечити врахування споживчих уподобань і підвищення загального рівня привабливості торговельного підприємства на електронному ринку товарів і послуг. Проблематика створення такої концепції і методики врахування споживчих пріоритетів в електронній торгівлі ще недостатньо досліджена, у зв'язку з низьким рівнем організації процесу комунікації між суб'єктами електронного ринку товарів і послуг, відсутністю адекватної і своєчасної відповіді на виникаючі потреби клієнтів, недостатньою вивченістю клієнтських уподобань, низьким рівнем впровадження автоматизованих систем управління взаємовідносинами з клієнтами (CRM-системи).

На практиці, досить часто, для визначення споживчих уподобань використовуються методи інтелектуального аналізу даних (Data Mining), за допомогою яких визначаються не явні, латентні закономірності, які знаходяться в даних моніторингу клієнтської активності, що складаються з набору споживчих і особистих характеристик клієнтів. В той же час, варто зазначити, що застосування методів інтелектуального аналізу даних (ІАД)

в процесі аналізу споживачів електронного ринку товарів і послуг носить неповний і фрагментарний характер: кластеризація без повної класифікації, або із уже відомою кількістю кластерів, загалом не використовується механізм пошуку оптимальних алгоритмів розбивки кластерів за споживчими пріоритетами, не застосовуються методи скорочення простору ознак, відсутнє розширення бази знань в процесі класифікації і т.п. При цьому застосування результатів ІАД недостатньо формалізоване і автоматизоване, що негативно позначається на своєчасності і оперативності прийняття управлінських рішень та підвищенню лояльності і ефективності роботи з клієнтами. Необхідний рівень формалізації і автоматизації може забезпечити імплементація системи управління бізнес-правилами (BRMS - Business Rules Management System) у рамках концепції прийняття управлінських рішень (EDM - Enterprise Decision Management) за рахунок її використання як частини методології CRM, впровадженої на типовому підприємстві електронної торгівлі.

Дослідженням економічних процесів в електронній торгівлі займалася значна кількість вітчизняних і закордонних науковців, зокрема: Геєць В.М. [1], Рамазанов С.К. [2], Вітлінський В.В. [3], Лисенко Ю.Г. [4], Базилевич В.Д. [5], Мазаракі А.А. [6], Порохня В.М. [7], Ситник В.Ф. [8], Хейг М. [9], Копитко Б.І. [10], Плєскач В.Л. [11], Дюбанов О.С. [12], Лаудон К.С. [13], Козьє Д. [14], Хартман А. [15]. Відзначаючи вагомий і важливий внесок згаданих науковців у дослідження економічних аспектів електронних засобів ведення торгівлі, варто зазначити, що питання функціонування електронного торговельного ринку залишається актуальними в силу появи нових механізмів і засобів електронної торгівлі.

Все це обумовлює необхідність удосконалення існуючих механізмів CRM з урахуванням споживчих пріоритетів на основі розробки комплексного методу і моделі, що представляють собою засоби

ефективного управління та аналізу клієнтської бази, ядром якої повинна стати наявність інформаційної системи підприємства електронної торгівлі на основі сучасних програмно-апаратних засобів і методів ІАД, що і визначає **актуальність** обраної теми випускної кваліфікаційної роботи.

Мета і завдання дослідження. Метою даного дослідження є розробка моделей, методів та програмно-інструментальних засобів визначення споживчих уподобань і інформаційної бази підприємства електронної торгівлі з використанням інтелектуального аналізу даних. Досягнення поставленої мети реалізується розв'язком наступних **завдань**:

- Дослідити механізми інтелектуального аналізу даних як засобу визначення споживчих пріоритетів в електронній торгівлі;
- Сформувати систему якісних і кількісних характеристик, що визначають активність споживачів на електронному ринку товарів і послуг;
- Розробити інформаційно-логічну модель визначення споживчих пріоритетів з урахуванням зворотного зв'язку та набору споживчих (споживчий профіль) і особистісних клієнтських характеристик (особистий профіль);
- Розробити метод виявлення клієнтських уподобань з урахуванням споживчого профілю і набору особистісних клієнтських характеристик на основі застосування методу k-means.
- Дослідити методи класифікації і кластеризації споживачів ринку електронної торгівлі;
- Здійснити обґрунтування вибору методів перевірки і візуалізації даних кластеризації;
- Дослідити інструментальні засоби, що забезпечують формалізацію застосування результатів кластеризації і класифікації уподобань клієнтів в електронній торгівлі;

- Розробити інформаційну технологію визначення споживчих пріоритетів в електронній торгівлі на основі методів інтелектуального аналізу даних;

Об'єктом дослідження є процеси формування споживчих пріоритетів в електронній торгівлі.

Предметом дослідження є методи, моделі та інформаційна технологія визначення споживчих пріоритетів в електронній торгівлі.

Методи дослідження: Методологічною і теоретичною основою дослідження є загальнонауковий аналітичний метод, а також системний підхід і праці провідних вітчизняних і закордонних вчених з проблем дослідження електронної торгівлі та побудови методів, моделей та інформаційних технологій управління діяльністю підприємств електронної торгівлі. Інформаційну базу дослідження становлять статистичні дані по електронній торгівлі представлені в глобальній мережі Інтернет. Для практичного вирішення поставлених задач використовувалися такі методи:

- статистичні методи обробки;
- метод модульного проектування;
- методи класифікації даних;
- методи кластеризації даних.

Наукова новизна роботи полягає в розробці інформаційно-логічної моделі виявлення і формування споживчих пріоритетів з урахуванням зворотного зв'язку та набору споживчих і особистісних клієнтських характеристик та у визначенні системи характеристик клієнтів і їх активності на електронному ринку товарів і послуг, яка здійснює комплексну оцінку типового користувача електронної торгівлі, з метою персоналізації набору товарів і послуг для кожного класу споживачів.

Практичне значення випускної кваліфікаційної роботи полягає в тому, що отримані теоретичні результати та програмно-апаратні засоби

можуть бути використані в процесі аналізу клієнтських уподобань на електронному ринку товарів і послуг в частині створення системи підтримки прийняття рішень по виявленню і формуванню споживчих пріоритетів в електронній торгівлі.

Апробація результатів випускної кваліфікаційної роботи. За результатами досліджень здійснено доповідь на конференції КНТЕУ «Соціально-економічні перспективи України у XXI столітті» на тему: «Специфіка електронної торгівлі в Україні: аналіз і перспективи розвитку», яка відбулася в квітні 2018р.

Публікації. Результати дослідження опубліковано у збірнику наукових статей студентів заочної форми навчання, які здобувають освітній ступінь магістра за освітньою програмою «Економічна кібернетика» КНТЕУ «Моделювання економічних та інформаційних процесів» на тему: «Методи кластеризації споживачів ринку електронної торгівлі», 2018 р.

РОЗДІЛ 1.

СУЧАСНИЙ СТАН І ПЕРСПЕКТИВИ РОЗВИТКУ СИСТЕМ ЕЛЕКТРОННОЇ ТОРГІВЛІ

1.1. Електронна торгівля: загальна проблематика, сутність і поняття

Протягом останніх десяти років сучасні інформаційні технології перетворили комп'ютерну мережу Інтернет у розвинену інфраструктуру, яка охоплює всі основні інформаційні центри, бази даних наукової і правової інформації, світові бібліотеки, багато державних та комерційних організацій. Сьогодні Інтернет може розглядатися як величезний електронний ринок, який потенційно здатний охопити практично все населення світу. Саме тому виробники програмних і апаратних засобів, торговельні і фінансові організації активно розбудовують різні види і методи ведення комерційної діяльності в глобальній мережі, так звану електронну комерцію. В результаті електронна торгівля перетворилася в дуже вигідну форму взаємин продавця з покупцем, яка розвивається, не прагнучи при цьому замінити собою інші форми торговельних контактів і зв'язків. Економічні властивості і особливості, які виявилися у процесі становлення інтернет-торгівлі, не тільки забезпечили їй можливість конкурентоспроможного існування у світі сучасного глобального бізнесу, але й створили передумови для цілком оптимістичних прогнозів щодо її майбутнього.

Електронна торгівля (ЕТ) за досить короткий проміжок часу стала явищем, яке почало претендувати на одну із провідних ролей у бізнесі і загалом в економічній діяльності. Вона здатна радикально змінити існуючі економічні й соціальні структури і механізми. Не дивно, що в останні кілька років вона стала одним з головних об'єктів уваги ділових кіл.

Електронна комерція (ЕК) - це щось більше, ніж технологія. Мова йде про всепроникаюче явище, що ґрунтується на сучасних інформаційно-комунікаційних технологіях (ІКТ) і здатному вплинути на всі аспекти ланцюжка створення вартості стосовно товарів і послуг. ЕТ може вестися по будь-яких мережах, які не обов'язково входять в Інтернет. Для цього застосовуються так звані стандарти електронного обміну даними між організаціями (EDI, Electronic Data Interchange) - набори правил електронного оформлення типових бізнес-документів: замовлень, накладних, митних декларацій, страхових форм, рахунків і т.д.

До кінця 60-х років у США вже було чотири індустриальні стандарти для обміну даними в системах управління авіаційним, залізничним і автомобільним транспортом [16]. Така множинність не сприяла розвитку економіки, тому був створений спеціальний Комітет узгодження транспортних даних (Transportation Data Coordination Committee, TDCC), який повинен був вирішити завдання об'єднання форматів. Розробки комітету лягли в основу нового Edі-Стандарту ANSI X.12. Приблизно в ті ж роки аналогічні події відбулися й в Англії, хоча в цій країні головною областю застосування EDI був не транспорт, а торгівля. Вироблений тут набір специфікацій Tradacoms був прийнятий Європейською економічною комісією ООН (UNECE, United Nations Economic Commission for Europe) у якості стандарту обміну даними в міжнародних торговельних організаціях. Цей набір форматів і протоколів називається GTDI (General-purpose Trade Data Interchange). В 80-х почалися роботи з об'єднання європейських і американських специфікацій. На базі GTDI міжнародна організація ISO сформувала новий стандарт Electronic Data Interchange for Administration, Commerce and Transport (EDIFACT, ISO 9735), що використовує в якості транспортного протоколу протокол електронної пошти X.400. В той же

час, в США більшість Edi-Систем і на сьогодні базуються на стандарті ANSI X.12.

Робота, вище зазначених комісій, і комітетів супроводжувалася бурхливим ростом оборотів ЕК і числа залучених до неї компаній. В 1996 р. (коли торгівля за допомогою мережі Інтернет була ще на початковому етапі розвитку) за допомогою Edi-Транзакцій було зроблено операцій на 300 млрд.дол., а в 1999 р. - на 1,1 трлн.дол. До 2003 р. цей показник досяг 2,3 трлн.дол. [17]. Розвиток мережі Інтернет з її низькою собівартістю передачі даних зробило актуальною модернізацію Edi-Систем для використання в новому інформаційно-комунікаційному середовищі. Тому в середині 90-х років був розроблений ще один стандарт — EDIFACT over Internet (EDIINT), що описує, як передавати Edi-Транзакції за допомогою протоколів безпечної електронної пошти SMTP/S-MIME.

Незважаючи на всі успіхи мережі Інтернет, компанії не поспішають змінювати традиційним технологіям торгівлі. В 1999 р. на частку Edi-Транзакцій через Інтернет доводилося всього 12% від загального обороту EDI-Бізнесу, а до 2003 р. ця частка збільшилася лише до 41% [17]. Однак саме Інтернет дає потужний поштовх до подальшого розвитку ЕК у силу доступності цієї глобальної мережі не тільки для великих корпорацій, підприємств малого і середнього бізнесу, але й для окремих осіб, і, як наслідок, можливості залучити до ЕК суттєво більш широку аудиторію постачальників і споживачів. Згідно з дослідженням фахівців. У порівнянні із Заходом у сценарії розвитку ЕК в Україні можна легко виявити наступну відмінність. На Заході електронна комерція в основному розвивалася шляхом заміщення. На момент появи Інтернету в США, наприклад, була налагоджена система електронної оплати, доставки, каталогів і всього іншого, що необхідно для покупок на відстані. Тому електронна комерція поступово заміщає частину традиційної торгівлі.

Дійсно, «пасивні» Web-Сайти (Web-Вітрини) замінили каталоги; електронна пошта і екранні форми - паперові бланки замовлень; платіжні системи - посилку чеків або факсів з номерами карток і так далі. Навіть у самій консервативній області - доставці матеріальних товарів - Інтернет знайшов свою нішу, наприклад у системах посилок, що відслідковують проходження товару. При цьому кількісний ріст перейшов у якісний. У якості маркера цього переходу можна відзначити ряд подій 1998 року, зокрема, доповідь Клінтона - Гора «A Framework for Global Electronic Commerce» ("Структурні основи електронної комерції") [18]. В Україні ситуація була принципово іншою – заміщувати практично не було чого. З одного боку, це добре - не потрібно долати опір існуючої системи. З іншого боку, це означало, що для розвитку комерції необхідно нагромадити критичну масу магазинів і покупців. Таким чином, глобальним завданням для українського ринку електронної комерції було і є нагромадження критичної маси користувачів, інвестицій, технологій, по досягненню якої процес розвитку електронної комерції стане самопідтримуючим.

Київ, Львів, Харків, Одеса, Дніпропетровськ відіграють роль свого роду інноваційних і телекомунікаційних центрів, звідки хвилі інновацій розходяться по інших містах і регіонах. Інші міста відстають за рівнем розвитку Інтернету, що виражається як у кількості користувачів, так і в ступені розвитку ринків послуг надання доступу в Інтернет, електронної комерції та інших напрямків електронного бізнесу. Якщо в містах-мільйонерах Інтернет став досить звичайним явищем, то в містах з населенням 50 - 100 тис. чоловік він проникає в суспільно-економічне життя в масовому порядку тільки в цей час (у сільській місцевості поки взагалі не отримав належного поширення, що пояснюється досить низьким

рівнем мерережево-комунікаційного розвитку, не говорячи вже про дуже низький рівень доходів населення).

Сам процес розвитку Інтернету і електронної комерції на території України представляє собою одну велику неперервну інноваційну хвилю, що поступово захоплює міста різних типів - від «мільйонерів» до малих. В 1991 - 1992 роках відбулося зародження комерційних мереж по наданню послуг електронної пошти, конференцій USENET, FIDO. Ці мережі функціонували в основному в Києві, Львові, Харкові і Одесі. 1993 - 1994 роки характеризувалися появою і розгортанням по всій країні мереж «Eunet/УкрТелеКом» і «Демос-Інтернет». Почали формуватися регіональні мережі передачі даних, що входять переважно в мережу «УкрТелеКом», на базі підприємств електрозв'язку або їх дочірніх компаній. Мережі використовувалися різними науковими організаціями. Новітні інформаційні технології почали впроваджуватися у великих організаціях (в своїй основі фінансовими), які підключали свої локальні мережі до глобальної мережі. В 1995 — 1996 роках стали з'являтися загальноукраїнські комп'ютерні мережі (UkrNet, Triolan та ін.). Регіональні оператори електрозв'язку ухвалюють рішення щодо виходу на ринок надання послуг доступу в мережу Інтернет. У ці роки починається підключення окремих фізичних осіб до мережі Інтернет. Мережні представництва могли дозволити собі лише далеко не всі навіть досить заможні компанії. На ранніх етапах впровадження інформаційно-комунікаційних технологій в економічну діяльність все починалося з невеликих статичних сайтів, без інтерактива, єдиною метою яких було привести потенційних покупців в існуючий фізично магазин. Під присутністю в Мережі розумілися лише контактні телефони, адреса, схема проїзду; навіть електронна пошта була рідкістю. Потім, поступово, з розвитком систем управління базами даних і збільшенням пропускних здатностей каналів обміну даних, почали

з'являтися періодично оновлювані інтернет-версії прайс-листів на товари або послуги компаній. Як правило, під Інтернет-Магазином тоді розуміли саме Інтернет-інтерпретацію реально існуючого магазину, що дозволяє переглядати прайс-листи цього магазину за допомогою Інтернету. В 1997 - 1998 роках відбувся перехід регіональних провайдерів на переважне використання наземних каналів зв'язку «УкрТелеКому» замість супутникових каналів зв'язку. В цей час починають формуватися регіональні центри обміну інтернет-трафіком. У Києві і Львові розвивається ринок веб-дизайну, формуються перші великі загальноукраїнські контент-проекти.

Починаючи з 1998 р. відбуваються конференції і семінари по електронній комерції, активізувалися Інтернет-магазини, було запущено в практичну експлуатацію кілька цікавих платіжних систем. У ці роки почали відбуватися процеси, що забезпечили повноцінне існування Інтернет-магазинів як повністю самостійних організацій, а саме:

- збагатилися засоби взаємодії із клієнтом, який одержав можливість не тільки переглядати, але й замовляти, а згодом і оплачувати товари або послуги, не відходячи від комп'ютера (поява електронних платіжних систем);
- відбувся розрив між реальними магазинами і їх Інтернет-версіями, тобто Інтернет-магазин зміг існувати без фізичного магазину.

У період з 1999 до середини 2000 року в Україні, як і в усьому світі відбувається так званий «інвестиційний бум». Цей бум був викликаний як західним ажіотажем навколо електронної комерції, так і бажанням, нарешті, знайти спосіб заробляти гроші на українському Інтернеті. Стало зрозуміло, що ні веб-дизайн, ні контент-провайдинг, ні показ банерів не приводять у сучасних умовах до масової прибутковості або хоча б до мінімальної рентабельності. Виявилося, що створення будь-якого

масового, особливо комерційного веб-сайту вимагає досить значних витрат, вимірюваних тисячами доларів, не тільки на апаратну частину, але й на дизайн і розкручування сайту. Тому є природною зміна інтересів на нові форми продажів і одержання доходу, зокрема електронну комерцію. Окрім цього, важливим позитивним моментом стало розуміння стабільності українського Інтернету. Інтернет-Провайдинг (ринок надання послуг доступу до мережі Інтернет) спокійно пережив усі потрясіння економічної кризи, що дозволяє розглядати український Інтернет як досить стійкий сегмент економіки, в який можна інвестувати фінанси. На тлі цих процесів відбувається інтенсивний розвиток великих інформаційних інтернет-проектів (у тому числі так званих порталів), створюються інтернет-холдинги. Виникають Інтернет-холдинги Slando.ua (OLX.ua), Aukro.ua, Prom.ua, які на сьогоднішній день є найбільшими роздрібними продавцями у українському Інтернеті. За рахунок збільшення числа операторів доступу до мережі Інтернет і посилення конкуренції між ними починає відбуватися поступове зниження вартості підключення до Інтернет і дана послуга поступово перестає бути «іграшкою для багатих». Усе більша кількість людей мають потенційну можливість підключитися до мережі Інтернет.

У середині 2000 року відбувається так звана «інвестиційна криза» на столичному ринку Інтернет-Контенту, однак на регіональних ринках Інтернет-Контенту відбувається інтенсивний розвиток. Уже до 2003 року остаточно сформувалися різні напрямки електронної комерції (B2C, B2B). Підприємства все більше використовують різні схеми аутсорсинга. Відбувається масовий розвиток «локальних мереж», великі московські провайдери виходять на регіональні ринки. Широкозмуговий кабельний доступ до Інтернету (за допомогою «будинкових мереж», технології ADSL та ін.) прискорив зростання Інтернет-трафіка і змінив характер

використання Інтернету звичайними користувачами та підприємствами. У сфері роздрібної електронної торгівлі (B2C) абоненти, що користуються широкосмуговим підключенням, з більшим ступенем імовірності включаються в електронну торгівлю, що спричинює швидке зростання інтернет-трафіка і онлайн-ових трансакцій. Разом з тим підприємства закупають у режимі онлайн набагато більше продуктів, ніж звичайні споживачі, а сучасні інформаційно-комунікаційні технології роблять такі продукти більш доступними, простими для використання, а отже, більш широко реалізованими, особливо для малих і середніх підприємств. Інформаційні технології дозволяють відразу декільком користувачам розподіляти підключення до Інтернету, що скорочує витрати на кожне індивідуальне підключення і є досить важливим рішенням для малих і середніх підприємств. Для великих підприємств централізація даних і додатків із застосуванням єдиного обладнання для зберігання інформації, забезпечує багатьом територіально віддаленим користувачам можливість одержувати доступ до значних масивів інформації, а також полегшує впровадження нових форм організації і ведення бізнесу.

Різні люди розуміють електронну комерцію по-різному. Для одних електронна комерція означала лише ведення торгівлі з використанням комп'ютерів, у той час як для інших до неї включалися види комерційної діяльності, що здійснюються через канали комп'ютерного зв'язку. У багатьох випадках визначення і показники електронної комерції були розроблені в маркетингових цілях, а не для проведення кількісного виміру або науково обґрунтованого аналізу цього явища.

Закордонні фахівці Саммер А. і Дункан Гр. під ЕК розуміють будь-яку форму бізнес-процесу, в якому взаємодія між суб'єктами відбувається електронним чином (з використанням Інтернет-технологій) [19]. Згідно з визначенням наданим в роботі [20], під ЕК розуміють технологію, що

забезпечує повний замкнений цикл операцій, який включає замовлення товару (послуги), проведення платежів, участь в управлінні доставкою товару (виконання послуги). Ці операції проводяться з використанням електронних засобів та інформаційних технологій і забезпечують перехід права власності або права користування від однієї юридичної (фізичної) особи до іншої. Дане визначення близьке до бачення електронної комерції рядом закордонних авторів, таких як Хейг М. [9], Лаудон К. [13], Козьє Д. [14], Хартман А. [15]. Так, згідно їхньої точки зору, поняття ЕК зводиться до будь-яких форм угод, при яких взаємодія сторін здійснюється електронними засобами замість фізичного обміну або безпосереднього фізичного контакту, і в результаті право власності або право користування товаром чи послугою передається від однієї особи іншій.

Звернемося до аналізу одного зі складових поняття «електронна комерція», а саме до «комерції», для того, щоб зрозуміти - чи є істотною різниця між поняттями «електронна комерція» і «електронна торгівля».

Слово «комерція» походить від латинського «коммерціум», що в перекладі означає - «торгівля». Однак треба мати на увазі, що термін «торгівля» має двояке значення: в одному випадку він означає самостійну галузь народного господарства (торгівлю), в іншому - торговельні процеси, що спрямовані на здійснення актів купівлі-продажу товарів. Комерційна діяльність пов'язана із другим поняттям торгівлі - торговельними процесами по здійсненню актів купівлі-продажу з метою одержання прибутку. Автор І.К.Беляєвський у своїй праці «Статистика ринку товарів і послуг» так визначає поняття «комерція»: «під комерцією розуміють діяльність, спрямовану на отримання прибутку» [21]. У книзі О.Є. Башиної «Статистика комерційної діяльності» надається наступне визначення: комерція це діяльність організацій і підприємств, а також окремих осіб по здійсненню операцій купівлі-продажу товарів і послуг з

метою задоволення попиту і одержання прибутку. У більш широкому сенсі під комерцією, автори розуміють, будь-яку підприємницьку діяльність, спрямовану на отримання прибутку [22].

Важливо пам'ятати, що комерційна діяльність - це не тільки торгівля, але й діяльність організацій, підприємств і приватних осіб, спрямована на здійснення операцій купівлі-продажу товарів і послуг з метою задоволення попиту населення і одержання прибутку. До комерційних також відносяться деякі процеси, що забезпечують здійснення операцій купівлі і продажу товарів у сфері товарообігу: дослідження споживчого попиту, організація господарських зв'язків між постачальниками і покупцями товарів, рекламування товарів, оптові закупівлі та ін.

Мета комерційної діяльності - максимізація прибутку. Досягнення цієї мети постійно супроводжує весь процес комерційної діяльності з кожним контрагентом по кожному конкретному випадку з використанням конкретних умов. Гарантією реалізації цих цілей є сенс комерційної діяльності, що полягає у вивченні процесу формування ринку товарів і послуг, обґрунтуванні напрямків і масштабів розвитку їх виробництва відповідно до потреб суспільства і окремих споживачів, доведення товарів до споживачів і організації самого процесу споживання, комерційне посередництво і організація договірних зв'язків на ринку товарів і послуг.

Враховуючи вище викладене, можна зробити висновок, що комерція — це діяльність економічних суб'єктів по здійсненню операцій купівлі-продажу товарів і послуг з метою задоволення попиту, спрямована на отримання прибутку. Фактично, це визначення повторює по змісту визначення поняття торгівля, із чого можна зробити припущення, що відмінностей між поняттями «електронна комерція» і «електронна торгівля» також бути не повинно. На нашу думку, з деякими припущеннями, можна поставити знак рівності між цими поняттями.

В той же час, якщо електронна комерція повинна охоплювати все, що передбачають терміни, які входять у визначення, то таке визначення ризикує стати досить широким і досить складним для розуміння. Тому ми вирішили зосередити свою увагу на чітко визначених операціях, які здійснюються комунікаційними мережевими засобами, а не на широкому понятті комерції або комерційної діяльності. Широке поняття електронної комерції як і раніше становить значний інтерес як для підприємств, так і для державних органів, однак для досягнення прогресу у визначенні і можливості кількісно осягнути електронну комерцію (для наступної, економіко-математичної оцінки) необхідно звужити і уточнити область охоплення економічних процесів, що підлягають під визначення електронної комерції.

Угода є кінцевою фіксованою подією, яка може бути виділена із цілого ряду інших подій комерційної діяльності. На нашу думку, при визначенні сутності електронної комерції слід звернутися до такого поняття як спосіб розміщення (або одержання) замовлення, яке визначає, чи відноситься угода до електронної комерції чи ні.

У широкому сенсі, електронна угода - це купівля-продаж товарів або послуг підприємствами, домашніми господарствами, фізичними особами, державними або приватними організаціями, що здійснюється за допомогою комп'ютерних мереж. Дане поняття містить у собі замовлення, які отримуються або розміщуються з використанням будь-якого он-лайнного програмного додатку, що застосовується для здійснення автоматизованих угод (наприклад таких, як інтернет-додатки, технологія EDI або інтерактивні телефонні системи). У більш вузькому сенсі інтернет-угоду потрібно розглядати як купівлю-продаж товарів або послуг підприємствами, домашніми господарствами, фізичними особами, державними або приватними організаціями, що здійснюється через Інтернет. В цьому випадку

поняття містить в собі замовлення, які отримуються або розміщуються з використанням будь-якого Інтернет-додатку, що застосовується для здійснення автоматизованих угод, таких як веб-сторінки або будь-які інші веб-додатки незалежно від способу доступу до цієї мережі (наприклад, через стільниковий телефон, кишеньковий комп'ютер і т.п.). Хоча основна увага в різних дослідженнях приділяється електронній комерції, загалом визнається, що електронна комерція є лише одним з важливих елементів більш широкого набору процесів при веденні електронного бізнесу.

Як і електронна комерція, електронний бізнес має глибокий і модифікуючий вплив на економіку і суспільство. Він впливають на структуру фірм, а також увесь цикл пропозиції і споживання, починаючи з розробки продукту і закінчуючи післяпродажним обслуговуванням. Електронні бізнес-процеси можуть здійснюватися не тільки через Інтернет, але й через будь-які інші комп'ютерні мережі. Прикладами електронних бізнес-процесів можуть бути:

- вивчення процесу формування ринку товарів і послуг, пошук комерційних партнерів, доведення товарів до споживачів і організація самого процесу споживання, комерційне посередництво і організація договірних зв'язків на ринку товарів і послуг — електронна комерція;
- придбання і збереження клієнтури - використання Customer Relationships Management або CRM-систем — систем управління взаємодією з клієнтами;
- управління фінансами, бюджетом і рахунками - електронний банкінг;
- управління людськими ресурсами - використання Human Resources або HR-систем - систем управління людськими ресурсами;

- проектування і розробка продуктів;
- виконання замовлень і відстеження їх виконання;
- логістика (вхідна і вихідна) і облік фінансвих і товарно-матеріальних запасів;
- обслуговування і підтримка продукції;
- післяпродажна підтримка й ін.

Ці бізнес-процеси здійснюються як всередині конкретних фірм, так і між ними, у зв'язку із чим, для розуміння мінливого характеру границь і взаємозв'язків фірм, дуже важливо провести чіткий розподіл між процесами, що відбуваються всередині фірм і між ними. На основі вищевикладеного була розроблено схематичне представлення співвідношення розглянутих понять на електронному ринку (рис. 1.1). Із схеми, що представлена на рис. 1.1 видно, що електронна-комерція є синонімом електронної-торгівлі і означає електронну торгівлю, що здійснюється за допомогою глобальної мережі Інтернет.



Рис. 1.1. Співвідношення понять електронного ринку

Електронний бізнес містить у собі як внутрішні (внутрішньофірмові), так і зовнішні бізнес-процеси (включаючи бізнес-процеси, що здійснюються економічними суб'єктами при здійсненні комерційної діяльності), які реалізуються з використанням будь-яких типів комп'ютерних мереж, що існують на сьогодні у світі (глобальні, локальні, корпоративні та ін.).

Остаточний варіант визначення електронна комерція буде звучати в такий спосіб: *електронна комерція (електронна торгівля)* - це діяльність економічних суб'єктів по здійсненню комерційних трансакцій з використанням електронних засобів обміну даними. Дане визначення добре узгоджується із визначенням електронної торгівлі наданим в роботі [23]. При цьому хочеться ще раз підкреслити, що торгівля може відбуватися як товарами (матеріальними/нематеріальними), так і послугами (страхові, банківські й ін.). Під електронними засобами обміну даних розуміються будь-які електронні засоби обміну даних (мережі EDI, телефон, телеграф і ін.), а не тільки Інтернет.

1.2. Теоретичні методи в дослідженні механізмів функціонування систем електронної торгівлі

Ефективне управління розвитком електронної комерції, а також процесами інформатизації економіки в цілому неможливо без повноцінного і всебічного економіко-математичного дослідження всього комплексу проблем, включаючи, з одного боку, діяльність підприємств галузі електронної торгівлі, а з іншого - використання інформаційних технологій на підприємствах і організаціях всіх галузей економіки [24-26].

У зв'язку з більш широким впровадженням технологій електронної комерції наукова розробка методик застосування економіко-математичних

методів у дослідженні стану і перспектив розвитку електронної комерції суттєво активізувалася і, що дуже важливо, підвищилася їхня затребуваність в практичній діяльності.

Застосування всього апарату відомих економіко-математичних методів для розв'язку багатьох конкретних завдань дозволяє підвищити ефективність діяльності економічних суб'єктів, які в своїй діяльності активно використовують електронну торгівлю. Загалом мова йде про розробку інструментарію, який може бути застосований для аналізу складних закономірностей в електронній торгівлі і який буде основою для розробки діючих механізмів ефективного управління та прийняття рішень на державному рівні.

Економіко-математичні методи представляють собою комплекс економіко-математичних дисциплін, що є симбіозом економіки, математики та новітніх інформаційних технологій [27-29]. В структурі економіко-математичних методів, що використовуються для дослідження і прогнозування стану та розвитку електронної комерції, можна виділити наступні їхні напрямки:

- математична статистика:
 - дисперсійний аналіз;
 - кореляційний аналіз;
 - регресійний аналіз;
 - кластерний аналіз та ін.
- методи прийняття оптимальних рішень:
 - лінійне програмування;
 - нелінійне програмування;
 - динамічне програмування;
 - стохастичне програмування;
 - математичне програмування;

- теорія й методи управління запасами;
- теорія масового обслуговування та ін.
- експериментальні методи вивчення економічних процесів:
 - імітаційне моделювання;
 - методи експертних оцінок і т.п.

Навіть коротка характеристика структури економіко-математичних методів, наведена вище, свідчить про надзвичайну широту наукового арсеналу теоретичних і прикладних математичних методів і моделей, які можуть використовуватися учасниками ринку електронної комерції в плануванні, управлінні і прогнозуванні своєї діяльності [30].

Дослідження, проведене в дисертаційній роботі, містять у собі кілька напрямків застосування економіко-математичних методів:

1. Економіко-математичне дослідження динаміки основних показників електронної комерції в Україні і залежності стану ЕК в Україні від макроекономічних показників з використанням кореляційного та регресійного методів;
2. Економіко-математичне прогнозування найважливіших показників стану і розвитку електронної комерції в Україні;
3. Економічна інтерпретація отриманих чисельних результатів і розробка рекомендацій з їхнього практичного застосування.

Основний зміст *першого напрямку* зводиться до розв'язку наступних завдань:

- оцінка взаємозв'язків між показниками за допомогою парних, частових і множинних коефіцієнтів кореляції;
- побудова багатofакторних регресійних моделей;
- оцінка істотності зв'язку (якості регресійних моделей).

Економіко-математичне прогнозування найважливіших показників стану і розвитку електронної комерції в Україні, що становить суть *другого напрямку* дослідження, полягає у розв'язку комплексу наступних завдань:

- прогнозування на основі регресійних моделей;
- прогнозування економічної динаміки на основі адаптивних моделей:
 - вибір оптимальної моделі;
 - побудова адаптивних моделей;
 - оцінка адекватності і точності адаптивних моделей;
 - одержання точкового і інтервального прогнозів;
 - верифікація прогнозу.

Третій напрямок передбачає повну економічну інтерпретацію отриманих чисельних результатів і їх практичне застосування. Воно містить у собі верифікацію і валідацію отриманих регресійних і прогностичних моделей, а також побудову оптимальних, управлінських рішень і розробку практичних рекомендацій.

Проведення економіко-математичного дослідження вимагає наявності певного набору первинної статистичної інформації для аналізу [31]. Наявні сьогодні дані статистичної звітності носять безсистемний характер і не охоплюють найважливіших аспектів функціонування електронної комерції. Спектр наявних статистичних даних обмежений в основному декількома показниками, і такий стан справ ускладнює проведення економіко-математичного аналізу стану і прогнозування тенденцій розвитку електронної комерції. Варто особливо підкреслити, що статистичне спостереження за станом і розвитком електронної комерції - досить новий напрямок дослідження, який вимагає детального опрацювання термінологічного апарата, складу використовуваних показників, створення інструментарію і організації збору та обробки

даних, формування зведених даних для прийняття управлінських рішень і т.д. Ця діяльність вимагає постійного розвитку і доповнення. У процесі вдосконалення інструментарію необхідно зберігати послідовність із попередніми роботами, виконаними раніше, що дозволить одержати тимчасові ряди, які характеризують процеси розвитку електронної комерції в Україні. На основі вище згаданого можна зробити висновок, що першочерговим завданням, яке передує економіко-математичному дослідженню механізмів функціонування ринку електронної торгівлі, є виявлення і розробка системи взаємозалежних показників електронної торгівлі.

1.3. Специфіка електронної торгівлі в Україні: аналіз і перспективи розвитку

Електронна комерція, як вид економічної діяльності, це наслідок глибоких соціально-економічних змін в суспільстві викликаних технічним прогресом і процесами глобалізації. У порівнянні з країнами Західної Європи та США рівень розвитку вітчизняного ринку електронної торгівлі значно відстає, оскільки в Україні знаходиться на початковому етапі інтенсивного розвитку цього виду економічної діяльності. Головна причина полягає у відносно нетривалому періоді часу широкого застосування мережевих інформаційних технологій для комерційної діяльності. За даними досліджень компанії GfK Ukraine [32], більше 50% населення України сьогодні є користувачами Інтернет-мережі (рис. 1.2), тоді як у Європі цей показник приблизно на 35% вище. Частка торгівлі онлайн в загальному обсязі українського торговельного ринку становить приблизно 2%, тоді як у країнах Євросоюзу середній показник 6,3%, а у Великобританії – 13% [33].

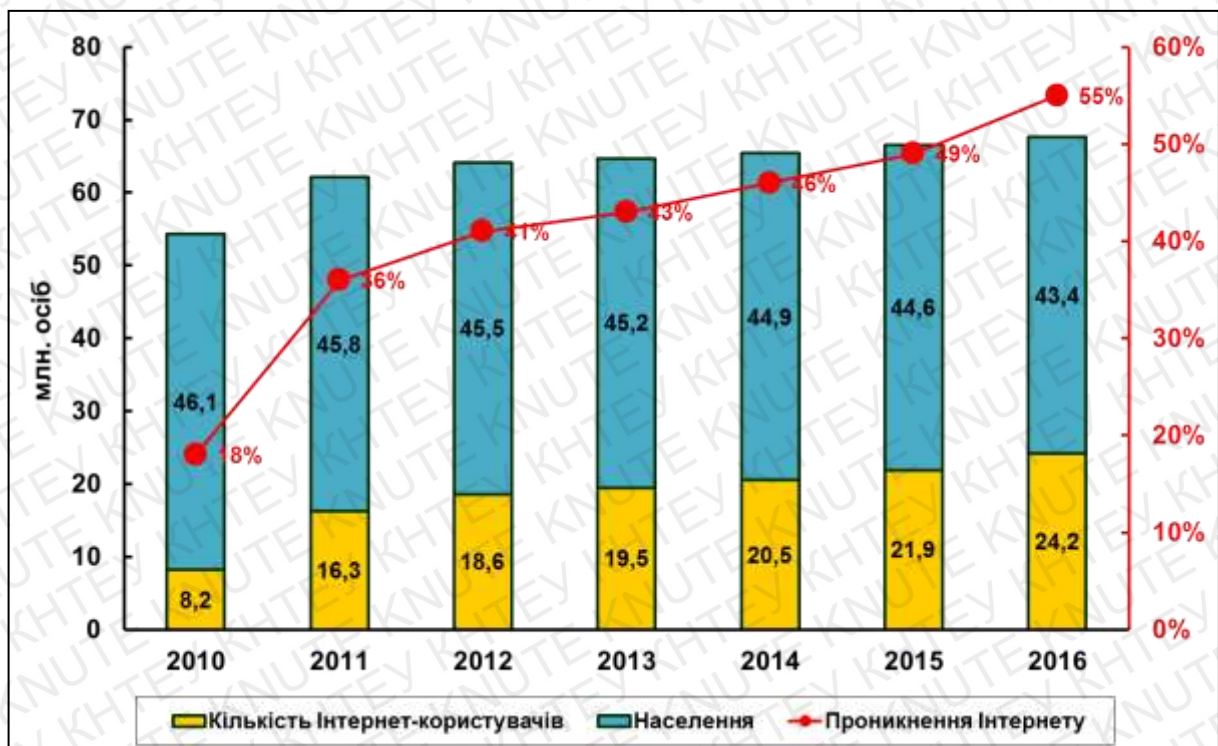


Рис. 1.2. Рівень проникнення Інтернету в Україні [32]

Україна має високі показники росту ринку електронної торгівлі [32]. На вітчизняному ринку електронної торгівлі спостерігається важливий тренд - з впровадженням високошвидкісного 3G Інтернету, все більше користувачів віддають перевагу мобільним додаткам в електронній торгівлі. Усе більшу частину ринку електронної комерції відвоює так звана мобільна комерція. За 2016-2017 роки оператори обіцяють покрити 3G всю територію України, тож ще через рік проникнення мобайл очікується на рівні 50% [34].

Український ринок e-commerce перебуває лише на етапі становлення, водночас має значний потенціал для розвитку [35]. Обсяг українського ринку електронної комерції в 2014 році в доларовому еквіваленті скоротився на 20% (у порівнянні з 2013р.) і склав 1,6 млрд. дол., повідомляється в дослідженні Української асоціації директ-маркетингу (УАДМ) [100].

Водночас гривневе зростання ринку склало 19%. За даними Української асоціації директ-маркетингу, у 2015 році обсяг ринку електронної комерції в доларах скоротився приблизно на 45% до 1,1 млрд. дол. [36]. Згідно з дослідженням, торік зросли продажі всіх груп товарів, незначне зростання показали тільки сегменти товарів для авто, косметика і парфумерія. Найбільшим сегментом залишається сегмент електроніки і побутової техніки: за підсумками року його обсяг склав 750 млн. дол., скоротившись на 25% у доларах і збільшившись на 11,3% в гривні [36]. Падіння продажів у вітчизняних онлайн-магазинах пояснюється девальвацією гривні та скороченням купівельної спроможності громадян. Частково через це триває переорієнтація на зарубіжні онлайн-магазини, зокрема, китайські [34]. Як зазначається в звіті Української асоціації директ-маркетингу [36], за умови стабілізації економічної ситуації в Україні, за обсягами ринок електронної торгівлі вийде на рівень 2013 року в середині 2018 року (рис. 1.3).

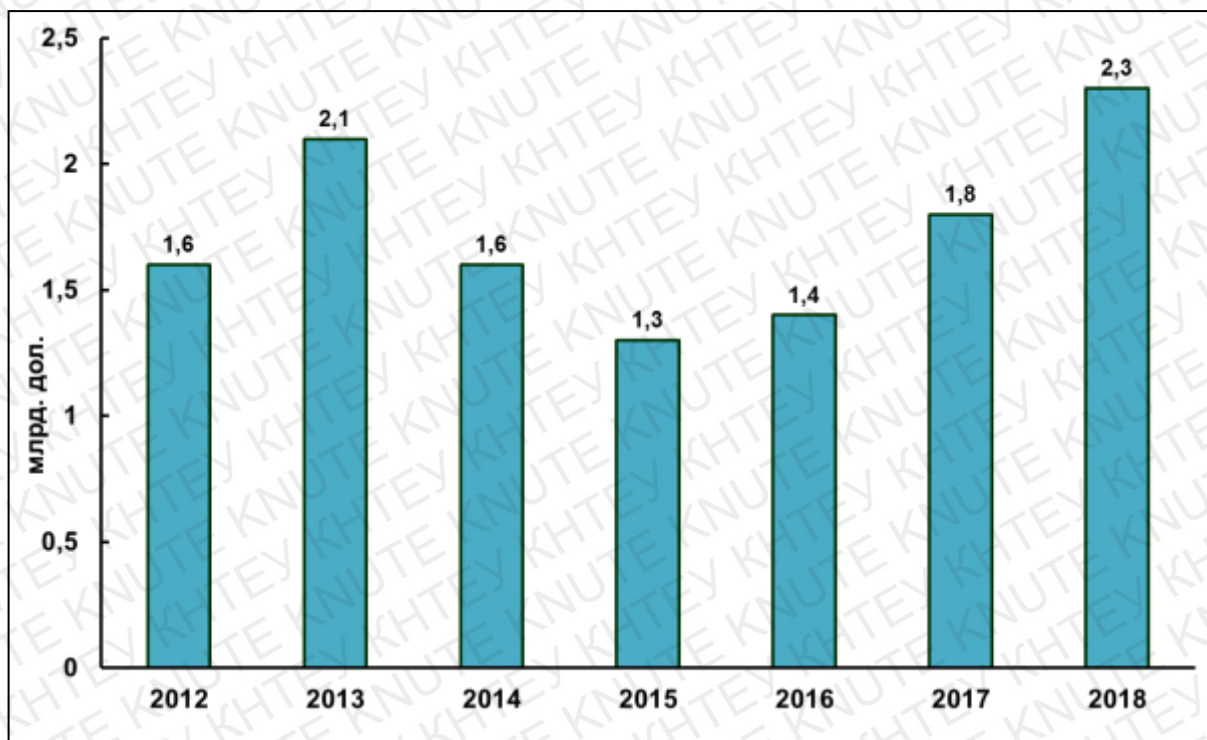


Рис. 1.3. Об'єм ринку електронної торгівлі в Україні [36]

У трійку лідерів за обсягом ринку в сегменті електроніки та побутової техніки (рис. 1.14) увійшли інтернет-магазин Rozetka.ua (частка 39,7%), “Алло” (частка 10,6%) і Fotos (6,6%) [36].

Обсяг сегменту Fashion (одяг та взуття) торік склав \$131 млн, скоротившись в доларах на 13% і збільшившись на 9,6% в гривні. Лідером у цьому сегменті став магазин Bon Prix (частка 25,7%).

На другому і третьому місці інтернет-магазини LeBoutique і ModnaKasta (рис. 1.14), частки яких становлять 22,5% і 18,6% відповідно [36].

В той же час, на ринку можуть виникнути досить серйозні зміни: практично монопольне становище Розетки, міноритарний пакет якої не так давно придбав фонд прямих інвестицій під управлінням компанії Horizon Capital [37], в сегменті електроніки, а також вихід маркетплейсу в інші ринкові сегменти, тим не менш, не означає, що компанія покриє весь ринок. У трійку лідерів за обсягом ринку в сегменті електроніки та побутової техніки (рис. 1.4) увійшли інтернет-магазин Rozetka.ua (частка 39,7%), “Алло” (частка 10,6%) і Fotos (6,6%) [36].

Обсяг сегменту Fashion (одяг та взуття) торік склав \$131 млн, скоротившись в доларах на 13% і збільшившись на 9,6% в гривні. Лідером у цьому сегменті став магазин Bon Prix (частка 25,7%). На другому і третьому місці інтернет-магазини LeBoutique і ModnaKasta (рис. 1.4), частки яких становлять 22,5% і 18,6% відповідно [36].

В той же час, на ринку можуть виникнути досить серйозні зміни: практично монопольне становище Розетки, міноритарний пакет якої не так давно придбав фонд прямих інвестицій під управлінням компанії Horizon Capital [37], в сегменті електроніки, а також вихід маркетплейсу в інші ринкові сегменти, тим не менш, не означає, що компанія покриє весь ринок.

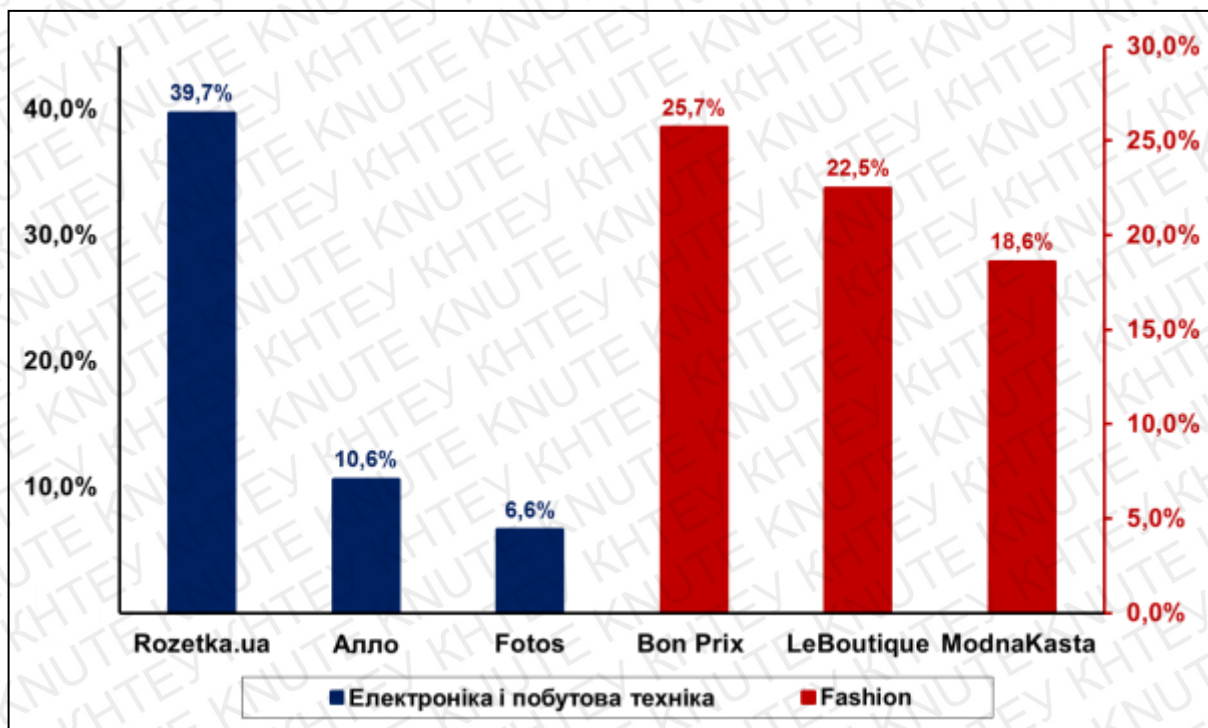


Рис. 1.4. Відсотковий розподіл лідерів ринку електронної торгівлі в Україні в сегментах fashion та електроніки і побутової техніки [36]

Залишаються можливості для розвитку нішевих е-ритейлерів. Наприклад, у вертикалях ринків одягу та взуття, продаж авіаквитків і сервісів бронювання готелів, косметики, дитячих товарів є свої визнані лідери. На відміну від розвинених ринків, в Україні все ще невелика частка онлайн-продажів товарів повсякденного попиту, будматеріалів, спортивних товарів, послуг [38]. З іншого боку, зростання кількості транскордонних угод з тими ж eBay і Amazon, а також розширення присутності на українському ринку AliExpress та JD.com, з їх широким асортиментом, низьким рівнем цін на товари та безкоштовною доставкою на більшість товарів по всьому світу, може суттєво вплинути на розстановку сил у цілих сегментах online-індустрії, зокрема, і на позиції лідерів ринку електронної торгівлі. Що ж стосується українських гравців, важливим трендом залишатиметься побудова бізнесу на базі

мультиканальності продажів: інтернет-магазини будуть відкривати шоу-руми (як робить та ж Розетка), у той час як традиційний рітейл буде йти в онлайн-торгівлю (приклади Foxtrot, Comfy, Eldorado, Helen Marlen, Intertop) [38].

Зростання прибутку підприємства при одночасному зниженні витрат і постійному вдосконаленні бізнес-процесів залишається основною метою будь-якого підприємства [35]. Один із способів досягнення цієї мети – електронні засоби ведення торгівлі. Використання електронної торгівлі розширює бізнес-простір і змінює організаційну структуру ведення бізнесу. Організація бізнесу за допомогою електронних засобів торгівлі дозволяє подолати традиційні бар'єри, які полягають в територіальній віддаленості і нестачі інформації щодо можливостей торговельного ринку. Відсутність територіальної залежності споживачів в електронній торгівлі сприяє значному збільшенню клієнтської бази [39]. Враховуючи вище зазначені обставини, дослідження ринку електронної торгівлі набуває не тільки теоретичного інтересу, але й практичного значення.

Одним із головних стримуючих факторів розвитку електронної торгівлі в Україні є недовіра споживачів до купівлі товарів через всесвітню мережу. Більша половина населення України вважає купівлю товарів в електронних засобах торгівлі ризикованою справою. Підвищення довіри споживачів в електронній торгівлі забезпечується двома факторами: зручність роботи з інтернет-магазином і позитивний досвід клієнтів використання електронних засобів торгівлі. Ці два фактори в електронній торгівлі формуються цілим набором позитивних характеристик роботи інтернет-магазину, зокрема: сучасний інтерактивний дизайн сайту, простота у використанні для пересічного користувача, широкий асортимент товарів і послуг, доступність і перевіреність інформації, наявність зворотного зв'язку, рівень роботи контакт-центру, інтерактивна

комунікація з покупцем, механізми доставки товарів, швидкість доставки, відповідність якості товарів до описаної на сайті, наявність широкого спектру різних видів оплати товарів і послуг і т.п.

На розвиток електронного торговельного ринку, також суттєво впливає ефективність служби доставки товарів, оскільки трансакційні витрати покупців в електронній торгівлі можуть бути суттєво зменшені за рахунок оптимізації маршрутів доставки товарів, що створює додаткові переваги на конкурентному торговельному ринку [40]. Важливу роль в процедурах доставки товару відіграє поведінка кур'єра. Цей фактор пов'язаний з тим, що в електронній торгівлі кур'єр одночасно виконує і роль продавця-консультанта звичайних магазинів. Від його професійності багато в чому залежить сприйняття електронних механізмів торгівлі і формування позитивного клієнтського досвіду та споживчих пріоритетів.

В розвинутих країнах Європи і Сполучених Штатах Америки більшість купівель в інтернеті оплачуються по безготівковому рахунку, в той час, як в Україні для оплати покупок в інтернеті цей вид розрахунку використовується мало (15% від загальної суми платежів) [41]. Цьому сприяють дві основні обставини: нерозвиненість електронних платіжних засобів і психологічний фактор, коли покупець готовий оплатити товар тільки після його отримання і перевірки його якості. Із розвитком ринку e-commerce в Україні цей вид оплати буде займати всі більшу частку на електронному ринку. На сьогодні, почали успішно впроваджуються системи багаторівневих затримок електронних платежів, які раніше в Україні взагалі не використовувалися. Такі системи дозволяють уникати негативних наслідків електронних покупок, оскільки в даному випадку, при здійсненні покупки онлайн платіж за товар або послугу затримується на рахунку банку. Це страхує покупця від недобросовісних інтернет-продавців і убезпечує від необхідності самому вирішувати питання по

поверненню витрачених грошових засобів. У впроваджені таких систем зацікавлені як інтернет-продавці, так і споживачі. За рахунок впровадження багаторівневих платіжних систем значно підвищується довіра до електронної торгівлі, що, в свою чергу, сприяє підвищенню активності онлайн споживачів.

Характерною особливістю електронних торговельних ринків є високе відношення рекламних витрат до отриманого прибутку. Високі витрати вказують на важливу роль реклами на електронних ринках. В електронній торгівлі збільшення торговельного ринку пов'язано із двома аспектами [42]: територіальне розширення місцезнаходження покупців і продавців та збільшенням цільової аудиторії. В інтернет-просторі, про перший аспект можна говорити тільки як про неявний параметр функції трансакційних витрат, що може корелювати з відстанню доставки товару до покупця [40]. Другий аспект функціонально залежить від рекламних витрат продавців, або іншими словами притоку на ринок нових покупців. Збільшення числа покупців на ринку підвищує обсяги продажу товарів, в той же час, разом з ростом обсягів продажів, загострюється конкуренція між інтернет-продавцями за клієнта, що призводить до значного зростання рекламних витрат. В результаті, з розширенням торговельного ринку, зростаючі витрати на рекламу зменшують прибутки інтернет-магазинів, як наслідок, скорочується і кількість продавців [42]. Якщо порівняти об'єми ринку інтернет-реклами в Україні і розвинутих країн світу то вони в середньому на два порядки менші [43]. Даний факт вказує на те, що ринок інтернет-реклами в Україні, ще далекий від свого граничного насичення, коли збільшення рекламних витрат вже не приводить до збільшення цільової аудиторії і рівня готовності купувати розрекламовану продукцію, або іншими словами до підвищення прибутку. Також, в Україні основна частка учасників електронного торговельного ринку невеликі інтернет-

продавці, які не можуть забезпечити необхідний рівень інвестицій в маркетинг, рекламу і логістику. Ці два факти говорять про те, що рівень конкуренції в електронній комерції поки що достатньо низький і за допомогою ефективної рекламної компанії можна в досить короткі терміни значно покращити свої позиції на електронному торговельному ринку України.

Економічна ситуація в країні сповільнила розвиток ринку e-commerce, в 2015 році ринок електронної торгівлі України може скоротитися на 20-30% у порівнянні з 2014 роком [32, 36]. Війна і економічна криза вплинули негативно на індустрію інформаційних технологій, даний сектор економіки у 2014 році отримав в два рази менше зовнішніх інвестицій (40 млн доларів США), ніж у попередньому році [44]. На думку експертів [36], найбільше від поточної кризи постраждають онлайн гіпермаркети, найменше - невеликі онлайн-магазини, які можуть замінювати падіння продажів в одних товарних категоріях, введенням нових товарних груп. Також, не варто очікувати росту окремих сегментів e-commerce в 2015 році, скоріше будуть спостерігатися тенденції до збереження існуючих обсягів продажів. Разом з тим, в Україні намітилися позитивні тенденції, які дозволяють ринку e-commerce вийти із кризи з найменшими втратами, зокрема: онлайн покупці в основному вже готові до проведення фінансових операцій онлайн, через цю систему оплати щомісяця проходить більш 2-х мільйонів онлайн транзакцій від близько 800 тисяч унікальних користувачів [45] та підвищення довіри споживачів за рахунок більш широкого впровадження систем багаторівневих затримок електронних платежів.

Основна причина активного росту показників електронної торгівлі в Україні полягає в тому, що електронна комерція, як вид економічної діяльності, знаходиться на етапі розвитку, практично за відсутності значної

конкуренції [39]. В той же час, стосовно перспектив такого активного росту електронної комерції в Україні, в найближче десятиліття в Україні може відбутися насичення торговельного ринку електронними засобами торгівлі [42], торгівля в основному переміститься на онлайн майданчики і інтенсивність розширення електронного ринку значно знизиться.

ВИСНОВКИ ДО РОЗДІЛУ 1

В останні роки електронна комерція стрімко розвивається і поширюється, пропонуючи усе більш різноманітний асортимент товарів і послуг. Електронна комерція стає інструментом інтеграції окремих осіб, підприємств, галузей, державних установ і держав у єдине співтовариство, всередині якого взаємодія партнерів ефективно та безперешкодно реалізується засобами інформаційних і телекомунікаційних технологій.

Враховуючи викладене в першому розділі, можна зробити висновок, що електронна комерція — це діяльність економічних суб'єктів по здійсненню операцій купівлі-продажу товарів і послуг з використанням електронних інформаційно-комунікаційних засобів, спрямована на отримання прибутку і задоволення попиту споживачів. Фактично, це визначення повторює по змісту визначення поняття торгівля, із чого можна зробити висновок, що відмінностей між поняттями «електронна комерція» і «електронна торгівля» також бути не повинно. На нашу думку, з деякими припущеннями, можна поставити знак рівності між цими поняттями. При цьому хочеться ще раз підкреслити, що торгівля може відбуватися як товарами (матеріальними/нематеріальними), так і послугами (страхові, банківські й ін.).

Проведене аналітичне дослідження дозволило виділити наступні основні перешкоди на шляху динамічного розвитку ринку електронної комерції в Україні:

- недосконалість законодавства в області електронної торгівлі;
- невелике число користувачів Інтернету;
- недостатній розвиток комунікацій;
- нерозвиненість платіжних систем.

Дослідження перспектив розвитку електронної торгівлі в Україні показує, що «чиста» модель B2C роздрібною торгівлі через Інтернет поступово заміщується «багатоканальною» формою торгівлі, оскільки, «багатоканальні» торговельні площадки більш ефективно використовують наявні складські і торговельні потужності, залучають покупців в реальні і віртуальні магазини з меншими витратами, забезпечують більш високу лояльність клієнтів і сприяють збільшенню придбань споживачів.

Ефективне управління розвитком електронної комерції, а також процесами інформатизації економіки в цілому неможливо без повноцінного і всебічного економіко-математичного дослідження усього комплексу проблем. У процесі вдосконалення інструментарію дослідження необхідно зберігати послідовність із попередніми роботами, виконаними раніше, що дозволить одержати часові ряди, які характеризують процеси розвитку електронної комерції в Україні. Першочерговим завданням, що передує економіко-математичному дослідженню, є виявлення і розробка системи взаємозалежних показників електронної комерції. В цілому мова йде про розробку методології оцінки стану і розвитку електронної комерції в Україні, що може застосовуватися для аналізу складних закономірностей в електронній комерції і яка стає основою для розробки діючих механізмів ефективного управління і прийняття рішень на державному рівні.

Основна причина активного росту показників електронної торгівлі в Україні полягає в тому, що електронна комерція, як вид економічної діяльності, знаходиться на етапі розвитку, практично за відсутності значної конкуренції. Стосовно перспектив активного росту електронної комерції в Україні, в найближче десятиліття в Україні може відбутися насичення торговельного ринку електронними засобами торгівлі, торгівля в основному переміститься на онлайн майданчики і інтенсивність розширення електронного ринку значно знизиться.

РОЗДІЛ 2.

МЕТОДИКА ВИЗНАЧЕННЯ СПОЖИВЧИХ ПРІОРИТЕТІВ В ЕЛЕКТРОННІЙ ТОРГІВЛІ НА ОСНОВІ ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

2.1. Інтелектуальний аналіз даних як частина методики визначення споживчих пріоритетів в електронній торгівлі

Сучасну економіку часто називають економікою послуг, цей факт пояснюється тим, що сфера надання послуг становить більше 50% частки ВВП більшості розвинених країн світу [46]. Щорічно збільшуються масштаби споживання і номенклатура послуг, що передбачає наявність великого обсягу інформаційних потоків, ефективно обробляти які підприємства не можуть без впровадження і наступного використання сучасних програмно-інструментальних засобів, на основі інтелектуальних методів і конкретних бізнес-рішень.

Розвиток цивілізованого ринку електронної торгівлі пов'язаний в першу чергу із збільшенням конкуренції. Виробники товарів і постачальники послуг високотехнологічних галузей (телекомунікація, Інтернет-Провайдери, провайдери «хмарного сервісу» і т.д.) намагаються запропонувати нові вирішення проблем споживачів. Це забезпечує споживачам право вибору і формує різноманіття продуктів і послуг, а як наслідок, стабілізує поступальний розвиток соціально-економічної системи. Найбільш ефективною стратегією в електронній торгівлі буде побудова довгострокових довірчих відносин з усіма існуючими і потенційними клієнтами (з кожним з них), щоб вони знали про можливості та конкурентних перевагах продуктів і послуг бізнесу і могли б у потрібний момент ними скористатися без зайвих витрат та зусиль із боку

менеджерів. Очевидно, що є галузі економіки, у яких побудова подібних відносин з кожним із клієнтів є практично неможливою - клієнтів багато, розмір прибутків від кожного невеликий і підтримувати інформацію про кожного із них економічно не вигідно. Інформацію про товари народного споживання і продукти харчування легше донести до мільйонів споживачів через засоби масової інформації шляхом шаблонних рекламних пропозицій, оскільки вартість доставки кожному покупцеві інформації про подібні товари явно перевищить прибуток від їхньої покупки конкретним покупцем.

З іншого боку, для галузей, у яких конкуренція вийшла на рівень взаємодії з кожним клієнтом, встановлення і підтримка відносин з ним стає ключовою, якщо не головною конкурентною перевагою. В першу чергу це стосується надання корпоративних послуг, де кожний клієнт - це компанія зі складною системою внутрішніх відносин і постійно мінливими вимогами. Знання цих вимог, уміння налагодити стосунки з організацією в цілому та з людьми, що приймають рішення, стає на сьогодні необхідною умовою ефективного бізнесу.

Незалежно від того, чи займається підприємство виробництвом складного промислового устаткування, чи надає високотехнологічні послуги, кожний його клієнт має певну цінність, і ця цінність тем вище, чим стійкіше і довірливіше складаються відносини з компанією-клієнтом і її ключовими співробітниками. Незнання їх поточних потреб, невміння знайти персоналізований підхід до вирішення їх проблем призводить до того, що клієнти йдуть до більш гнучкого і лояльного конкурента.

У сфері роботи із приватними клієнтами можливість побудови персоналізованих відносин теж стає ключовою конкурентною перевагою, частиною психології ведення електронного бізнесу [47-49]. Споживач із більшим задоволенням віднесе свої засоби в банк, де в нього працює

знайомий банківський працівник, або купить автомобіль там, де в нього встановився контакт із менеджером з продажу, які б пояснили йому чесно і відверто переваги та недоліки кожного з випадків купівлі і запропонували оптимальні цінові умови.

У першу чергу це стосується дорогих покупок (транспорт, нерухомість, меблі, відпочинок і т.д.), однак фактор персоналізованих взаємовідносин з клієнтами, поступово переміщається в сферу більш масових покупок (зв'язок, розваги, служба побуту). Виходячи з методики взаємин «споживач-постачальник», все вище згадане, в рівній мірі, актуально і для ринку електронної торгівлі. В той же час, через високу динаміку ринку і значне збільшення кількості суб'єктів взаємодії, окрім основних каналів взаємодії постачальників і споживачів електронної торгівлі, виникає необхідність наявності каналу оперативного зворотного зв'язку.

Зворотній зв'язок, організовується за допомогою впровадження інтелектуальної інформаційної системи підтримки прийняття рішень, що складається із блоку інтелектуального аналізу даних і відповідних інструментальних засобів BRMS. Їх впровадження дозволяє споживачеві максимально впливати на пропозиції ринку, а постачальникові пропонувати в автоматичному режимі найбільш раціональні і клієнтоорієнтовані товари і послуги, що орієнтовані не тільки на загальну динаміку споживання товарів і послуг на ринку електронної торгівлі, але й динаміку споживання послуг окремими клієнтами та клієнтськими групами (кластерами) для об'єктивного виявлення виникаючих на ринку клієнтських вподобань.

Типові компанії, що працюють на ринку електронної торгівлі в механізмах свого функціонування містять велику кількість різних бізнес-процесів, і для близько 70% з них потрібна наявність оперативних рішень,

наприклад, у таких процесах, як процедури купівлі/продажу або оформлення послуги. До таких оперативних рішення можуть також відноситися питання: «чи раціональна дана послуга для даної групи споживачів», «чи раціональний набір послуг для клієнта з погляду його платоспроможності, профілю його внутрішнього споживання, розміру підприємства» і т.д.

У сукупності дані бізнес-рішення представляють собою важливий фактор успіху компанії - виконання корпоративної політики, забезпечення маневреності, оперативної реакції на зміни в бізнесі у порівнянні з конкурентами і партнерами на ринку електронної торгівлі. До того ж ринок електронної торгівлі є динамічним і регулярно поповнюється новими товарами і послугами: послуги для фізичних осіб, для малого бізнесу, партнерські програми для корпоративних клієнтів, VIP-клієнтів і інших юридичних осіб, що містять велику кількість умов і пропозицій у рамках кожної категорії. Виникає закономірна проблема вибору конкретного товару і послуги для конкретного профілю клієнта на ринку електронної торгівлі, який визначається набором його споживчих пріоритетів [50].

Таким чином, політика середньостатистичного пакета послуг для безлікої маси клієнтів стає неефективною для підтримання, або, тим більше, максимізації доходу з одного споживача (при мінімізації витрат). Відділу менеджменту послуг підприємства доводиться стикатися з новими умовами функціонування, конкуренцією у боротьбі за прихильність споживача [51, 52].

Вартість доставки товарів постійно змінюється, росте попит на нові інноваційні послуги, міняються склад і умови надання послуг в електронній торгівлі. Крім пошуку затребуваних товарів і більш раціональних послуг для групи клієнтів (з погляду споживача і

постачальника послуг), усі зміни необхідно вносити в положення, інструкції, бізнес-процеси, а потім, і в інформаційні системи.

Тому для підприємств електронної торгівлі, проблема формування клієнтських профілів з відповідними наборами споживчих вподобань, за своєю актуальністю виходить на перше місце, і вони починають переглядати свої стратегії в області інформаційних технологій і систем, шукати нові підходи для швидкого реагування на зовнішні та внутрішні зміни, з використанням адаптивних математичних моделей і сучасних інформаційних технологій [53].

Щоб успішно конкурувати з іншими підприємствами ЕТ, важливо знати своїх споживачів і мати повне уявлення про бажання та потреби поточних і потенційних клієнтів, вивчати і аналізувати ринок товарів послуг [53]. Для реалізації даної стратегії необхідно розділити сукупність клієнтів на кластери, кожний з яких буде відповідати унікальному для його кластера клієнтському профілю споживання [54].

Виникає необхідність у вивченні і розумінні потреб існуючих кластерів клієнтів з їхнім наступним аналізом і створенням принципово нової стратегії утримання та залучення нових клієнтів. Процес розподілу клієнтів на однорідні групи на основі узагальнених для кожного кластера атрибутів (звички, смаки, динаміка споживання товарів і послуг) повинен стати частиною стратегії і методики просування товарів і послуг підприємства на ринку електронної торгівлі.

Термін «кластеризація клієнтів» означає процес поділу клієнтів на однорідні групи на основі загальних атрибутів (звички, смаки, динаміка споживання і т.д.). «Особистий профіль клієнта» описує клієнта за його особистими атрибутами, такими як вік, стать, дохід і спосіб життя. «Профіль споживання клієнта» - часткове або відсоткове споживання ним товарів і послуг у середині клієнтського кластера, також може

вимірюватися в абсолютних величинах. Маючи ці два профіля, маркетингологи можуть приймати рішення, стосовно маркетингові дій по кожному кластеру клієнтів, а потім розподіляти ресурси між кластерами з метою вирішення конкретних бізнес-процесів [55].

2.2. Система характеристик клієнтів. Збір, підготовка і аналіз інформації про споживачів та їх активності на ринку електронної торгівлі

Найважливішим фактором в процесі інтелектуального аналізу даних, є розуміння даних та їх взаємозв'язків. Для отримання адекватного і якісно правильного опису клієнта вирішальне значення має вибір змінних (характеристик) [56]. Для визначення характеристик необхідно вибрати найменший набір змінних, що повноцінно описують поведінку покупця. Допомогти в процесі визначення характеристик можуть такі ключові слова:

1. **Як?:** Як клієнт сприяє створенню реєстраційного запису даних про купівлю товару? Він робить голосове замовлення по телефону, здійснює замовлення засобами сайту інтернет-магазину або використовує для цього e-mail?
2. **Кому?:** Кому робить замовлення клієнт? Перелік потенційно цікавих для покупця інтернет-магазинів;
3. **Який?:** Який тип товару (види товарів) цікавий для покупця?
4. **Де?:** Де перебувають клієнт і інтернет-магазин? Чи може він робити міжнародні замовлення товарів і послуг?
5. **Коли?:** Коли клієнт робить замовлення? Клієнт може робити виклики в денний робочий час або в неробочий час увечері, уночі чи у вихідні;

6. **Куди?:** Куди клієнт замовляє доставку? Він може замовляти доставку в іншу країну?

7. **Як довго?:** Скільки часу триває замовлення?

8. **Як часто?:** Як часто клієнт робить і підтверджує замовлення?

Використовуючи ці ключові слова та критерії, що пропонуються у літературних джерелах [57, 58], побудуємо список характеристик, які можуть застосовуватися в якості короткого опису клієнта по споживанню ним послуг інтернет-магазину (у відносному та абсолютному значенні):

Механізми замовлення:

1. Замовлення в своєму (домашньому) регіоні (середня кількість замовлень в одиницю години);

2. Замовлення через сайт інтернет-магазину (середня кількість замовлень в одиницю години);

3. Замовлення через стаціонарні телефони (середня кількість замовлень в одиницю години);

4. Замовлення через мобільні телефони (середня кількість замовлень в одиницю години);

5. Замовлення через e-mail;

6. Замовлення в інших регіонах України (середня кількість замовлень в одиницю години);

7. Міжнародні замовлення (середня кількість замовлень в одиницю години).

Послуги, що приносять прибуток:

8. Обсяги замовлень (загальна сума в одиницю часу);

9. Обсяги замовлень по каналах замовлень (загальна сума для кожного окремого каналу в одиницю часу);

10. Обсяги замовлень по типу доставки (загальна сума для кожного окремого типу в одиницю часу);

11. Обсяги замовлень за місцем доставки (загальна сума для кожного окремого типу місця доставки в одиницю часу);
12. Обсяги замовлень по виду товарів (загальна сума для кожного окремого виду товару в одиницю часу);
13. Обсяги замовлень по конкретним інтернет-магазинам (загальна сума для кожного окремого інтернет-магазину);
14. Обсяги замовлень по виду оплати (загальна сума для кожного окремого виду оплати);
15. Тривалість замовлення (середня тривалість одного замовлення в одиницях часу за визначений період);
16. Кількість замовлень (середня кількість в одиницю часу);
17. Замовлення в робочі дні тижня (виміряються у відсотках);
18. Замовлення в денний час, в години з 9:00 - 18:00 год. (виміряються у відсотках);
19. Завершені покупки (виміряється у відсотках);
20. Співвідношення між завершеними і незавершеними покупками;
21. Співвідношення між замовленнями в денний - не денний час та в робочі – не робочі дні тижня.
22. Співвідношення між покупками (завершеними замовленнями) в денний - не денний час та в робочі – не робочі дні тижня.

Більшість із 22 перерахованих вище характеристик можуть генеруватися безпосередньо за допомогою вихідних даних, отриманих із сховища даних конкретного інтернет-магазину, але для деяких характеристик потрібен творчий підхід і виконання певних дій з даними [59]. Наприклад, клієнти, які користуються телефоном тільки на робочому місці, і клієнти, які також використовують телефон в особистих цілях, можуть перебувати в різних кластерах.

У цьому випадку кластеризація ґрунтується на відсоткових значеннях викликів в робочі дні тижня і в денний час. Ясно, що генерування корисних характеристик, включаючи зведені характеристики, є важливим фактором в процесі ІАД. Якщо генеруються неправильні характеристики, то інтелектуальний аналіз даних закінчиться невдачею.

Профілі клієнтів є базовою основою для встановлення ефективної взаємодії з існуючими клієнтами з метою збереження клієнтів, надання їм більш якісних послуг та залучення нових клієнтів. Створення профілів клієнтів здійснюється шляхом синтезу зібраної інформації про клієнтів, зокрема, демографічні та персональні дані. Вибудовані профілі клієнтів також можуть застосовуватися для пошуку нових клієнтів з використанням характерних механізмів взаємодії типових для даного профілю клієнтів, наприклад, на основі демографічних даних, отриманих з різних джерел [60-62]. Ці дані використовуються для встановлення зв'язку із кластерами клієнтів, які формуються на початковому етапі.

Це дозволяє формувати для кожного профілю клієнта (що представляє собою притаманну тільки цьому профілю комбінацію демографічних і персональних даних) відповідні кластери (дані про споживання послуг електронної торгівлі) і навпаки. Для кожного профілю може бути виконана оцінка поведінки при використанні клієнтами засобів і механізмів електронної торгівлі під час здійснення замовлення і купівлі товару. В залежності від конкретної мети, необхідно вибирати профілі, що мають безпосереднє відношення до проекту. Звичайний профіль клієнта представляє собою файл, що містить, принаймні, інформацію про вік і стать.

Якщо потрібні профілі клієнтів по конкретних товарах і послугах, то файл буде містити інформацію про товар (послугу) і/або суму витрачених

коштів. Характеристики клієнтів, які можуть використовуватися для створення профілів, розглянуті в роботах [63]:

1. Географічні відомості. Чи згруповані клієнти на регіональному, національному або глобальному рівнях?
2. Культурні та етнічні характеристики. Якою мовою розмовляють клієнти? Чи має місце етнічний вплив на вподобання клієнтів або їх поведінку при здійсненні купівлі товарів?
3. Економічні умови, дохід і/або купівельна спроможність. Який середній дохід або купівельна спроможність клієнтів? Чи є в них які-небудь труднощі, пов'язані з оплатою? Яку кількість коштів і як часто клієнт витрачає на кожний продукт?
4. Вік і стать. Яка вікова група є переважною серед ваших цільових клієнтів? Скільки в сім'ї дітей і якого вони віку? Серед клієнтів, що використовують певний продукт або послугу, переважають жінки або чоловіки?
5. Цінності, погляди і переконання. Яке відношення клієнтів до вашого товару або послуги?
6. Життєвий цикл. Наскільки регулярно клієнт купує у вас продукцію і користуються вашими послугами?
7. Знання і розуміння. Які знання є в клієнтів щодо товарів, послуг, або галузі? Який потрібен рівень освіти? Яка необхідна реклама брэнда, для того щоб клієнти були інформовані про пропозиції на ринку електронної торгівлі?
8. Спосіб життя. Скільки є корисних характеристик, що стосуються способу життя покупців?
9. Метод залучення клієнта. Як здійснювалося залучення клієнта?

Вибір характеристики залежить також від наявності даних про клієнтів. За допомогою цих характеристик може бути створена модель

оцінки клієнта (профіль клієнта). З цією метою може використовуватися метод інтелектуального аналізу даних, що називається методом опорних векторів (SVM).

На рис. 2.1 наведено розроблену структуру клієнтів підприємства електронної торгівлі [59]. Клієнти компанії можуть бути поділені на корпоративних і приватних. В свою чергу поділимо клієнтів на два основних профілі: прогресивний і консервативний.

Прогресивний профіль споживання характеризується високим споживанням товарів і послуг на ринку електронної торгівлі. Консервативний профіль споживання характеризується помірним і низьким споживанням товарів і послуг на ринку електронної торгівлі.

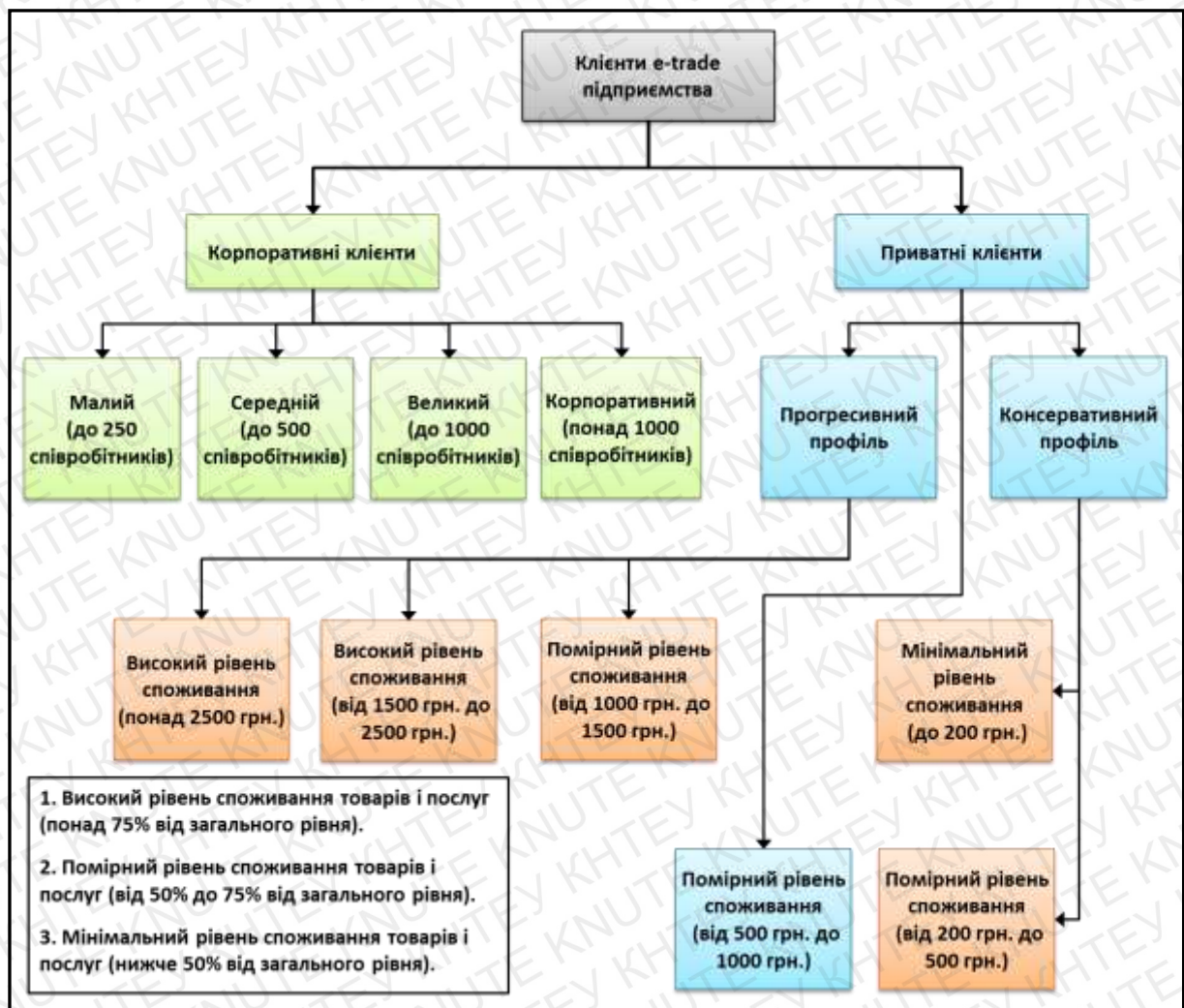


Рис. 2.1. Структура клієнтів підприємства електронної торгівлі

Також крім профілів необхідно виділити «золоту середину» - користувачів, що споживають помірну кількість товарів і послуг на ринку електронної торгівлі. Крім цього, у відповідності із профілями клієнтів, розділяються і їхні витрати на товари і послуги (до 200 грн. на місяць, проміжні значення від 500 до 2500 грн. на місяць, і більше 2500 грн. на місяць). Корпоративні клієнти діляться на чотири групи в залежності від розміру компанії та кількості співробітників: малий, середній, великий бізнес і корпорації [60-62]. Рахунок корпорації за спожиті послуги виписується на юридичну особу. Найчастіше, якщо мова йде про особисті, конфіденційні дані, вони бувають недоступні в повному обсязі. Інформація про спосіб життя та доходи клієнтів відсутня. Проте, можна віднайти деяку інформацію в БД підприємства електронної торгівлі. Інформація, яка знаходиться в БД, містить в собі ім'я та адресу, а також інші відомості, наприклад, план обслуговування, інформацію про контракт і телефонний номер. Володіючи цими даними, можна використовувати такі показники для створення профілів клієнтів [59]:

1. Вікова група: «<25, 25-40, 40-55, >55 років»;
2. Стать: чоловіча, жіноча;
3. Тип телефону: мобільний, стаціонарний;
4. Тип приватного клієнта: консервативний, помірний, прогресивний;
5. Розмір корпоративного клієнта: малий, середній, великий, корпорація (дані про фіз. особу контактера);
6. Місце проживання: велике місто, невелике місто.

Оскільки відносно невелика різниця у віці клієнтів може означати схожу поведінку споживача, клієнти повинні групуватися за віком. Інакше результат алгоритму класифікації є занадто специфічним для формування бази знань [64]. Основна ціль групування показників полягає в зменшенні їх числа до керованої кількості та усуненні кореляції між показниками.

Склад груп повинен вибиратися з особливою ретельністю. Надзвичайно важливо, щоб розміри груп, по можливості, були практично рівними [65]. Якщо будуть групи, кількість клієнтів у якій перевищує число клієнтів в інших групах, це приведе до зниження ефективності класифікації. Цей факт пов'язаний з тим, що в цій групі презентовано відносно більша кількість клієнтів з кожного кластера. Тому на основі цієї характеристики визначити кластер клієнта буде неможливо.

Для оцінки кластерів клієнтів застосовується метод k-середніх з використанням профілів клієнтів. Далі безпосередньо розглянемо методи кластерного аналізу, які застосовуються для вивчення споживчих пріоритетів клієнтів.

2.3. Методи кластеризації споживачів ринку електронної торгівлі

Необхідно зазначити, що мета кластерного аналізу полягає в поділі об'єктів на групи з урахуванням подібностей об'єктів. Кластеризацію можна вважати найбільш важливим в інтелектуальному аналізі даних методом навчання без учителя [66]. Як і будь-який інший метод навчання без учителя, кластеризація не використовує ідентифікатори класу для виявлення основної структури в зборі даних.

Кластер може бути визначений як сукупність об'єктів, які є «подібними» один до одного і «неподібними» для об'єктів, що належать до інших кластерів (рис. 2.2). В даному випадку (рис. 2.2) легко ідентифікуються 3 кластера, на які можуть бути поділені дані. Критерієм подібності є відстань: два або більш об'єкта належать одному кластеру, якщо вони є «близькими» відповідно до заданої відстані (в даному випадку геометричної відстані). Така кластеризація називається кластеризацією на основі відстані.

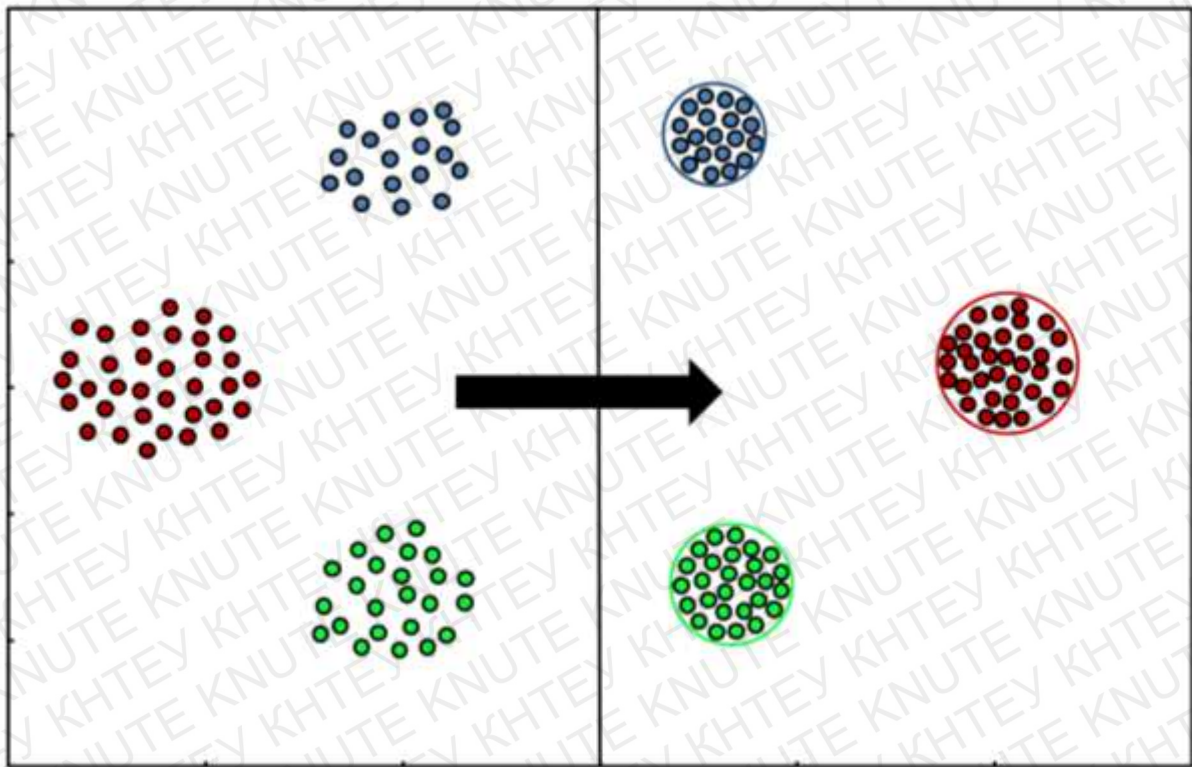


Рис. 2.2. Приклад процесу кластеризації даних

Іншим методом кластеризації є концептуальна кластеризація. В цьому методі два або більше об'єкти належать одному кластеру, якщо він визначає концепцію, загальну для об'єктів. Інакше кажучи, об'єкти групуються у кластери відповідно до заданої концепції, а не відповідно до простої подібності (наприклад кольору, рис. 2.2).

Визначення кластера може бути сформульоване по-різному, в залежності від мети кластеризації. Можна дати загальне визначення кластера, згідно з яким кластер являє собою групу об'єктів, які більше подібні один до одного, ніж до представників інших кластерів. Термін «подібність» може сприйматися як математична подібність, що вимірюється в деяких певних параметрах.

В метричних просторах подібність часто визначається за допомогою відстані (абсолютної, відносної або конкретних значень). Відстань можна вимірювати по-різному. Перший спосіб полягає в проведенні вимірів серед

векторів даних. Другий спосіб пов'язаний з вимірюванням відстані від вектора даних до деякого типового об'єкта кластера. Центри кластерів, як правило, [67-69] (і в даному дослідженні також) невідомі апіорі і будуть визначатися за допомогою алгоритмів кластеризації одночасно з розподілом даних. Центрами кластерів можуть бути вектори такого ж розміру, як об'єкти даних. Також центри кластерів можуть визначатися геометричними об'єктами «більш високого рівня», наприклад, лінійними або нелінійними підпросторами або функціями. Дані можуть формувати кластери з різними геометричними формами, розмірами і щільністю, як це показано на рис. 2.3, 2.4. Кластери можуть бути сферичними, витягнутими та порожніми. Кластери можуть існувати в будь-якому n -вимірному просторі. Кластери a , c і d можна охарактеризувати як лінійні та нелінійні підпростори в просторі даних (в даному випадку просторі R^2). Алгоритми кластеризації здатні виявляти підпростори у просторі даних.

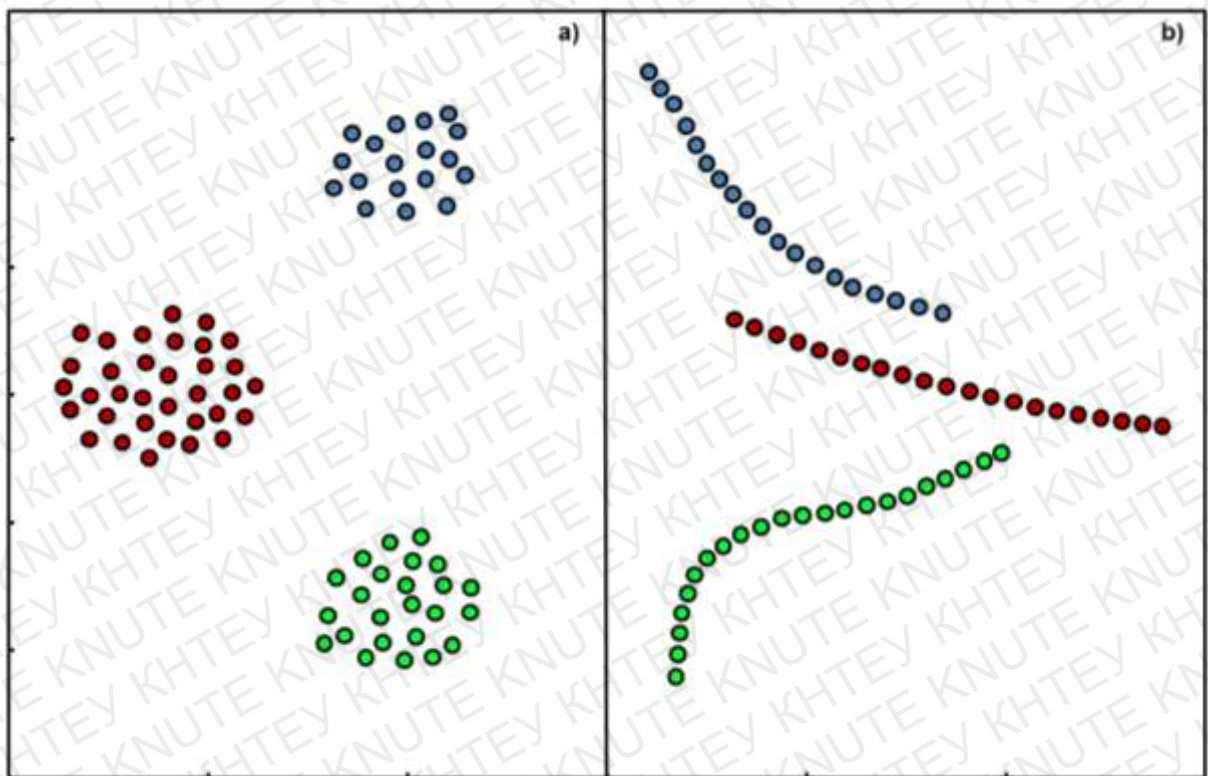


Рис. 2.3. Кластери сферичної (а) та витягнутої (б) форми

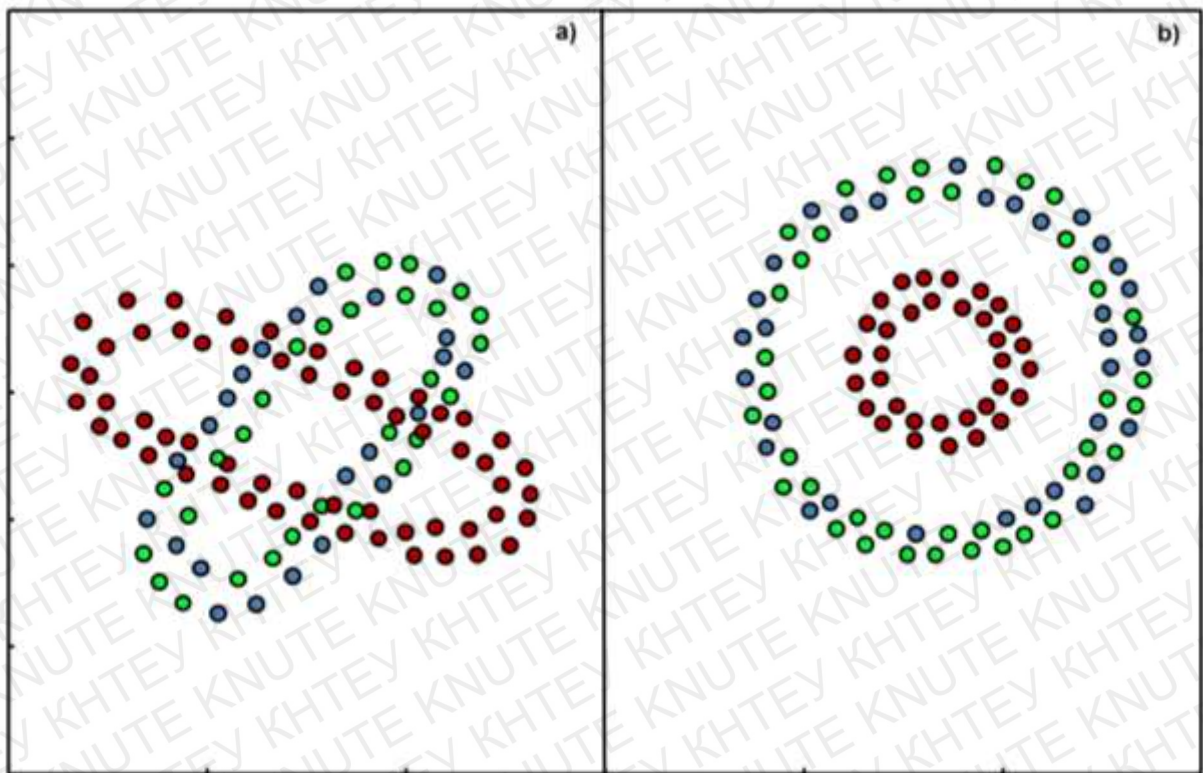


Рис. 2.4. Поржні кластери овальної (а) та колоподібної (б) форми

Ефективність більшості відомих алгоритмів кластеризації даних залежить не тільки від геометричної форми та щільності окремих кластерів, але також від положення в просторі і відстані між кластерами. Кластери можуть бути розділені, пов'язані один з одним або накладені один на одного. Формально кластери можуть розглядатися як підмножини в множинах даних. Можна виділити два можливі результати класифікації методів кластеризації, підмножини можуть бути нечіткими або чіткими.

Методи чіткої кластеризації засновані на класичній теорії множин, яка вимагає, щоб об'єкт або належав, або не належав кластеру. Чітка кластеризація в множині даних X означає розбивку даних на задану кількість підмножин множини X . Число підмножин (кластерів) позначаємо символом s .

Методи нечіткої кластеризації дозволяють об'єктам належати одночасно декільком кластерам з різним ступенем приналежності.

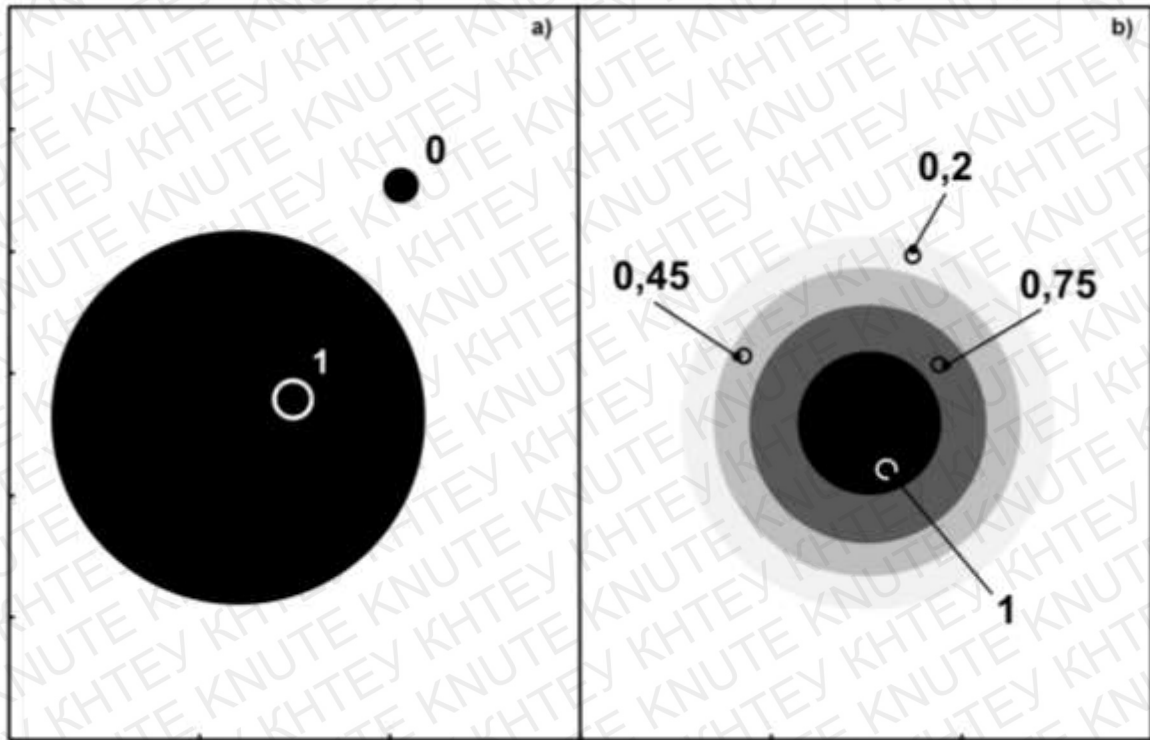


Рис. 2.5. Різні форми кластерів і ступінь приналежності об'єктів в R^2 -просторі. а) – чітка кластеризація, б) – нечітка кластеризація

Таким чином, множина даних X розбивається на нечіткі підмножини. У багатьох реальних ситуаціях нечітка кластеризація є більш природною, ніж чітка кластеризація, оскільки об'єктам, які перебувають на границях між декількома класами, не визначається належність до одного із класів, при цьому їм також властива міграція між кластерами. Таким об'єктам задається ступінь приналежності між 0 і 1 , що визначає їх часткову приналежність (див. рис. 2.5).

Дискретний характер чіткої розбивки приводить до неможливості аналітичного і алгоритмічного розв'язку алгоритмів, що ґрунтуються на аналітичних функціоналах, оскільки ці функціонали не є диференційованими.

Матриця розбивки кластерів $Y=[\mu_{ik}]$ ($i=1,2, \dots, N, k=1,2, \dots, c$) має наступний вигляд:

$$Y = \begin{pmatrix} \mu_{11} & \dots & \mu_{1k} & \dots & \mu_{1c} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{i1} & \dots & \mu_{ik} & \dots & \mu_{ic} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{N1} & \dots & \mu_{Nk} & \dots & \mu_{Nc} \end{pmatrix}, \quad (2.1)$$

Мета чіткої кластеризації полягає в розбивці множини даних X на кластери c . Припустимо, що значення c є відомим, наприклад, на основі попередніх даних, або представляє собою пробне значення, результати розбивки якого повинні бути перевірені. Використовуючи класичні множини, чітку розбивку можна розглядати як сімейство підмножин $\{A_i | 1 \leq i \leq c \subset P(X)\}$ з наступними властивостями:

$$\bigcup_{i=1}^c A_i = X, \quad (2.2)$$

$$A_i \cap A_j, 1 \leq i \neq j \leq c, \quad (2.3)$$

$$\emptyset \subset A_i \subset X_i, 1 \leq i \leq c, \quad (2.4)$$

Зазначені вище умови означають, що підмножини A_i містять усі дані множини X , вони повинні не перетинатися і жодне із них не є порожнім та не може містити усі дані множини X . Функції, виражені при врахуванні приналежності:

$$\bigvee_{i=1}^c \mu_{A_i} = 1, \quad (2.5)$$

$$\mu_{A_i} \vee \mu_{A_j}, 1 \leq i \neq j \leq c, \quad (2.6)$$

$$0 \leq \mu_{A_i} < 1, 1 \leq i \leq c, \quad (2.7)$$

де μ_{A_i} являє собою характеристичну функцію підмножини A_i , значення якої дорівнює нулю або одиниці. Для спрощення запису, замість

μ_{A_i} будемо використовувати μ_i . Використання позначення μ_{ik} замість $\mu_i(x_k)$ дозволяє представити розбивки в матричному вигляді $U = [\mu_{ik}]$, матриця N_{xc} , є представленням чіткої розбивки тоді й тільки тоді, коли її елементи задовольняють наступній умові:

$$\mu_{ik} \in \{0, 1\}, 1 \leq i \leq N, 1 \leq k \leq c, \quad (2.8)$$

$$\sum_{k=1}^c \mu_{ik} = 1, 1 \leq i \leq N, \quad (2.9)$$

$$0 < \sum_{i=1}^N \mu_{ik} < N, 1 \leq k \leq c, \quad (2.10)$$

Простір чіткої розбивки на кластери може бути визначений із наступних положень: нехай X – кінцева множина даних, а кількість кластерів $2 \leq c < N \in \mathbb{N}$. Тоді простір чіткої розбивки для X , можна задати таким чином:

$$M_{hc} = \left\{ U \in R^{N_{xc}} \mid \mu_{ik} \in \{0, 1\}, \forall i, k; \sum_{k=1}^c \mu_{ik} = 1, \forall i; 0 < \sum_{i=1}^N \mu_{ik} < N, \forall k \right\}, \quad (2.11)$$

Нечітка розбивка може бути визначена як узагальнення чіткої розбивки, в цьому випадку μ_{ik} може набувати всіх дійсних значень від 0 до 1. Розглянемо матрицю $U = [\mu_{ik}]$, що містить нечіткі розбивки, при цьому задані наступні умови:

$$\mu_{ik} \in [0, 1], 1 \leq i \leq N, 1 \leq k \leq c, \quad (2.12)$$

$$\sum_{k=1}^c \mu_{ik} = 1, 1 \leq i \leq N, \quad (2.13)$$

$$0 < \sum_{i=1}^N \mu_{ik} < N, 1 \leq k \leq c, \quad (2.14)$$

Варто звернути увагу на те, що є всього одна відмінність від умов чіткої розбивки (2.8). Крім того, визначення простору нечіткої розбивки мало відрізняється від визначення простору чіткої розбивки. Простір нечіткої розбивки може бути визначений у такий спосіб: нехай X – кінцева множина даних, а кількість кластерів $2 \leq c < N \in N$. Тоді простір нечіткої розбивки для X можна задати в такий спосіб:

$$M_{hc} = \left\{ U \in R^{N \times c} \mid \mu_{ik} \in [0, 1], \forall i, k; \sum_{k=1}^c \mu_{ik} = 1, \forall i; 0 < \sum_{i=1}^N \mu_{ik} < N, \forall k \right\}, \quad (2.15)$$

Тобто i -й стовпець матриці U містить значення функцій приналежності i -ої нечіткої підмножини в множині даних X . Відповідно до рівняння (2.13) сума кожного стовпця повинна бути рівною 1. Це означає, що сумарна приналежність усіх x_k в X дорівнює одиниці. Ніяких обмежень на розподіл приналежностей серед нечітких кластерів не існує. Дане дослідження сфокусовано на чіткій розбивці. Проте, алгоритми нечіткої кластеризації також застосовуються. Для того щоб вирішити проблему нечітких приналежностей, в якості кластера, якому належить об'єкт, буде обиратися кластер з найвищим ступенем приналежності. Результатом використання даного методу стане чітка розбивка кластерів. Окрім цього, в дослідженні також будуть застосовані методи чіткої кластеризації даних.

При аналізі початкових даних дослідження, також використовуються наступні алгоритми кластеризації:

1. Алгоритм К-середніх. Алгоритм К-середніх (K-means) є одним із найбільш простих алгоритмів навчання без учителя, що застосовується для вирішення проблеми кластеризації даних. Проте, результати цього методу чіткої розбивки не завжди надійні, а сам алгоритм має ряд проблем. Процедура являє собою простий спосіб класифікації заданої множини

даних $N \times n$ для заздалегідь визначеної кількості кластерів c . Алгоритм К-середніх відносить кожен елемент даних до одного із кластерів c для мінімізації суми квадратів:

$$\sum_{i=1}^c \sum_{k \in A_i} \|x_k - v_i\|^2, \quad (2.16)$$

В даному випадку, A_i – являє собою множину елементів даних у кластері i , а v_i – є середнім значенням елементів даних у кластері i . Варто звернути увагу на те, що квадрат різниці $\|x_k - v_i\|^2$ фактично є нормою вибраної відстані. В алгоритмах кластеризації v_i – є також центром (називається також прототипом) кластера i :

$$v_i = \frac{\sum_{k=1}^{N_i} x_k}{N_i}, \quad x_k \in A_i, \quad (2.17)$$

де N_i – кількість елементів даних в кількість точок даних в A_i .

2. Алгоритм К-медоїд. В алгоритмі кластеризації К-медоїд (K-medoids), як і в інших алгоритмах чіткої розбивки, використовуються ті ж самі рівняння, що і в алгоритмі К-середніх. Різниця полягає лише в тому, що в алгоритмі кластеризації К-медоїд центрами кластерів є елементи даних найближчі до середнього значення даних у кластері $V = \{v_i \in X \mid 1 \leq i \leq c\}$. Цей метод може бути корисний, у випадку коли, наприклад, в просторі даних відсутня неперервність. Це означає, що середнього значення в одному кластері фактично не існує.

3. Алгоритм нечітких С-середніх. Алгоритм нечітких С-середніх (Fuzzy C-Means - FCM) мінімізує цільову функцію, що називається функціоналом С-середніх, для визначення кластерів. Функціонал С-середніх J визначається по формулі, яка була запропонована Бездеком і Даном у роботі [70]:

$$J(X;U;V) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m \|x_k - v_i\|_A^2, \quad (2.18)$$

де

$$V = [v_1, v_2, \dots, v_c], v_i \in R^n, \quad (2.19)$$

В формулі (2.18) V – позначає вектор із центрами кластерів, які повинні бути визначені. Нормою відстані $\|x_k - v_i\|_A^2$ - називають квадрат норми, що утворений скалярним добутком і визначається по формулі:

$$D_{ikA}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^T A(x_k - v_i), \quad (2.20)$$

З погляду статистики, рівняння (2.18) визначає загальну кількість дисперсій x_k від v_i . Мінімізацію функціонала С-середніх можна розглядати як нелінійну задачу оптимізації, яка може бути вирішена за допомогою різних методів. Прикладами методів, що дозволяють вирішувати нелінійні задачі оптимізації, є групова координатна мінімізація і генетичний алгоритм. Найбільш простим способом розв'язку цієї задачі є використання ітерації Пікара [71] через умови першого порядку для стаціонарних точок рівняння (2.18). Цей метод називається алгоритмом оптимізації нечітких С-середніх. Для знаходження стаціонарних точок функціонала С-середніх можна для J , застосувати вираз (2.13) з використанням множників Лагранжа:

$$J(X;U;V;\lambda) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ikA}^2 + \sum_{k=1}^N \lambda_k \left(\sum_{i=1}^c \mu_{ik} - 1 \right), \quad (2.21)$$

а також? заданням нульових градієнтів (J) по відношенню до U , V і X . В цьому випадку, при $D_{ikA}^2 > 0$, \forall_i, k і $m > 1$, функціонал С-середніх може бути мінімізований за допомогою $(U, V) \in M_{fc} \times R^{n \times c}$, якщо

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c (D_{ikA} / D_{jkA})^{2/(m-1)}}, 1 \leq i \leq c, 1 \leq k \leq N, \quad (2.22)$$

і

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^m x_k}{\sum_{k=1}^N \mu_{ik}^m}, 1 \leq i \leq c, \quad (2.23)$$

Розв'язок цих рівнянь задовольняє обмеженням, які були задані в рівняннях (2.12) і (2.14). Відзначимо, що v_i у рівнянні (2.23) є середньозваженим значенням елементів даних, що належать кластеру, а вага елемента представляє собою ступінь приналежності до кластеру. Це пояснює, чому алгоритм називається саме «С-середніх». Алгоритм нечітких С-середніх фактично є ітерацією між рівняннями (2.22) і (2.23). Алгоритм нечіткої кластеризації FCM (Fuzzy C-Means) використовує для обчислень стандартну евклідову відстань. Тобто, даний алгоритм здатний визначати гіперсферичні кластери. Варто звернути увагу на той факт, що алгоритм може виявляти тільки кластери, що мають однакову форму, яка є результатом вибору норм-створюючої матриці $A = 1$. У якості норм-створюючої матриці може бути обрана діагональна матриця $n \times n$ виду:

$$A_D = \begin{pmatrix} \left(\frac{1}{\sigma_1}\right)^2 & 0 & \dots & 0 \\ 0 & \left(\frac{1}{\sigma_2}\right)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \left(\frac{1}{\sigma_n}\right)^2 \end{pmatrix}, \quad (2.24)$$

Ця матриця враховує дисперсії по координатних осях для X . Іншим варіантом є вибір A в якості матриці, оберненої до коваріаційної $n \times n$, виду $A = F^{-1}$, де

$$F = \frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})(x_k - \bar{x})^T, \quad (2.25)$$

де \bar{x} - середнє значення даних. В цьому випадку матриця A заснована на нормі відстані Махаланобіса [26]: $D^2 = (\mu_1 - \mu_2)^T \sum^{-1} (\mu_1 - \mu_2)$, де μ_1 і μ_2 – вектори середніх, а \sum^{-1} – коваріаційна матриця (вважається однаковою для обох розподілів).

4. Алгоритм Густафсона-Кесселя. Алгоритм Густафсона-Кесселя (GK) є варіацією алгоритму нечітких С-середніх [72]. В ньому використовується інша і до того ж адаптивна норма відстані для розпізнавання геометричних форм в даних. Кожний кластер буде мати власну норм-створюючу матрицю A_i , що задовольняє наступній нормі, яка породжується скалярним добутком:

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), 1 \leq i \leq c, 1 \leq k \leq N, \quad (2.26)$$

Матриці A_i використовуються в якості параметрів оптимізації в функціоналі С-середніх. Це означає, що для кожного кластера норма відстані може адаптуватися відповідно до локальної топологічної структури даних. В даному випадку, c -кортеж норм-створюючих матриць визначається за допомогою \mathbf{A} , де $\mathbf{A} = (A_1, A_2, \dots, A_c)$. Тому, цільовий функціонал алгоритму Густафсона-Кесселя може бути розрахований у такий спосіб:

$$J(X; U; V; A) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ikA}^2, \quad (2.27)$$

Якщо A є фіксованим, то умови (2.12), (2.13) і (2.14) можуть застосовуватися без будь-яких обмежень. В той же час, рівняння (2.27) не може бути прямо мінімізоване, оскільки воно є лінійним відносно A_i . Це

означає, що J можна зробити як завгодно малим, роблячи A_i менш позитивно визначеним.

Щоб уникнути цього, A_i необхідно обмежити для одержання припустимого розв'язку. Основний спосіб вирішення даної задачі полягає в обмеженні визначника матриці. Взаємозв'язок між A_i з фіксованим визначником та оптимізацією кластера з фіксованим об'ємом виглядає в таким чином:

$$\|A_i\| = p_i, p > 0, \quad (2.28)$$

де p_i – об'єм i -го кластера, p – залишкова константа для кожного кластера.

При застосування множника Лагранжа, A_i можна виразити у такий спосіб:

$$A_i = [p_i \det(F_i)]^{1/n} F_i^{-1}, \quad (2.29)$$

$$F_i = \frac{\sum_{k=1}^N (\mu_{ik})^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^N (\mu_{ik})^m}, \quad (2.30)$$

де F_i – називається нечіткою коваріаційною матрицею. Тобто, це рівняння (2.30) в комбінації з рівнянням (2.29) можна підставити в рівняння (2.26). В результаті виразом для норми, що утворюється скалярним добутком (2.26), є узагальнений квадрат норми Махаланобіса між елементами даних і центром кластера. Ваговий коефіцієнт коваріації визначається ступенем приналежності до кластера.

5. Алгоритм Гаф-Гева.

В своїй роботі Гаф і Гева [73] запропонували нечіткий алгоритм оцінки максимальної правдоподібності (FMLE - fuzzy maximum likelihood estimation, правдоподібність найближчого сусіда) з відповідною нормою відстані, що визначається за формулою 2.31:

$$D_{ik}(x_k, v_i) = \left(\frac{\sqrt{\det(F_{wi})}}{\alpha_i} \right)^{\frac{1}{2}(x_k - v_i)^T F_{wi}^{-1} (x_k - v_i)}, \quad (2.31)$$

У порівнянні з алгоритмом Густафсона-Кесселя, норма відстані містить у собі експонентний член. Це означає, що дана норма відстані буде зменшуватися швидше, ніж норма, породжена скалярним добутком [71]. В цьому випадку нечітка коваріаційна матриця F_i , буде визначатися за такою формулою:

$$F_{wi} = \frac{\sum_{k=1}^N (\mu_{ik})^w (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^N (\mu_{ik})^w}, \quad 1 \leq i \leq c, \quad (2.32)$$

Причиною використання змінної w є узагальнення виразу (2.32). В оригінальному алгоритмі FMLE $w = 1$. В даному дослідженні змінна w приймається рівною 2, з метою компенсації впливу експонентного члена і одержання більш нечітких кластерів. Внаслідок узагальнення утворюються дві зважені коваріаційні матриці. Змінна α_i , у рівнянні (2.31) представляє собою ймовірність вибору кластера і визначається в такий спосіб:

$$\alpha_i = \frac{1}{N} \sum_{k=1}^N \mu_{ik}, \quad (2.33)$$

В роботі [73] встановили, що алгоритм FMLE здатний виявляти кластери різних форм, розмірів та густини, і що кластери не обмежені в об'ємі. Головним недоліком алгоритму Гаф-Гева є стійкість, оскільки експонентна норма відстані може сходиться в локальний оптимум.

Таким чином, з проведеного аналізу можна зробити висновок, що вибір алгоритмів кластеризації повинен здійснюватися на основі застосування методів перевірки, візуалізації і розпізнавання даних кластеризації.

2.4. Обґрунтування вибору методів перевірки і візуалізації даних кластеризації.

Задача перевірки процесу кластеризації пов'язана з визначенням і оцінкою правильності розбивки та вибору кількості кластерів. Алгоритм кластеризації призначений для параметризації кластерів з метою забезпечення найкращої відповідності. Проте, найкраща відповідність не завжди є найбільш ефективним рішенням. Кількість кластерів може виявитися неправильною, або ж форми кластерів можуть не відповідати фактичним групам у наявних початкових даних. В найгіршому варіанті дані взагалі неможливо згрупувати. Загалом, можна виділити два основні способи визначення належної кількості кластерів у даних:

- Почати з досить великої кількості кластерів і послідовно зменшувати їхню кількість шляхом об'єднання кластерів з однаковими властивостями;
- Поєднувати в кластери дані для різних значень центрів кластерів c і перевіряти правильність отриманих кластерів, використовуючи методи перевірки.

Для застосування другого способу необхідно розробити методи перевірки. В наявній літературі пропонуються різні методи перевірки, проте, жоден з них не є ідеальним. Тому в даному дослідженні використовується ряд параметрів, які описуються нижче:

1. Коефіцієнт розподілу (PC) - вимірює кількість «перекриттів» між кластерами. Згідно з визначенням Бездека [74, 75], визначається у такий спосіб:

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (u_{ij})^2, \quad (2.34)$$

де u_{ij} – приналежність елемента даних j в кластері i . Головним недоліком цього методу перевірки є відсутність прямого зв'язку з наявними даними. Оптимальний показник визначається мінімальним значенням коефіцієнта.

2. Класифікаційна ентропія (CE) - вимірює лише розмитість кластера і має незначні відмінності від коефіцієнта розподілу. Оптимальний показник визначається максимальним значенням коефіцієнта. Визначається за формулою:

$$CE(c) = -\frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N u_{ij} \log(u_{ij}), \quad (2.35)$$

3. Індекс розподілу (PI) - показує відношення сумарних значень компактності і поділу кластерів. Кожний окремих кластер вимірюється за допомогою методу перевірки кластерів. Значення нормується діленням на нечітку потужність кластера. Для одержання індексу розподілу використовується сумарне значення по всіх кластерах. Індекс розподілу визначається у такий спосіб:

$$PI(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (u_{ij})^m \|x_j - v_i\|^2}{N_i \sum_{k=1}^c \|v_k - v_i\|^2}, \quad (2.36)$$

PI в основному використовується для порівняння різних розбивок з однаковою кількістю кластерів. Мінімальне значення індексу визначає найкращу розбивку.

4. Індекс поділу (SI) - на відміну від індексу розподілу (PI), індекс поділу використовує для перевірки розбивки поділ з мінімальною відстанню. Оптимальний показник визначається мінімальним значенням індексу. Індекс розраховується за формулою:

$$SI(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (u_{ij})^2 \|x_j - v_i\|^2}{N \min_{i,k} \|v_k - v_i\|^2}, \quad (2.36)$$

5. Індекс Се-Бені (XB) - це метод визначення відношення повної варіації в межах кластерів і поділів кластерів [68]. Оптимальний показник визначається мінімальним значенням індексу. Визначається за формулою:

$$XB(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (u_{ij})^m \|x_j - v_i\|^2}{N \min_{i,k} \|v_k - v_i\|^2}, \quad (2.37)$$

6. Індекс Дана (DI) – індекс призначений для ідентифікації кластеризації із чіткою розбивкою. Тому результат кластеризації завжди перераховується. Індекс визначається за формулою:

$$DI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in c_i, y \in c_j} d(x, y)}{\max_{k \in c} \left\{ \max_{x, y \in c} d(x, y) \right\}} \right\} \right\}, \quad (2.38)$$

Основним недоліком індексу Дана є значне збільшення складності обрахунку із збільшенням c і N . Оптимальний показник визначається мінімальним значенням індексу.

7. Альтернативний індекс Дана (ADI) - для спрощення процедур визначення індексу Дана був розроблений механізм розрахунку альтернативного індексу Дана. В цьому випадку відмінність між двома кластерами, що вимірюються при обмірюваними при $\min_{x \in c_i, y \in c_j} d(x, y)$, розраховується відповідно до нерівності:

$$d(x, y) \geq \left| d(y, v_j) - d(x, v_j) \right|, \quad (2.39)$$

де v_j – є центром j -го кластера. Сам альтернативний індекс Дана розраховується за формулою:

$$ADI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in c_i, y \in c_j} \left| d(y, v_j) - d(x, v_j) \right|}{\max_{k \in c} \left\{ \max_{x, y \in c} d(x, y) \right\}} \right\} \right\}, \quad (2.40)$$

Варто звернути увагу на той факт, що коефіцієнт розподілу і класифікаційна ентропія ефективні тільки для кластеризації з нечіткою розбивкою. Для нечітких кластерів індекс Дана та альтернативний індекс Дана є ненадійними. Це пов'язане з необхідністю перетворення результатів, отриманих за допомогою методу чіткої розбивки. Після перевірки правильності розбивки кластерів, необхідно визначитися з методами візуального відображення даних кластеризації, що допомагає краще зрозуміти дані та результати використання методів кластеризації. Але множини даних, які використовуються в дослідженні, є багатомірними, тому відобразити їх безпосередньо у двовимірному просторі неможливо.

Тому, в дисертації для візуалізації результатів кластеризації використовується три методи відображення. Першим методом є метод головних компонент (Principal Component Analysis) [76] - стандартний і найпоширеніший метод відображення багатомірних даних у просторі низької розмірності. Також розглянуті метод відображення Семона та метод нечіткого відображення Семона [68].

Метод головних компонентів (Principal component analysis- PCA).

Даний метод містить у собі математичну процедуру перетворення певного числа корельованих змінних у меншу кількість некорельованих змінних, які називаються головними компонентами. Перша головна компонента максимально описує мінливість даних. Наступні компоненти описують мінливість, що залишається. Основні цілі методу PCA [76]:

- Ідентифікація нових значимих базових змінних;
- Виявлення і/або зменшення розмірності множини даних.

Головні компоненти будуть обчислюватися на основі аналізу власних векторів і власних значень матриці попарних кореляцій. Напрямок першого головного компонента задається власним вектором з

максимальним власним значенням. Друге по величині власне значення відповідає другому головному компоненту і т.д. В нашому дослідженні ми прагнемо використати метод головних компонент з точки зору його застосування для вирішення задачі виявлення і/або зменшення розмірності множини даних. У цьому випадку коваріаційна матриця множини даних може бути описана наступним виразом:

$$F = \frac{1}{N} \sum_{k=1}^N (x_k - \nu)(x_k - \nu)^T, \quad (2.41)$$

де $\nu = \overline{x_k}$. Метод головних компонент заснований на проектуванні корельованих багатомірних даних на гіперплощину [68]. Цей метод використовує тільки q перших ненульових власних значень і відповідних значень власних векторів коваріаційної матриці нормованих даних:

$$F_l = U_l \Lambda_l U_l^T, \quad (2.42)$$

де l – кількість головних компонент, яка визначається через введення граничного значення пояснюваної компонентами дисперсії нормованих даних; Λ_l – матриця, що містить власні значення λ_{ij} із F_l ; в своїй діагоналі в порядку зменшення; U_l – матриця, що містить власні вектори, які відповідають власним значенням. Окрім того, ще існує q -вимірний вектор y_{lk} , що представляє собою вектор x_k в X і визначається таким чином:

$$y_{lk} = W_l^{-1}(x_k) = W_l^T(x_k), \quad (2.43)$$

В стовпці вагової матриці W_l , міститься q головних ортонормованих вісей [75, 76]:

$$W_l = U_{j,q} \Lambda_{l,q}^{\frac{1}{2}}, \quad (2.44)$$

Метод відображення Семона. Перевагою методу відображення Семона є здатність зберігати відстані між образами. Таке відображення

відстаней значно більше відповідає кластеризації, ніж визначення дисперсій (використовується при застосуванні методу PCA).

Як було зазначено раніше, у відображенні Семона використовуються відстані між елементами для виявлення в q -вимірному просторі даних N елементів, які є представницькими для « n -вимірної множини даних більш високої розмірності. Відстані між елементами в n -вимірному просторі, що визначаються як $d_{ij}=d(x_i, x_j)$, відповідають відстаням між елементами в q -вимірному просторі, $d_{ij}^*=d^*(y_i, y_j)$. Це досягається за рахунок «стресу» Семона - критерію мінімізації помилки E :

$$E = \frac{1}{\lambda} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{(d_{ij} - d_{ij}^*)^2}{d_{ij}}, \quad (2.45)$$

де λ - є константою.

$$\lambda = \sum_{i>j} d_{ij} = \sum_{i=1}^N \sum_{j=i+1}^{N-1} d_{ij}, \quad (2.46)$$

Варто звернути увагу на те, що перераховувати λ не потрібно, оскільки ця константа не міняється в результаті процесу оптимізації. Мінімізація помилки E представляє собою задачу оптимізації в N, q змінних y_{il} , при $i \in \{1, 2, \dots, N\}$ та $l \in \{1, 2, \dots, q\}$, припускаючи, що $y_i = \{y_{i1}, \dots, y_{iq}\}^T$. Значення y_{il} в t -ій ітерації визначається в такий спосіб:

$$y_{il}(t+1) = y_{il}(t) - \alpha \begin{bmatrix} \frac{\partial E(t)}{\partial y_{il}(t)} \\ \frac{\partial^2 E(t)}{\partial y_{il}^2(t)} \end{bmatrix}, \quad (2.47)$$

де α - додатна скалярна константа з рекомендованим значенням $\alpha \sim 0,3 - 0,4$. Ця скалярна константа являє собою розмір кроку для градієнтного пошуку в напрямку:

$$\frac{\partial E(t)}{\partial y_{il}(t)} = -\frac{2}{\lambda} y \sum_{k=1, k \neq i}^N \left[\frac{d_{ki} - d_{ki}^*}{d_{ki} d_{ki}^*} \right] (y_{il} - y_{kl}), \quad (2.48)$$

$$\frac{\partial^2 E(t)}{\partial y_{il}^2(t)} = -\frac{2}{\lambda} \sum_{k=1, k \neq i}^N \frac{1}{d_{ki} d_{ki}^*} \left[(d_{ki} - d_{ki}^*) - \left(\frac{y_{il} - y_{kl}}{d_{ki}^*} \right)^2 \left(1 + \frac{d_{ki} - d_{ki}^*}{d_{ki}} \right) \right], \quad (2.49)$$

Використовуючи даний метод градієнтного спуску, неможливо досягти локального мінімуму на поверхні помилок при пошуку мінімуму E . Це є суттєвим недоліком, оскільки для пошуку мінімуму необхідно виконати ряд експериментів з різними випадковими початковими значеннями. Таким чином, підсумуємо недоліки методу відображення Семона:

1. Відображення Семона представляє собою метод проектування, заснований на збереженні евклідової норми відстані між елементами. Це означає, що відображення Семона може застосовуватися тільки для алгоритмів кластеризації, в яких при розрахунках кластерів використовується евклідова норма відстані між елементами даних;
2. Відображення Семона направлено на знаходження в n -вимірному просторі високої розмірності N елементів у q -вимірному підпросторі більш низької розмірності, так, щоб відстані між елементами відповідали відстаням виміряним в n -вимірному просторі. Для вирішення даної задачі необхідний складний обчислювальний алгоритм, оскільки в кожній ітерації потрібне обчислення $N(N-1)/2$ відстаней.

Проте, використовуючи метод відображення Семона та ґрунтуючись на інформації, що отримується з даних, можна визначити правильні початкові значення кластеризації.

Метод нечіткого відображення Семона. Щоб уникнути проблем, пов'язаних з використанням відображення Семона, у дисертаційному дослідженні застосовується третій алгоритм - модифікований алгоритм, який називається нечітким відображенням Семона. Він враховує основні властивості алгоритмів нечіткої кластеризації, в яких важливими вважається тільки відстань між елементами даних і центрами кластерів [77]. У модифікованому алгоритмі, названому нечітким відображенням Семона, використовуються $N \cdot c$ відстаней, зважених по ступеню приналежності аналогічно рівнянню FCM (2.18):

$$E_{fuzz} = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ki})^m (d(x_k, v_i) - d_{ki}^*)^2, \quad (2.50)$$

де $d(x_k, v_i)$ - відстань між елементом даних x_k і центром кластера v_i , у початковому n -вимірному просторі. Евклідова відстань між центром кластера z_i та даними y_k проектованого q -вимірного простору, представлена у вигляді $d^*(y_k, z_i)$. Відповідно до цих умов, у проектованому двовимірному просторі кожний кластер представлено однією точкою, незалежно від форми вихідного кластера.

Алгоритм нечіткого відображення Семона аналогічний алгоритму відображення Семона, але в цьому випадку центр проектованого кластера буде перераховуватися в кожній ітерації після адаптації проектованих елементів даних. Перерахунок буде виконуватися по формулі середньозваженого значення для алгоритмів нечіткої кластеризації, наведеній в рівнянні (2.18). Графік значень приналежності проектованих даних може бути побудований з використанням стандартного рівняння для розрахунків значень приналежності:

$$\mu_{ki}^* = \frac{1}{\sum_{j=1}^c \left(\frac{d^*(x_k, \eta_i)}{d^*(x_k, \nu_j)} \right)^{\frac{2}{m-1}}}, \quad (2.51)$$

де $U^* = [\mu_{ki}^*]$ - матриця розбивки з перераховуваними приналежностями. На графіку буде відобразитися тільки наближення кластеризації для простору великої розмірності до кластеризації для двовимірного простору. Для виміру якості результатів використовується наступна функція, яка визначає середньоквадратичну помилку між початковими і перераховуваними:

$$P = \|U - U^*\|, \quad (2.52)$$

У свою чергу, недоліком нечіткого відображення Семона є зниження точності для відстані, оскільки враховується тільки відстань між елементами даних і центрами кластерів.

2.5. Визначення методу розпізнавання даних кластеризації споживачів електронної торгівлі

Одним із найбільш вдалих методів розпізнавання даних в кластерах є метод опорних векторів (SVM) алгоритм, що представляє собою механізм присвоювання міток об'єктам за допомогою навчання на прикладах [78, 79]. В даному дослідженні метод опорних векторів використовується для розпізнавання кластерів клієнтів на основі вивчення сотень клієнтів кожного кластера (аналізу характеристик клієнтів). Як правило, даний метод використовується в математиці в якості алгоритму максимізації

певної математичної функції відносно заданої множини даних. Основні ідеї механізму опорних векторів можна пояснити без використання рівнянь. Розглянемо чотири основні поняття:

1. Поділяюча гіперплощина;
2. Гіперплощина з максимальним зазором;
3. М'який зазор;
4. Функція ядра («кernфункція»).

Для геометричної інтерпретації даних припустимо, що існують тільки два кластери. Розглянемо випадок, коли дані про клієнтів містять 2 характеристики - вік і дохід. На рис. 2.6, зеленими маркерами показані клієнти, які перебувають у кластері 1, а червоними позначені клієнти, що перебувають у кластері 2.

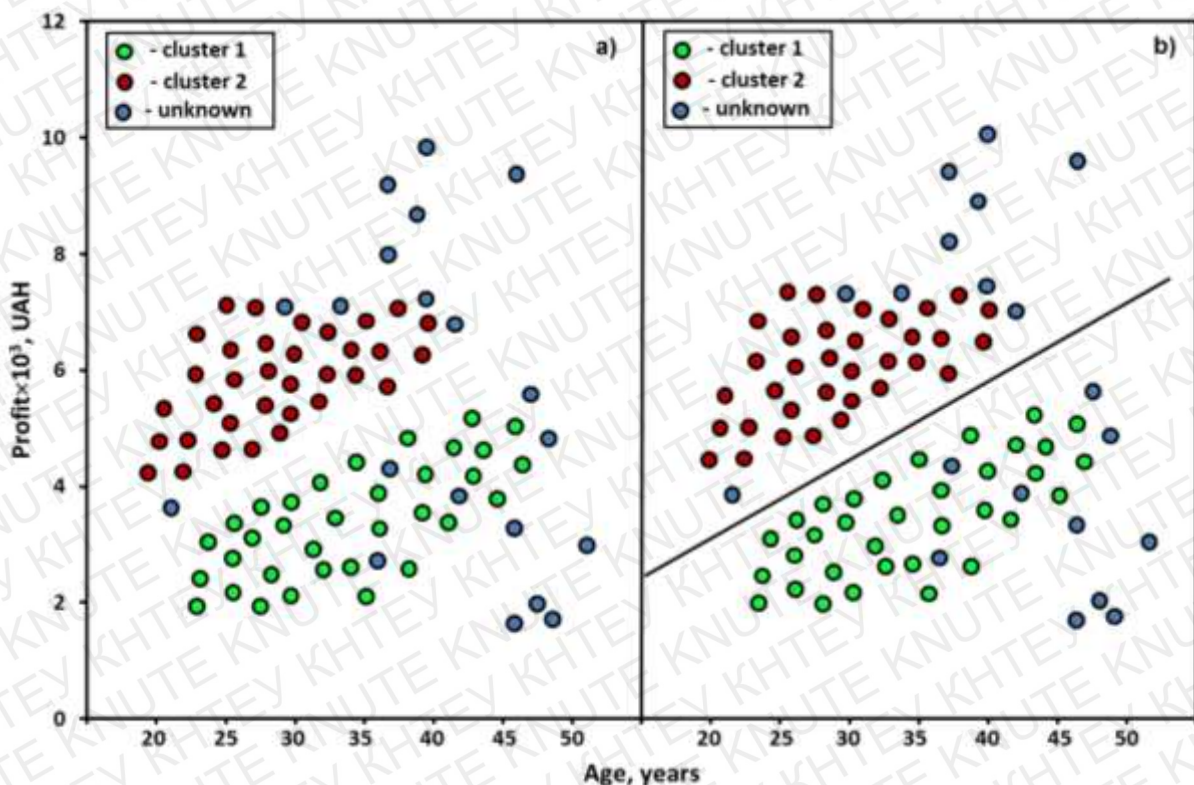


Рис. 2.6. Представлення даних про клієнтів кластера 1 та 2. а) Двовимірне представлення клієнтів. б) Поділяюча гіперплощина.

Ціль методу опорних векторів (SVM) полягає в тому, щоб навчитися визначати відмінність між групами і передбачати для невідмічених клієнтів (сині маркери з позначенням «невідомо» на рис. 2.6) відповідність кластеру 1 або кластеру 2.

Ми можемо легко розпізнавати образи. На рис. 2.6(a) добре видно, що зелені маркери утворюють одну групу, а червоні маркери утворюють іншу групу і ці групи можна легко розділити, провівши між ними лінію (рис. 2.6(б)).

Прогнозування мітки невідомого клієнта виконується просто: потрібно з'ясувати, чи попадає новий клієнт в кластер 1 або кластер 2, які перебувають із різних сторін поділяючої лінії. Для того щоб визначити поняття поділяючої гіперплощини, в даному випадку розглянемо ситуацію, в якій для опису клієнта використовуються не тільки два значення характеристик.

Наприклад, якщо використовується тільки одне значення характеристики для опису клієнта, то простір в якому знаходиться відповідна одномірна характеристика, представляє собою одномірну лінію. Ця лінія може бути розділена навпіл одним маркером. У двох вимірах, на координатній площині, пряма лінія ділить простір на дві половини (рис. 2.6(б)).

У тривимірному просторі для поділу простору необхідна площина. Ця процедура може математично екстраполюватися на більш високі розмірності. У багатовимірному просторі прямій лінії відповідає гіперплощина. Таким чином, термін «поділяюча гіперплощина», по суті є лінією, яка розділяє кластери.

Концепція, пов'язана з представленням об'єктів у вигляді точок в багатовимірному просторі і побудовою поділяючої їх лінії, є розповсюдженим способом класифікації і отже, не унікальним для SVM. В

той час, SVM відрізняється від усіх інших методів класифікації принципом вибору гіперплощини. Ще раз розглянемо проблему класифікації, яка показана на рис. 2.6(а).

Ціль застосування SVM полягає в тому, щоб побудувати лінію, що розділяє кластери клієнтів 1 і 2. Існує нескінченне число можливих ліній, питання полягає в тому, яка лінія повинна бути обрана в якості оптимального класифікатора, і як визначити оптимальну лінію.

Логічним способом визначення оптимальної лінії є вибір лінії, що розташована «у середині». Або іншими словами, це - лінія, яка розділяє два кластери і перебуває на максимальній відстані від маркерів. Не дивно, що теорема теорії статистичного навчання підтримує саме цей вибір [80]. Визначивши відстань від гіперплощини до найближчого клієнта (в загальному випадку вектор) як зазор гіперплощини, SVM вибирає поділяючу гіперплощину з максимальним зазором (рис. 2.7(б)).

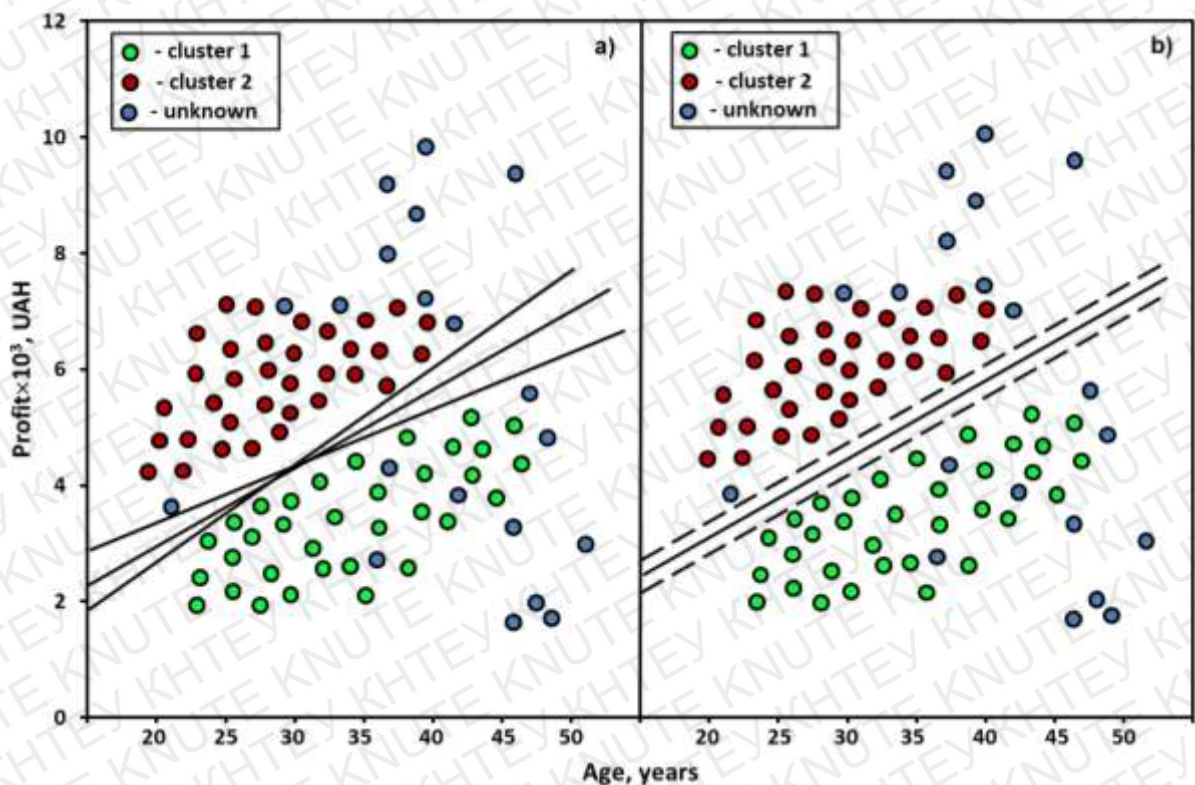


Рис. 2.7. а) Декілька можливих варіантів кластеризації. б) Гіперплощина з максимальним зазором.

При виборі цієї гіперплощини, SVM може робити прогноз щодо невідомих кластерів клієнтів за пріоритетами споживчого профілю, як показано на рис. 2.6(а,б). Вектори (маркери), які обмежують ширину зазору, є опорними векторами.

Ця теорема є основою для успішного використання методу опорних векторів. Проте, є деякі зауваження і застереження, про які варто пам'ятати. Перш за все потрібно зазначити, що теорема заснована на тому, що дані, по яких здійснюється навчання SVM, беруться з того ж самого розподілу, що й дані, які цей метод повинен класифікувати. Це є цілком логічно, оскільки є неправильною класифікація, наприклад, різних типів автомобілів за допомогою SVM, навченого на основі даних про клієнтів. Не варто також очікувати, що SVM забезпечить якісну класифікацію, якщо дані для навчання підготовлені з використанням протоколу, що відрізняється від досліджуваних даних. З іншого боку, згідно з теоремою SVM, два набори даних повинні бути взяті з одного розподілу. Також, застосування SVM не обмежується даними, що беруться тільки з нормального розподілу. Дотепер теорія припускала, що дані можуть розділятися прямою лінією. Проте, більшість реальних наборів даних неможливо чітко розділити прямою лінією, наприклад рис. 2.8(а).

На рис. 2.8 показані дані, що містять помилковий об'єкт. Інтуїтивний спосіб вирішення проблем, пов'язаних з помилками, полягає в розробці алгоритму методу опорних векторів таким чином, щоб він допускав влучення декількох аномальних клієнтів на «інший бік» поділяючої лінії. Це може бути досягнуто шляхом додавання в SVM «м'якого зазору». М'який зазор дозволяє невеликому відсотку точок даних виходити за край поділяючої гіперплощини, не впливаючи на кінцевий результат. При використанні м'якого зазору множина даних, показана на рис. 2.8(а), буде розділена так, як це показано на рис. 2.7(б).

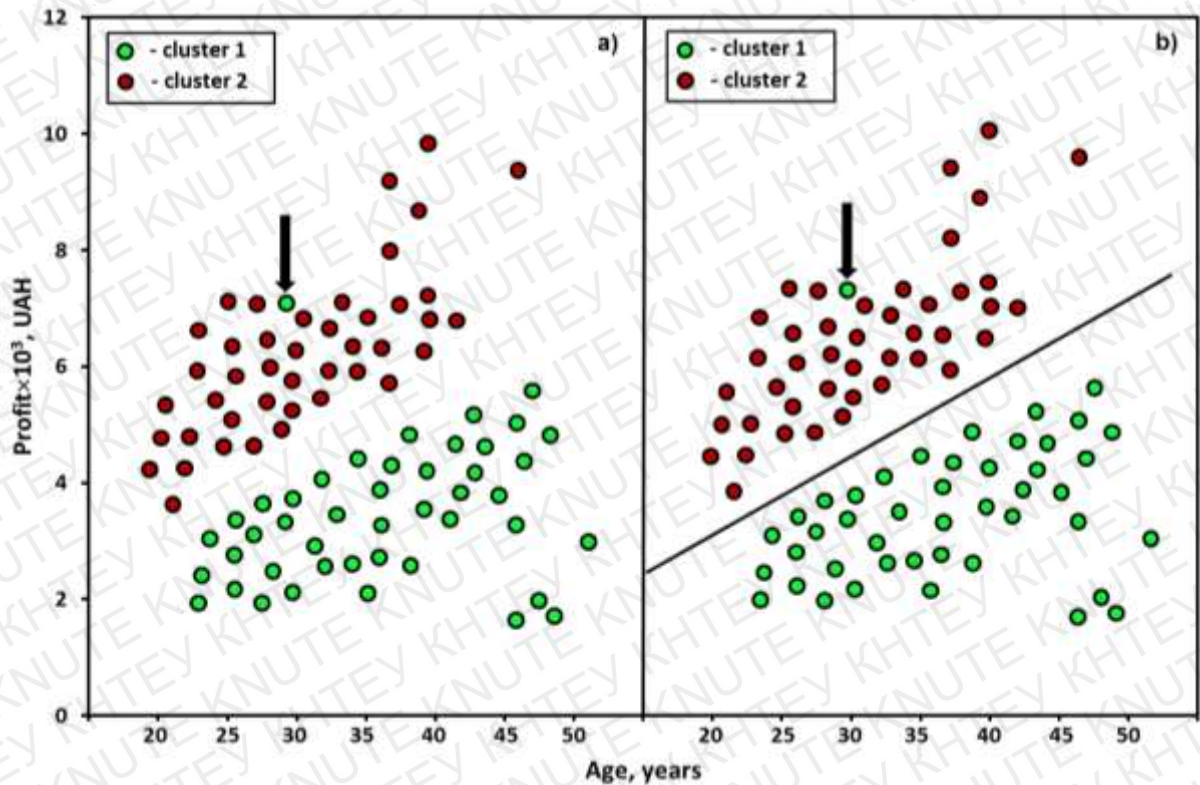


Рис. 2.8. а) Множина даних, що містить одну помилку. б) Приклад м'якого зазору.

Помилковий об'єкт буде розглядатися, як різке відхилення і буде знаходитися на боці кластера 1. Зрозуміло, що SVM не повинен допускати занадто великої кількості помилок класифікації.

Варто звернути увагу на те, що при використанні м'якого зазору вводиться параметр, який контролює м'який зазор та кількість клієнтів, яким дозволено порушувати поділяючу лінію, а також визначає, наскільки далеко вони можуть заходити за лінію. Установка цього параметра представляє собою складний процес, оскільки для певного числа правильних класифікацій зазор буде великим. Іншими словами, м'який зазор визначає компроміс між порушеннями гіперплощини і розміром зазору. Для розуміння функції ядра (кernфункції) спростимо дані, що використовувалися, як приклад ще більше. Припустимо, що замість двовимірної множини даних є одновимірною множиною даних, показана на рис.

2.9(a). У цьому випадку поділяючою гіперплощиною буде єдина точка. Розглянемо тепер ситуацію, показану на рис. 2.9(a), який ілюструє неподільну множину даних. Жодна із точок не може розділити дані на два кластери, а введення м'якого зазору також не допомагає впоратися з даною проблемою. В той же час, функція ядра дозволяє розв'язати цю задачу. Функція ядра додає в дані новий вимір, у нашому випадку піднесенням до другого ступеня одновимірної множини даних. Результат показано на рис 2.9(б). У новому, двовимірному просторі SVM може розділити дані на два кластери однією прямою. Взагалі, функцію ядра можна розглядати для SVM як математичний прийом по проектуванню даних із простору низької розмірності в простір більш високої розмірності (по суті, цей метод обернений до методу головних компонент, де багатовимірний простір даних стискався в простір низької розмірності).

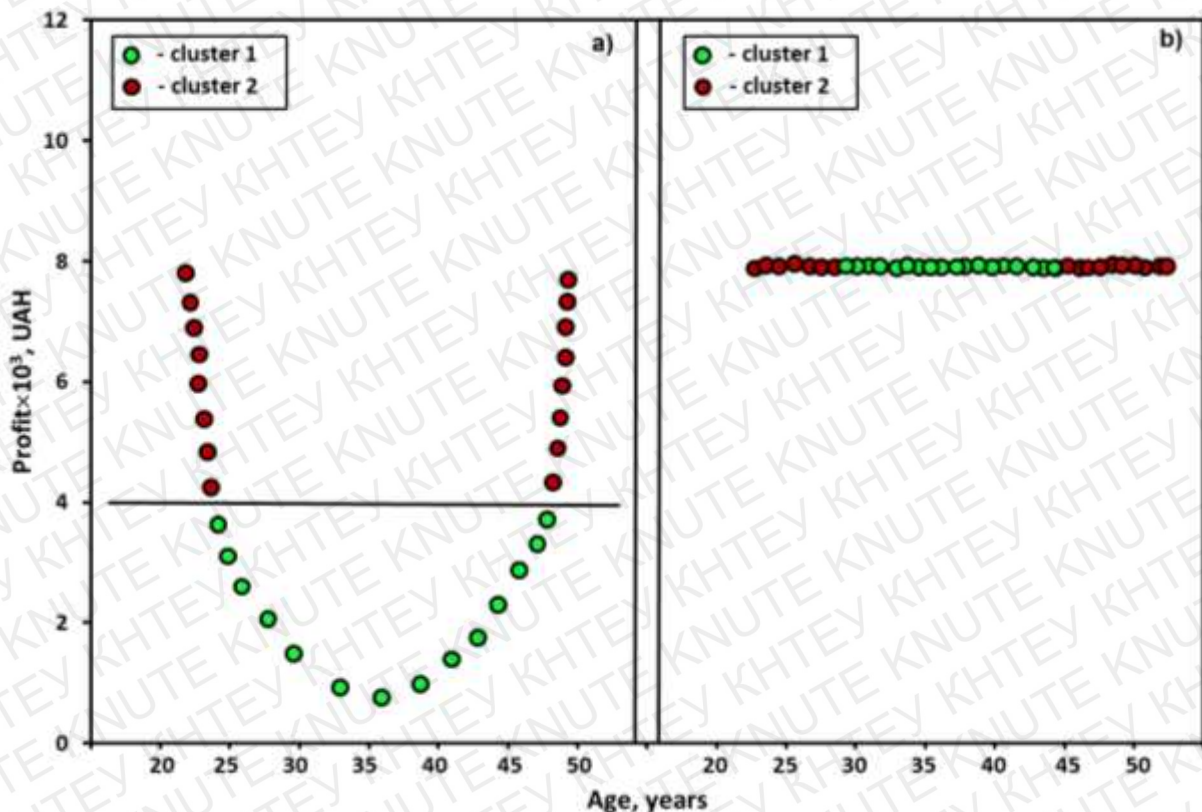


Рис. 2.9. Демонстрація функції ядра. а) Поділ не неподільної множини даних через збільшення кількості вимірів. б) Неподільна множина даних.

Вибір належної функції ядра дозволить розділяти дані у відповідних більш високих просторових вимірах. Для кращого розуміння функції ядра розглянемо деякі приклади. На рис. 2.8, показано множину даних у двовимірному просторі. За допомогою відносно простої функції ядра, ці дані можна спроектувати в чотиривимірний простір.

Відобразити дані в чотиривимірному просторі неможливо, але, використовуючи проекцію гіперплощини SVM у чотиривимірному просторі обернено в початковий двовимірний простір, одержимо криву, показану на рис. 2.10(a). Можна довести, що для будь-якої множини даних існує функція ядра, яка дозволяє SVM розділяти дані лінійно у вимірах більш високого порядку. Звичайно, множина даних повинна містити послідовні мітки, тобто два однакових елементи даних не можуть мати різні мітки. Таким чином, теоретично SVM є ідеальним класифікатором.

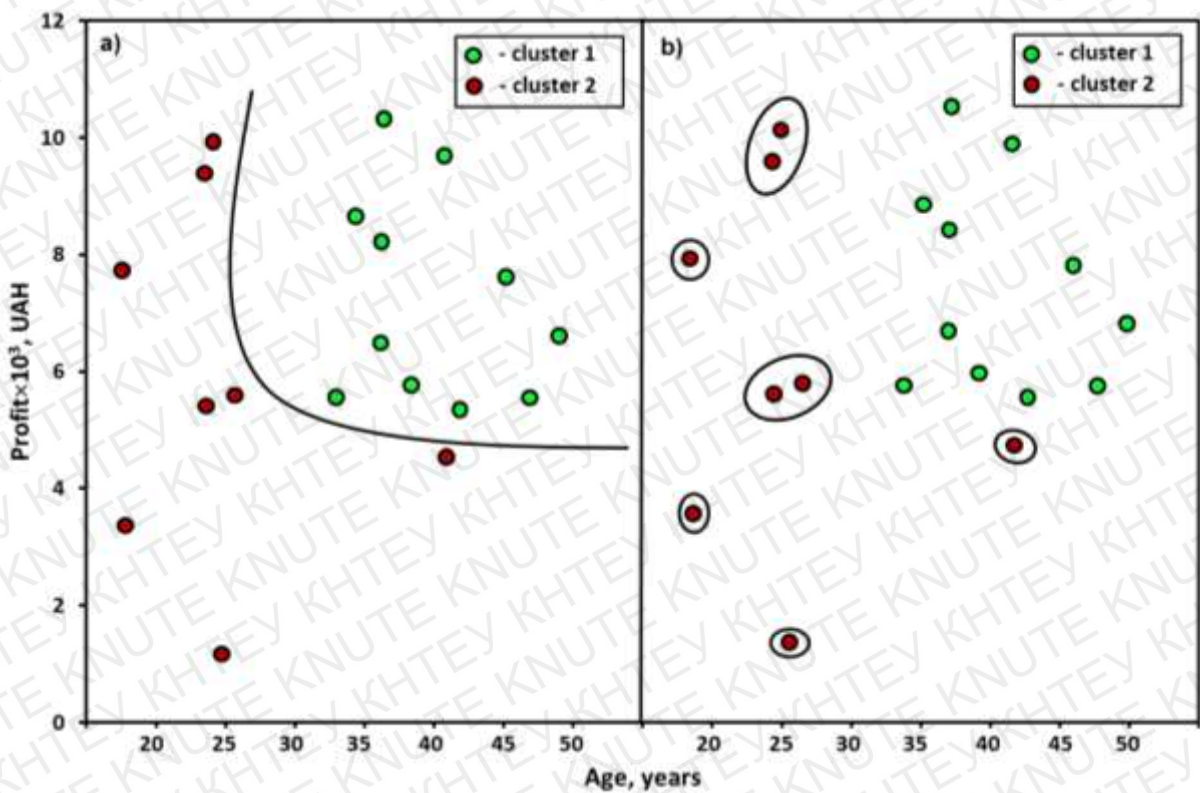


Рис. 2.10. Приклад поділу з використанням функції ядра. а) Лінійний поділ в чотиривимірному просторі. б) SVM з перенавчанням даних.

Але існують певні недоліки проектування даних у простори великої розмірності для визначення поділяючих гіперплощин. Першою проблемою є так звана «проблема розмірності»: коли із збільшенням кількості змінних збільшується кількість можливих розв'язків, при чому в геометричній прогресії.

Тобто, знайти правильний розв'язок стає все складніше при використанні будь-якого алгоритму. На рис. 2.10, показана ситуація, коли дані проектуються в простір великої розмірності. На цьому рисунку представлені ті ж дані, що й на рис. 2.9, але гіперплощина, що проектується знаходиться за допомогою функції ядра дуже великої розмірності. Це приводить до утворення специфічних границь для взятої як приклад множини даних. Такий феномен називається перенавчанням. Метод опорних векторів також неефективний і для нових невидимих невідмічених даних.

Використання нових невидимих даних для SVM пов'язано із суттєвою проблемою. Ця проблема стосується вибору функції ядра, яка б розділяла дані, але не приводила б до занадто значного збільшення розмірності. На жаль, розв'язати дану задачу в більшості випадків можна тільки методом проб і помилок. В дослідженні метод опорних векторів використовується з різними «стандартними» функціями ядра [80]. За допомогою методу перехресної перевірки, статистично буде обрана оптимальна функція ядра [80]. Але цей процес займає багато часу і не гарантує, що виявлена в ході перехресної перевірки оптимальна функція ядра насправді виявиться найкращою із існуючих функцій ядра. Більш імовірно, що існує не перевірена функція ядра, яка є більш ефективною, ніж обрана функція ядра. Описаний вище метод, як правило, дає задовільні результати. У загальному вигляді функція ядра визначається за наступною формулою:

$$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j), \quad (2.53)$$

де x_i і x_j - навчаючі вектори. Вектори переходять у простір більш високої розмірності за допомогою функції Φ . Може використовуватися велика кількість функцій ядра, можливо навіть нескінченна, але раніше вже було виявлено, що деякі функції ядра є ефективними для широкого спектра додатків [80]. В дослідженні використовуються такі рекомендовані функції ядра:

Лінійна, яка визначається за формулою:

$$K(x_i, x_j) = x_i^T x_j, \quad (2.54)$$

Поліноміальна функція ядра ступеня d , що має вигляд:

$$K(x_i, x_j) = (\gamma x_i^E x_j + C_0)^d, \quad (2.55)$$

Радіальна базисна функція, також відома як гаусівська функція ядра і може бути представлена у вигляді:

$$K(x_i, x_j) = \exp(\gamma \|x_i - x_j\|^2), \quad (2.56)$$

Сігмоїдальна функція ядра, яка також використовується в нейронних мережах і має вигляд:

$$K(x_i, x_j) = \tanh(\gamma x_i^E x_j + C_0), \quad (2.57)$$

В нашому дослідженні константа C_0 приймається рівною 1. Концепція застосування функції ядра є надзвичайно перспективною і ефективною [78], вона дозволяє SVM виконувати поділ класів навіть для складних границь [80], як це показано на рис. 2.11. Розглянута нами ідея використання гіперплощини для поділу векторів характеристик на дві групи, може застосовуватися тільки для двох цільових категорій. Як алгоритм SVM застосовувати для більшої різноманітності класів? Для розв'язку цієї задачі було запропоновано кілька методів, два з яких є самими популярними і найбільш часто використовуються [80].

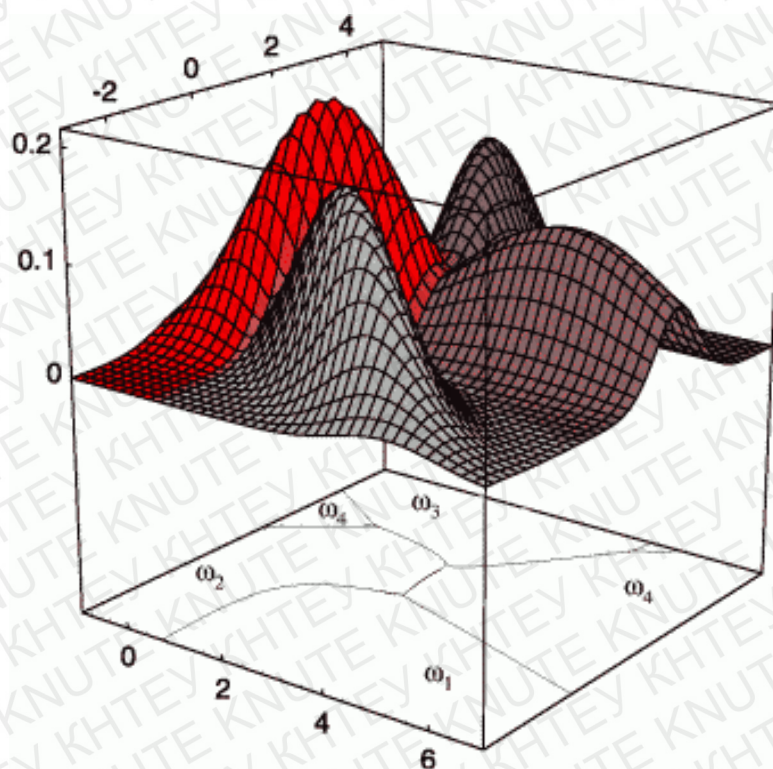


Рис. 2.11. Поділ класів із складними границями [80]

Перший метод заснований на навчанні розпізнаванню класів за допомогою класифікатора «один проти всіх». Наприклад, якщо SVM повинен розпізнати три класи: А, В і С, те можна просто навчити три окремі SVM відповідати на бінарні питання «Це - А?», «Це - В?» і «Це - С?». Другий метод - «один проти одного», в якому створюються моделі $k(k-1)/2$, де k - число класів. При проведенні безпосередніх досліджень використовується метод «один проти одного».

Таким чином, в результаті виконаних досліджень визначено кращі методи кластеризації для досліджуваної сукупності клієнтів, виділені їхні сильні і слабкі сторони. Також розглянуті методи візуалізації даних, при цьому найменш придатним виявився метод відображення Семона, оскільки він може застосовуватися тільки для алгоритмів кластеризації, в яких при розрахунках кластерів використовується евклідова норма відстані між елементами даних. Відповідно, запропоновано застосовувати алгоритм, що

називається k-means, який враховує основні властивості алгоритмів нечіткої кластеризації та відстань між елементами даних і центрами кластерів. Розглянуто метод класифікації з навчанням на прикладах (SVM із учителем). Даний метод доцільно застосовувати для класифікації та присвоєння міток об'єктам за допомогою навчання на прикладах.

ВИСНОВКИ ДО РОЗДІЛУ 2

Споживча поведінка суб'єктів сучасного ринку електронної торгівлі більшою мірою залежить від наявності ефективної стратегії довгострокових довірчих відносин з кожним із існуючих та потенційних клієнтів, забезпечення для них можливості одержання своєчасної інформації про конкурентні переваги пропонованих продуктів і послуг та своєчасного їхнього використання без зайвих витрат і зусиль, як з боку клієнтів, так із боку відділу менеджменту послуг підприємства електронної торгівлі.

Запропоновано раціональний набір параметрів інформації про клієнтів і їх активності на ринку електронної торгівлі. Охарактеризовано процес збору, підготовки та аналізу інформації про клієнтів ринку для наступної її кластеризації і класифікації. Виявлені джерела інформації, проілюстрований зв'язок між обраними характеристиками (описано 22 характеристики споживачів ринку електронної торгівлі). Охарактеризоване поняття споживчого профілю на ринку електронної торгівлі і профілю клієнта, як описової складової елемента в середині клієнтського кластера. В сукупності вони відображають цілісну модель типового клієнта ринку електронної торгівлі і можуть бути використані з метою виявлення та аналізу клієнтських пріоритетів.

Розроблено методику виявлення клієнтських пріоритетів, що включає в себе такі етапи: збору і поділу статистичної клієнтської інформації; вибору і обґрунтування оптимальних методів кластеризації (без учителя і попередньої класифікації); візуалізації результатів; побудови споживчих профілів в електронній торгівлі на основі даних кластеризації і класифікації отриманих кластерів, з урахуванням особистого профілю клієнта. З цією метою проаналізовано і структуровано математичний

апарат методів чіткої і нечіткої кластеризації (без учителя і попередньої класифікації). Визначені недоліки і переваги кожного з методів. Проаналізовані і структуровані методи перевірки, візуалізації та розпізнавання даних кластеризації. Методи перевірки визначаються коефіцієнтами: *PC*, *CE*, *PI*, *SI*, *XB*, *DI* і *ADI*. Для візуалізації результатів кластеризації використовуються три методи відображення: метод головних компонент (PCA), метод Семона (SM) і метод нечіткого відображення Семона (FSM). Для розпізнавання особистих даних клієнтів у кластерах (формування клієнтського профілю) використовується механізм k-means і його основні поняття: поділяюча гіперплощина, гіперплощина з максимальним зазором, м'який зазор і функція ядра як функції навчання (керна функція). Визначені недоліки і переваги кожного із методів з точки зору їхнього застосування в рамках досліджуваної сукупності споживачів і побудови клієнтських профілів.

Застосування розробленої методики виявлення клієнтських пріоритетів показало, що в рамках динамічних умов розвитку і змін ринку електронної торгівлі доцільно для моделювання процесів виявлення клієнтських пріоритетів застосовувати алгоритми нечіткої кластеризації.

РОЗДІЛ 3.

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЗНАЧЕННЯ СПОЖИВАЧИХ ПРІОРИТЕТІВ В ЕЛЕКТРОННІЙ ТОРГІВЛІ.

3.1. Специфіка програмно-апаратних засобів та методів для розробки інформаційної технології

Найпопулярнішим прикладом реалізації інформаційної системи визначення споживчих пріоритетів в електронній торгівлі є впровадження її на веб-сайт за допомогою методів машинного навчання. З таким способом реалізації ми стикаємось досить часто в мережі Інтернет, коли хочемо здійснити будь-яку онлайн-покупку. Проте останнім часом традиційні інформаційні системи, які ми звикли спостерігати на веб-сайтах онлайн-магазинів, втрачають свою актуальність: компанії активно використовують персональних онлайн-консультантів, які також накопичують дані про покупців. Їхня розробка на порядок важча від звичайного впровадження рекомендацій на веб-сайт: система потребує великої кількості інформації у вигляді повноцінної бази знань як про предмети, так і про користувачів, часу для розробки та впровадження, а також постійної подальшої підтримки. Не зважаючи на це, он-лайн-консультанти стають все більш популярним способом реалізації класифікаційних інформаційних системи, оскільки, зважаючи на постійне збільшення кількості інформації, що пропонується користувачам, персональний онлайн-помічник – незамінний інструмент для швидкого пошуку потрібного товару. Типом класифікаційної інформаційної системи є система на основі знань, що використовує дані про предмети (споживчий профіль) та користувачів (особистий профіль) для здійснення рекомендацій. Розробка класифікаційної інформаційної технології для

визначення споживчих пріоритетів здійснена за допомогою одного із методів машинного навчання – неконтрольованого методу. Алгоритмом машинного навчання без вчителя, що використовувався під час розробки є алгоритм *k-means*. Побудова та тестування алгоритму *k-means* здійснювались з використанням технології Microsoft Azure, зокрема інструменту, який розроблений з метою полегшення розробки систем на основі машинного навчання, – Microsoft Azure Machine Learning Studio.

Алгоритм *k-means*, використовуючи базу знань про товари, допомагає класифікувати інформацію без необхідності маркування вхідних даних. Такий алгоритм часто використовують у рекомендаційних системах на основі знань, оскільки при постійному зростанні інформації про користувача чи предмети алгоритм сам визначає залежності між ними. Наприклад, у розробленій інформаційній технології алгоритм *k-means* використано для розподілу на кластери за ціною предметів. Використання алгоритмів класифікації вимагало жорсткого маркування і потребувало тривалого тестування. Одним із недоліків алгоритму *k-means* є проблема вибору значення *k*. Завдяки Azure Machine Learning Studio тестування цього алгоритму перетворюється у швидку та достатньо просту задачу – одночасно можна виконувати *n* кількість тестувань із різними значеннями *k*. Остаточний результат реалізації алгоритму у вигляді тієї ж вхідної вибірки значень, але які вже поділені на кластери, можна завантажити та імпортувати у базу даних, яка використовується для збереження інформації про предмети та користувачів, а також доступу до цієї інформації для надання рекомендацій.

Компанія Microsoft надає доступ до технологій машинного навчання, які значно спрощують розробку рекомендаційних систем. Для реалізації класифікаційного механізму, окрім Azure Machine Learning Studio, було використано ще один з інструментів, що доступні у цій сфері від Microsoft,

– Microsoft Framework. Microsoft Framework надає можливість побудови інформаційних систем як “з нуля” за допомогою інструменту Builder, що включає в себе набори із засобів розробки, утиліт та документації (Software development kit – SDK), так і використовуючи Azure Service, що надає п’ять готових шаблонів для створення інформаційної технології з класифікації пріоритетів надає безліч функцій, за допомогою яких взаємодія між інформаційною системою і користувачем реалізується простіше та швидше.

Framework Emulator – інструмент, що дозволяє розробнику тестувати та налагоджувати роботу інформаційної системи. Емулятор відображає повідомлення так само як вони б виглядали в інтерфейсі веб-сайту. Запити користувача та відповіді реєструються у форматі JSON та відображаються у вікні емулятора. JSON (JavaScript Object Notation) – синтаксис для зберігання та обміну даними з позначенням об’єкту JavaScript.

Node.js – це програмна платформа, що перетворює мову програмування JavaScript у мову загального призначення: додає можливість взаємодіяти з пристроями вводу-виводу через API (програмний інтерфейс - набір готових класів, процедур, функцій і т.д.), підключати зовнішні бібліотеки, написані на різних мовах програмування та викликати їх із JavaScript-коду. Node.js надає можливість підключення бібліотек Builder та Restify (для створення веб-сервісу), тому платформа дуже зручна у використанні для побудови класифікаційних систем.

Розробка будь-якої інформаційної класифікаційної системи ґрунтується на інформації – вхідних даних. Для зберігання та швидкого доступу до даних використано документоорієнтовану систему управління базами даних – MongoDB. Ця система для збереження інформації використовує JSON-подібні документи. MongoDB – найпопулярніша система управління базами даних для платформи Node.js. Для того, щоб

була змога відповідно до запитів відповідати на отримані від користувача повідомлення у MongoDB зберігається інформація про предмети, які рекомендуються, та про користувача – його профіль та попередні замовлення. За допомогою цієї інформації формуються класи та надсилаються цільовому користувачеві системи (менеджера). Доступ до бази знань відбувається за допомогою побудови запитів, які ґрунтуються на вимогах користувача. Запит формується із даних, що надійшли від цільового користувача в повідомленні.

3.3. Інформаційно-логічна модель процесу визначення споживчих пріоритетів

Для знаходження балансу між постачальником і споживачем товарів і послуг, задоволенням потреб клієнтів, автоматизацією добору клієнтові набору товарів і послуг в залежності від динаміки споживання і соціально-демографічних характеристик, розроблено наступну інформаційно-логічну модель реалізації процесу формування клієнтських вподобань у рамках реалізації бізнес-процесу підбору раціонального набору товарів і послуг в контексті клієнтських характеристик (рис. 3.1).

Постачальник товарів і послуг в електронній торгівлі розробляє спеціальні набори послуг, засновані на профілі споживання елементів клієнтських кластерів і їх центрів. Кожному елементу кластера споживачів ринку електронної торгівлі визначається один зі спеціально створених наборів послуг n (сформований із шаблонів).

Якщо змінюється профіль споживання елемента кластера - змінюється набір послуг, що визначає двосторонній зв'язок інформаційних потоків у схемі, реалізованій за допомогою впровадження інформаційної системи підтримки прийняття рішень. В роботі розглядається питання

кластеризації клієнтів без використання знань експертів і без попередньої класифікації. Кластеризація здійснюється на основі даних про споживання товарів і послуг за досліджуваний період часу (місяць) і додаткової інформації про споживачів на основі випадкової вибірки для 1000 клієнтів з 5000 клієнтів досліджуваної загальної сукупності споживачів.

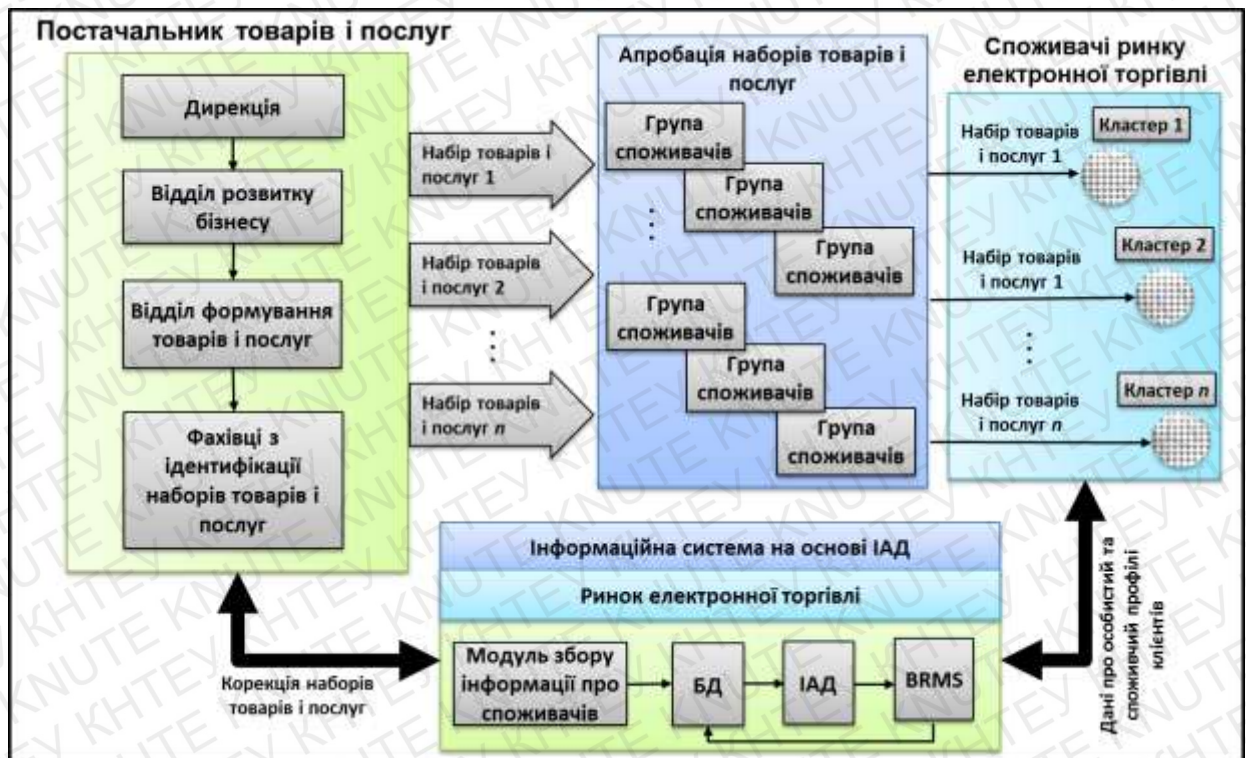


Рис.3.1. Інформаційно-логічна модель реалізації процесу визначення клієнтських вподобань в рамках взаємодії постачальника і покупця з використанням кластерного розподілу і класифікації на основі особистого та споживчого профілю клієнта

Для досягнення даної мети розроблена методика виявлення клієнтських вподобань в основі якої лежить комплексне використання ІАД. Випробувані, перевірені і співставленні один з одним різні методи аналізу даних, названі методами кластеризації, застосовані методи перевірки, розпізнавання і візуалізації даних кластеризації. В результаті

кластеризації, на основі отриманих особистих даних клієнтів, для кожного кластера створений особистий профіль клієнта. На основі особистої інформації і профілю споживання клієнта виконана його оцінка, визначено його поведінку і здійснено підбір товарів і послуг за допомогою встановлених правил.

3.2. Побудова інформаційної бази системи і програмна реалізація

Розроблена інформаційна технологія використовує базу знань, на основі якої здійснює класифікацію споживачів. Ця база спроектована за допомогою системи управління базами даних MongoDB. Побудована база знань складається із двох баз даних: про користувачів та про товари. Створена класифікаційна система є допоміжним інструментом для будь-якого онлайн-магазину. Вона побудована на основі використання інформації про товари які закупають споживачі та взята із мережі Інтернет. Вибір такої інформаційної бази обґрунтований великим вибором товарів у сфері електронної торгівлі, а різноманіття брендів, моделей та інших характеристик передбачає пошук того, що користувачеві потрібно.

Програмна реалізація інформаційної технології здійснювалась на платформі Node.js. Перш за все потрібно побудувати підключення до локального сервера за допомогою фреймворку Restify.

На рисунку 3.2 показано локальне підключення до бази даних за допомогою драйвера MongoDB та класу MongoClient. Таким чином створюється база даних під назвою “items” та всередині неї колекції з товарами. Проектування бази даних розпочинається із внесення інформації та визначення кількості необхідних таблиць. База даних товарів містить дві колекції: сукупність усіх доступних товарів та сукупність товарів, кластеризованих за ціною.


```
db_products.js
1  var MongoClient = require('mongodb').MongoClient;
2
3  var url = "mongodb://localhost:27017/items";
4
5  MongoClient.connect(url, function(err, db) {
6    if (err) throw err;
7    console.log("Database created!");
8    db.close();
9  });
10
11 MongoClient.connect(url, function(err, db) {
12   if (err) throw err;
13   var dbase = db.db("items");
14   dbase.createCollection("phones", function(err, res) {
15     if (err) throw err;
16     console.log("Collection created!");
17     db.close();
18   });
19 });
```

Рис. 3.2. Підключення до MongoDB та створення колекції товарів.

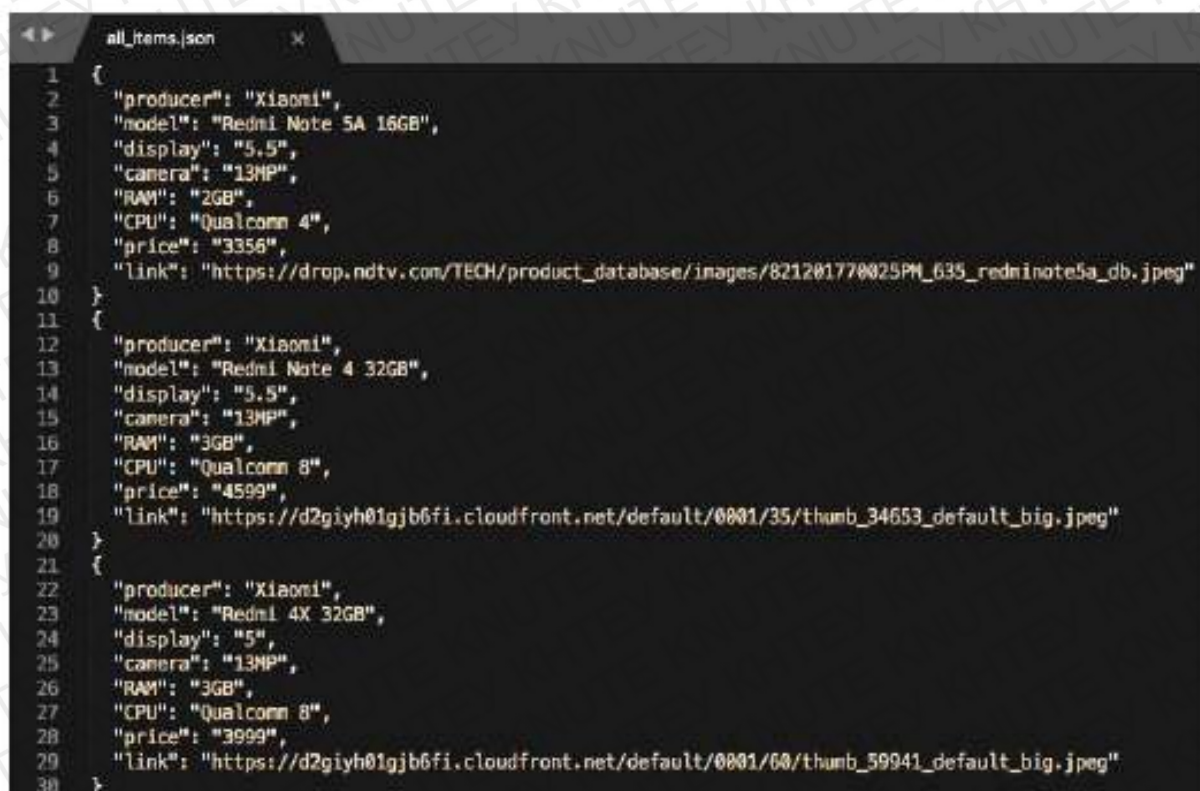
Для локального доступу до бази даних необхідно встановити допоміжний драйвер MongoDB для Node.js. Після цього створюється необхідна база даних та пусті колекції товарів за допомогою Node.js. Створення колекції користувачів здійснюється аналогічно на основі даних, що згруповані у відповідності до табл. 2.1.

За допомогою вікна терміналу – інтерфейсу командного рядка можна перевірити чи створились необхідні бази даних та колекції (рис 3.3).

```
MongoDB Enterprise e-trade > show dbs
admin 0.000GB
config 0.000GB
items 0.000GB
local 0.000GB
MongoDB Enterprise > use items
switched to db items
MongoDB Enterprise > show collections
customers
MongoDB Enterprise > db.products.find()
MongoDB Enterprise e-trade > █
```

Рис. 3.3. Перевірка статусу баз даних у вікні терміналу

Таким чином можна переглянути список баз даних, колекцій, що вони містять, а також наповнення цих колекцій. На рисунку 3.4 показано, що створення бази даних товарів і користувачів пройшло успішно, проте колекції поки що пусті та не містять ніяких записів. Підготовка даних до запису в базу є важливим етапом розробки інформаційної технології. Товари які закупають споживачі можуть мати безліч різноманітних характеристик, за якими користувачі здійснюють запити. Розробка та впровадження повноцінної бази знань є тривалим процесом, навіть після впровадження бази знань її потрібно постійно доопрацьовувати та переглядати, оскільки неможливо відразу врахувати усі можливі запити користувачів. Система на основі знань передбачає розширення. Інформаційна технологія передбачає запам'ятовування запитів користувачів та пошук за встановленим пріоритетом по класу кластера.



```
all_items.json
1 {
2   "producer": "Xiaomi",
3   "model": "Redmi Note 5A 16GB",
4   "display": "5.5",
5   "camera": "13MP",
6   "RAM": "2GB",
7   "CPU": "Qualcomm 4",
8   "price": "3356",
9   "link": "https://drop.ndtv.com/TECH/product_database/images/821201770025PM_635_redminote5a_db.jpeg"
10 }
11 {
12   "producer": "Xiaomi",
13   "model": "Redmi Note 4 32GB",
14   "display": "5.5",
15   "camera": "13MP",
16   "RAM": "3GB",
17   "CPU": "Qualcomm 8",
18   "price": "4599",
19   "link": "https://d2giyh01gjb6fi.cloudfront.net/default/0001/35/thumb_34653_default_big.jpeg"
20 }
21 {
22   "producer": "Xiaomi",
23   "model": "Redmi 4X 32GB",
24   "display": "5",
25   "camera": "13MP",
26   "RAM": "3GB",
27   "CPU": "Qualcomm 8",
28   "price": "3999",
29   "link": "https://d2giyh01gjb6fi.cloudfront.net/default/0001/60/thumb_59941_default_big.jpeg"
30 }
```

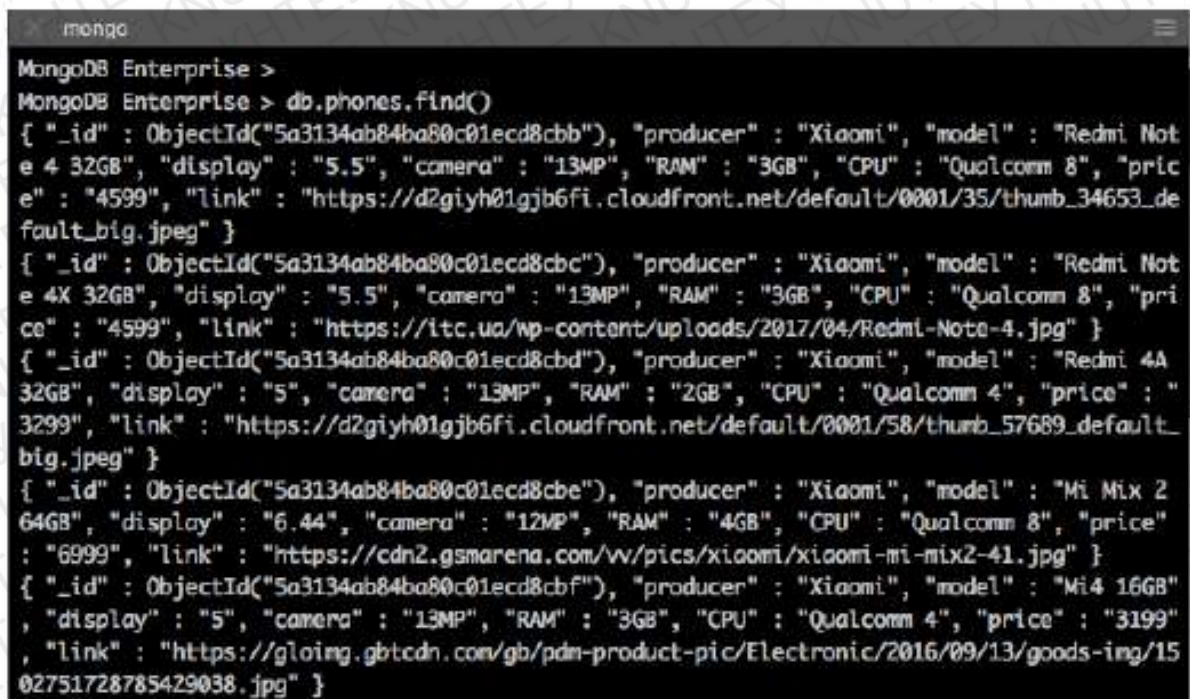
Рис. 3.4. Структура колекції JSON-документів про різні товари бази даних

MongoDB

Тому підготовка даних для розробки системи є дуже непротим та тривалим етапом, але й не менш важливим, оскільки саме на основі використання бази знань ґрунтується вся інформаційна технологія.

Кожна колекція бази даних MongoDB складається із записів, які називаються документами. Створена колекція товарів складається із наступних документів: виробник, тип товару, ціна, кількість, тип доставки, термін гарантії. Створена колекція покупців складається з таких документів: стать тип телефону, тип приватного клієнта, розмір корпоративного клієнта, місце проживання, вікова група. Приклад структурованого JSON-документу, де зберігається інформація, яка буде завантажена в колекцію бази даних, зображений на рис. 3.4.

Створений JSON-документ готовий до завантаження у базу даних товарів. Переглянути результат завантаження можна за допомогою вікна терміналу, що показано на рисунку 3.5.



```
mongo
MongoDB Enterprise >
MongoDB Enterprise > db.phones.find()
{ "_id" : ObjectId("5a3134ab84ba80c01ecd8cbb"), "producer" : "Xiaomi", "model" : "Redmi Note 4 32GB", "display" : "5.5", "camera" : "13MP", "RAM" : "3GB", "CPU" : "Qualcomm 8", "price" : "4599", "link" : "https://d2giyh01gjb6fi.cloudfront.net/default/0001/35/thumb_34653_default_big.jpeg" }
{ "_id" : ObjectId("5a3134ab84ba80c01ecd8cbc"), "producer" : "Xiaomi", "model" : "Redmi Note 4X 32GB", "display" : "5.5", "camera" : "13MP", "RAM" : "3GB", "CPU" : "Qualcomm 8", "price" : "4599", "link" : "https://itc.ua/wp-content/uploads/2017/04/Redmi-Note-4.jpg" }
{ "_id" : ObjectId("5a3134ab84ba80c01ecd8cbd"), "producer" : "Xiaomi", "model" : "Redmi 4A 32GB", "display" : "5", "camera" : "13MP", "RAM" : "2GB", "CPU" : "Qualcomm 4", "price" : "3299", "link" : "https://d2giyh01gjb6fi.cloudfront.net/default/0001/58/thumb_57689_default_big.jpeg" }
{ "_id" : ObjectId("5a3134ab84ba80c01ecd8cbe"), "producer" : "Xiaomi", "model" : "Mi Mix 2 64GB", "display" : "6.44", "camera" : "12MP", "RAM" : "4GB", "CPU" : "Qualcomm 8", "price" : "6999", "link" : "https://cdn2.gsmarena.com/vv/pics/xiaomi/xiaomi-mi-mix2-41.jpg" }
{ "_id" : ObjectId("5a3134ab84ba80c01ecd8cbf"), "producer" : "Xiaomi", "model" : "Mi4 16GB", "display" : "5", "camera" : "13MP", "RAM" : "3GB", "CPU" : "Qualcomm 4", "price" : "3199", "link" : "https://gloing.gbtcn.com/gb/pdn-product-pic/Electronic/2016/09/13/goods-ing/1502751728785429038.jpg" }
```

Рис. 3.5. Відображення у вікні терміналу колекції “products” бази даних “items”

Як видно з рисунку 3.5 зручнішим для відображення колекції є JSON-документ, ніж проглядати її за допомогою вікна терміналу, але потрібно мати змогу перевірити результат імпорту в базу даних. База даних товарів – основа створеної інформаційної технології. Саме завдяки інформації, що в ній зберігається, інформаційна технологія здійснюватиме класифікацію клієнтів, особливо, якщо даних про користувача ще немає у базі знань. Якщо база даних товарів проектується безпосередньо перед початком роботи, то база даних споживачів створюється безпосередньо під час роботи інформаційної системи. На основі колекції товарів було створено ще одну колекцію, яка наповнена цією ж інформацією. Відмінність полягає у тому, що за допомогою алгоритму кластеризації k-means неконтрольованого машинного навчання усі товари у новій колекції мають ще одну характеристику: кластер, до якого вони належать. Кластеризація здійснювалась за допомогою хмарної платформи Microsoft Azure Machine Learning Studio. Для здійснення кластеризації потрібно підготувати вхідні дані у форматі CSV, що, як і JSON, призначений для збереження табличних даних. Після цього потрібно завантажити цей файл у Machine Learning Studio. Платформа дозволяє обрати необхідний алгоритм машинного навчання, протестувати його, та завантажити результат у форматі CSV. Для поставленої задачі – групування товарів за ціною – було обрано алгоритм k-means. Побудова алгоритму за допомогою платформи Microsoft здійснюється за допомогою створення своєрідної схеми із блоків, кожен з яких здійснює певну задачу. Для того, щоб вказати з якими вхідними даними алгоритм повинен працювати, потрібно перемістити із бокового меню блок із завантаженими вхідними даними (рис. 3.6). Кожного разу, коли до моделі додається новий блок, потрібно її запускати для перескладання, щоб впевнитись, що всі блоки побудовані правильно та виконують поставлені задачі.



Рис. 3.6. Завантаження вхідних даних, необхідних для реалізації алгоритму машинного навчання на Microsoft Azure Machine Learning Studio

Для того, щоб зрозуміти та переглянути результат роботи того чи іншого блока, існує опція візуалізації. На рисунку 3.7 показано візуалізацію вхідних даних, і, що головне, статистичні дані, які необхідні для побудови алгоритму кластеризації.

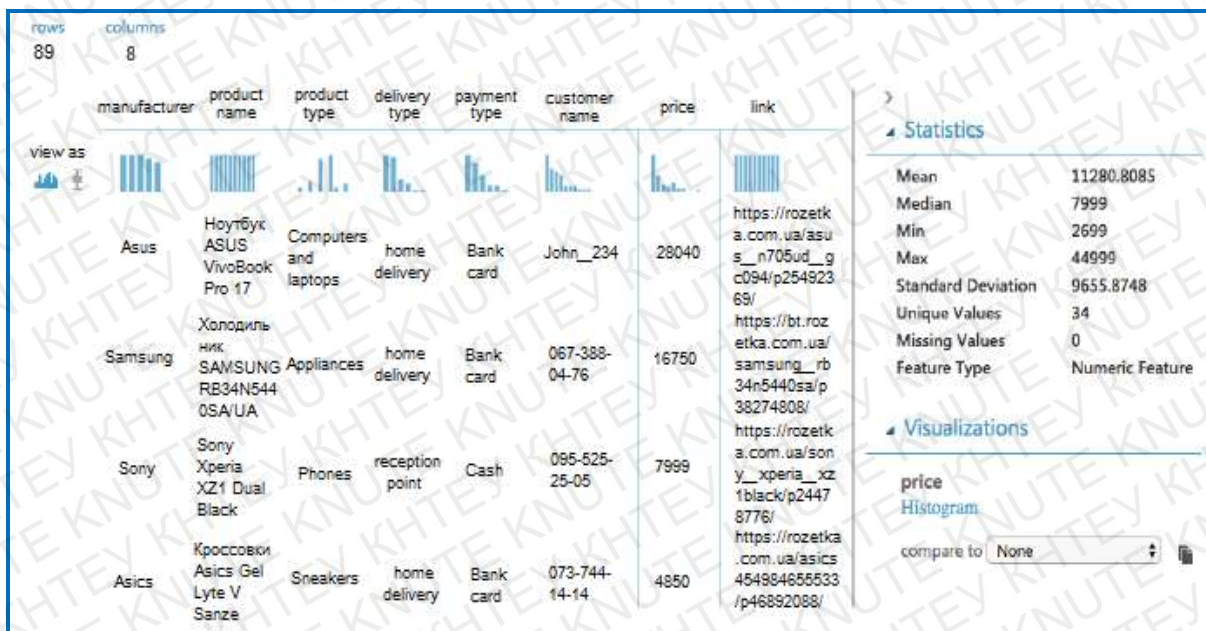


Рис. 3.7. Візуалізація колекції вхідних даних

Статистичні дані за ціною товарів містять в собі інформацію про найвище та найнижче значення у колонці, середнє значення, відхилення тощо. Усі ці значення необхідні для подальшого коригування вхідних даних, наприклад, якби у даних, які ми надаємо були пропущені значення – їх потрібно було б відкинути. Також за допомогою статистичної інформації можна підібрати приблизне початкове значення k для реалізації алгоритму k-means.

Оскільки, завантажені вхідні дані не містять пропущених значень, а колонка ціни є числовим значенням, не потрібно здійснювати додаткову обробку даних. Блок-схема реалізованого алгоритму кластеризації зображена на рисунку 3.8. Machine Learning Studio дозволяє тестувати не один алгоритм, тому до вхідних даних було застосовано алгоритм k-means із двома різними значеннями k . Алгоритм k-means при $k = 4$ поділив вхідні дані за ціною на 4 нерівні кластери, один з яких містив занадто багато значень, порівняно з іншими.

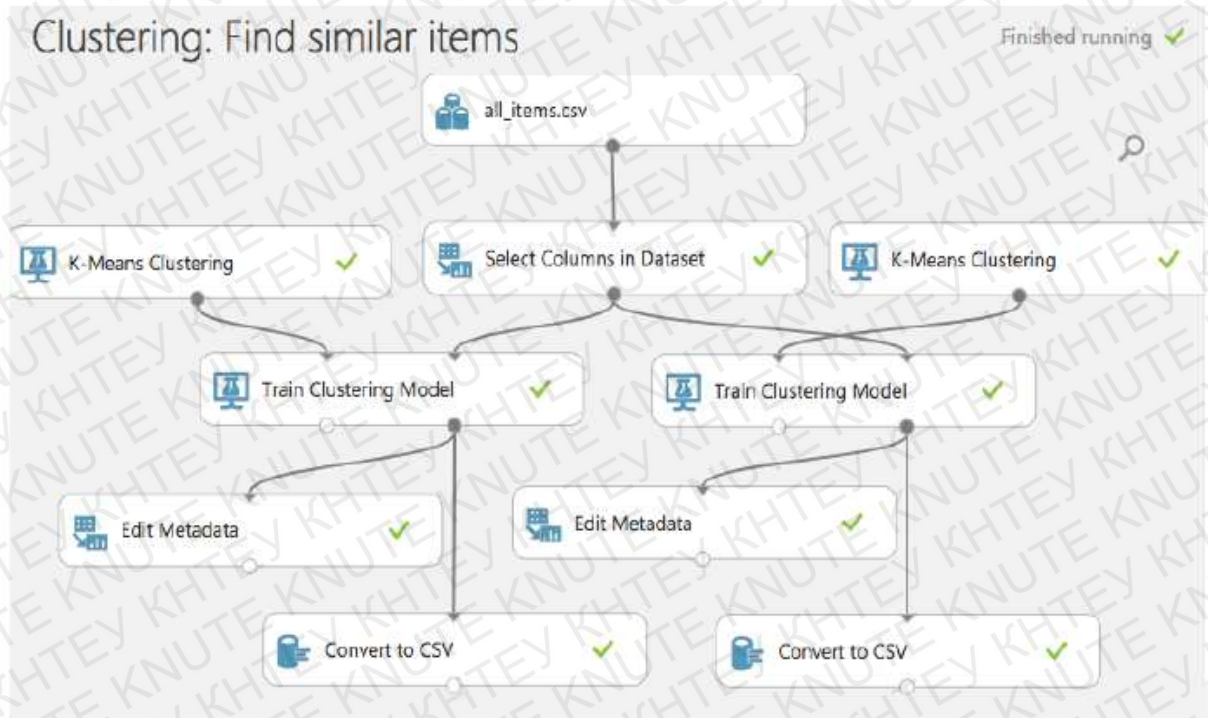


Рис. 3.8. Схема алгоритму кластеризації k-means

При $k = 5$ розмір найбільшого кластеру зменшується, проте все одно у порівнянні з іншими містить велику кількість товарів. Це зумовлено тим, що ціна товарів у вхідній сукупності даних, яка зустрічається найчастіше, знаходиться у межах від 7000 грн до 11000 грн. При $k = 6$ розмір найбільшого кластеру не зменшується, проте здійснюється розбиття інших кластерів, що зумовлює створення груп товарів, у яких міститься лише декілька товарів. Тестування алгоритму з різними значеннями k дозволило обрати найкращий результат – при $k = 5$. Графічне зображення результату дії алгоритму k-means представлено на рисунку 3.9. Вісь X – показує значення ціни, за якою здійснювалась кластеризація, а вісь Y – показує обсяги кластеру розподіл клієнтів за профілями.

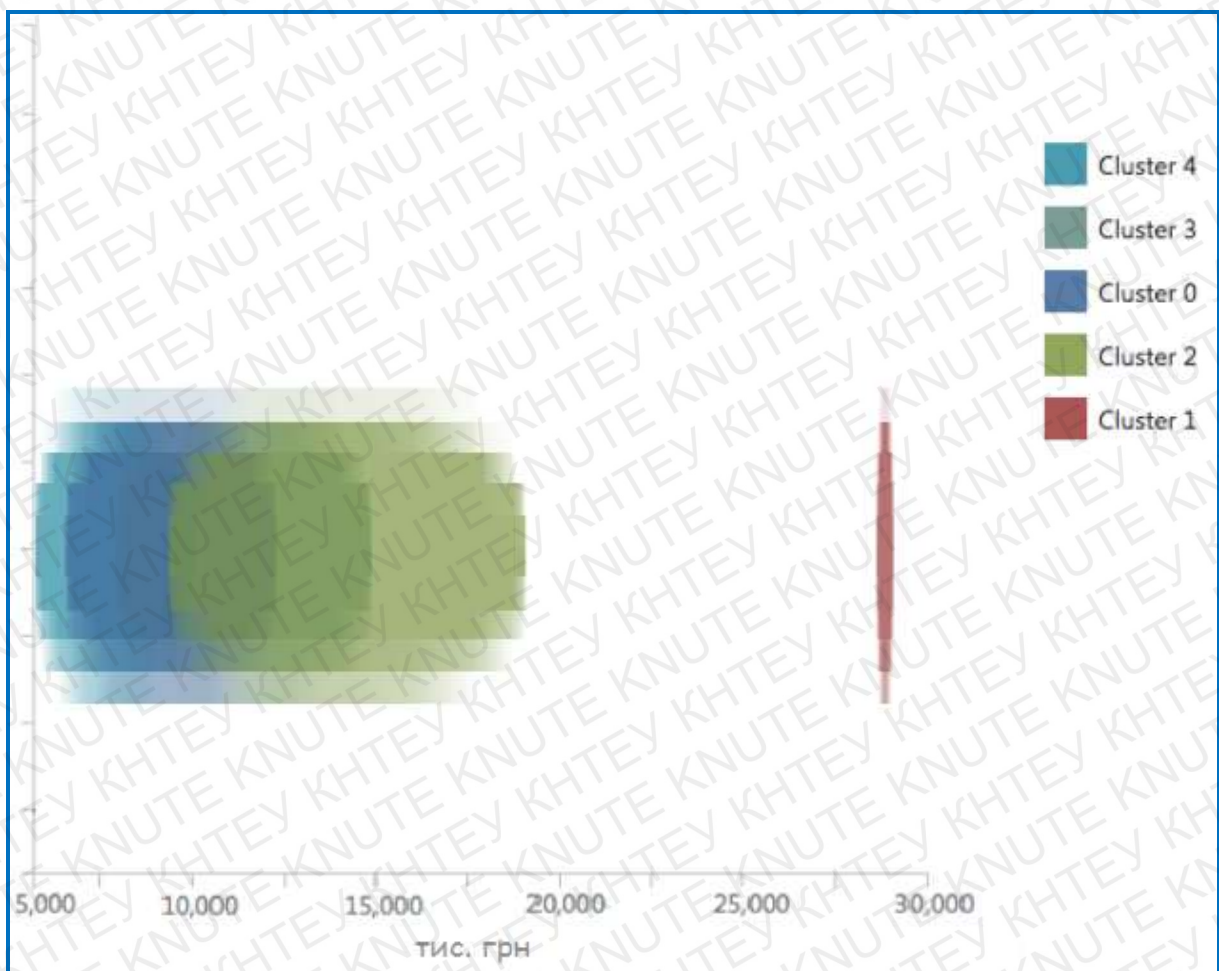


Рис. 3.9. Графічне зображення результату дії алгоритму кластеризації k-means за ціною товарів

На рис. 3.9 кластер №1 містить лише одне значення – це приклад точки-викиду вхідних даних. Ціна одного із ноутбуків знаходиться у межах 28000 грн, тому порахована евклідова відстань до центру найближчого кластеру занадто велика і алгоритм не зміг віднести це значення ні до одного з існуючих кластерів.

Під час розробки інформаційної технології база даних товарів дуже велика, тому ймовірність знайти точки-викиди значно зменшується. Відповідно, залежність між варіативністю компоненти колекції бази даних, за якою буде здійснюватися кластеризація, та кількістю кластерів прямо пропорційна. Іншою базою даних, що необхідна для роботи розробленої інформаційної технології, є база даних користувачів, яка складається з наступних колекцій: профілі, замовлення та нерозпізнані запити. Схема бази даних користувачів зображена на рисунку 3.9.

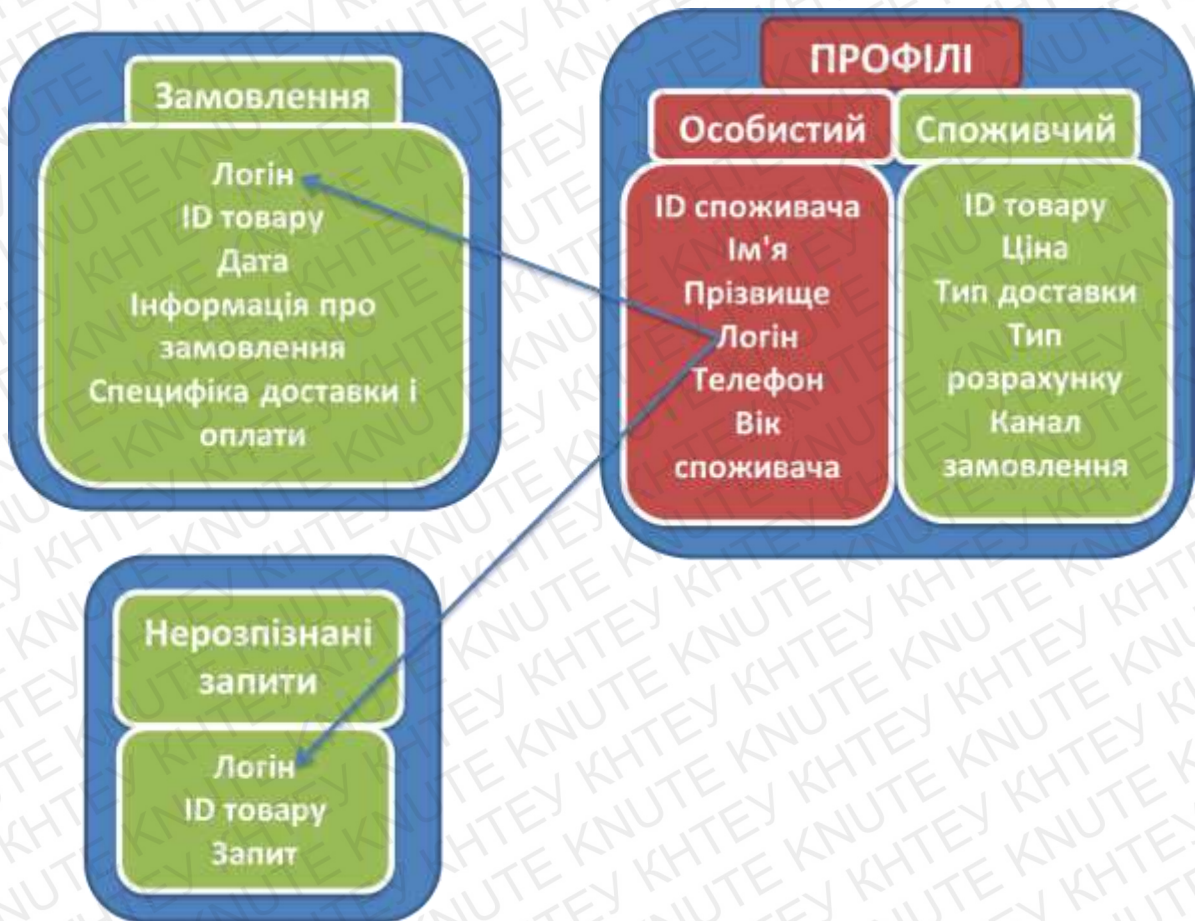


Рис. 3.9. Схема бази даних споживачів

Усі колекції бази даних користувачів заповнюються інформацією у процесі роботи системи, тому, зазвичай, колекція даних про користувачів пуста до моменту впровадження системи. Колекція “профілі” складається із особистого профіля клієнта та його споживчого профіля і містить наступні документи: ID споживача, ім'я, прізвище, вік споживача, ID товару, ціна, тип доставки, тип розрахунку, канал замовлення та логін, за допомогою якого він зможе здійснювати покупки та переглядати інформацію про попередні замовлення. Колекція замовлень містить такі документи: логін, ID товару, дату, інформацію про замовлення, специфіку доставки і оплати. Колекція нерозпізнаних запитів містить документ, в якому зберігається нерозпізнане повідомлення від користувача та документ з логіном користувача, від якого надійшло повідомлення. Завдяки збереженню замовлень легко знайти інформацію про споживача на основі його покупки, будуючи запити за датою або за даними про товар.

Після запуску сервера, відкривши емулятор, потрібно з'єднатися з цим сервером, вказавши URL. Якщо з'єднання успішне, в журналі подій (Log) зареєструється оновлення поточної розмови та відкриється доступ до введення повідомлень користувачем. Покупець повинен зареєструватися або увійти. При реєстрації споживачем вводиться інформація яка є основою особистого профілю клієнта (додаток А). А в процесі купівлі товарів формується споживчий профіль клієнта на основі товарів які він купив і процедур які він вибрав в процесі купівлі товару. У межах кожного кластеру існує мінімальна та максимальна ціна (рис. 2.1), вікова група, тип телефону, тип приватного клієнта (табл. 2.1), розмір корпоративного клієнта, місце проживання, тип оплати, тип доставки. Програмний код виділення кластерів представлено у додатку Б.

Для визначити в якому кластері здійснювати пошук, потрібно знайти ці значення та записати у поточну сесію. Після цього необхідно здійснити

сортування кластерів – від кластеру із найменшими цінами до того, що із найбільшими. Програмний код виділення товарів за пріоритетами в кластери наведено у додатку В. Після того, як користувач вибере товар, ціну та виробника товару, система формує групи товарів по клієнтам, які відповідають обмеженнями та вимогами користувача. Надіславши повідомлення “замовити”, користувач отримує відповідь у вигляді посилання для здійснення оплати. Після цього користувач може вибрати один із варіантів оплати: картка, переказ, готівка.

Колекція кластеризованої інформації про товари та споживачів формується за допомогою алгоритму кластеризації. Алгоритм k-means використано для пошуку кластерів, товари в яких близькі за ціною та з близькими показниками особистого профілю клієнта. На основі цієї інформації формуються характеристики пріоритетів кластерів споживачів за цінами і групами товарів, формою замовлення, типами доставки та розрахунків (табл. 3.1).

Табл. 3.1. Відсоткові співвідношення різних класифікаційних груп [15 -17]

Вікова група	<25 26,22%	25-40 34,5%	40-55 27,9%	>55 11,4%
Стать	Чоловіки 60,2%	Жінки 39,8%		
Тип телефону	Мобільний 73,4%	Стаціонарний 26,6		
Тип приватного клієнта	Консервативний 34,9%	Помірний 36,0%	Прогресивний 29,1%	
Розмір корпоративного клієнта	Малий 31,5%	Середній 34,4%	Великий 17,2%	Корпорація 16,9%
Місце проживання	Велике місто 42,0%	Невелике місто 38,0%		

У табл. 3.1. наводяться результати аналізу досліджуваних клієнтів (вся сукупність, у відсотках) в обраній групі відповідно до обраних показників. З табл. 3.1, добре видно, що розміри груп обґрунтовано підібрані, а отримані значення можуть використовуватися для створення профілів клієнтів.

ВИСНОВКИ ДО РОЗДІЛУ 3

Для знаходження балансу між постачальником і споживачем товарів і послуг, задоволенням потреб клієнтів, автоматизацією добору клієнтового набору товарів і послуг в залежності від динаміки споживання і соціально-демографічних характеристик, розроблено інформаційно-логічну модель реалізації процесу формування клієнтських вподобань у рамках реалізації бізнес-процесу підбору раціонального набору товарів і послуг в контексті клієнтських характеристик.

Побудована інформаційна технологія реалізована на основі інструменту Microsoft Net Framework. Розробка системи здійснювалася за допомогою методу неконтрольованого машинного навчання. У процесі розробки інформаційної технології були використані наступні інструментальні засоби та технології: Microsoft Azure Machine Learning Studio; Microsoft Net Framework; Builder SDK на платформі Node.js; Net Framework Emulator; MongoDB.

Інформаційна технологія є системою на основі знань, оскільки використовує різноманітні бази даних для надання рекомендацій. Управління базами даних здійснюється за допомогою системи MongoDB. Інформаційна база складається із двох компонент: база даних товарів та база даних користувачів. База даних товарів побудована на основі інформації із доступних баз даних, що розміщені в мережі Інтернет.

База знань системи вміщує колекції, що дозволяють будувати конкретні запити, не здійснюючи пошук по всьому масиву даних. База даних складається з таких колекцій: товари, особистий і споживчий профіль клієнта. Інформація, що зберігається колекціях створюється у форматі JSON, завдяки чому її можна переглянути у структурованому та зручному для аналізу вигляді.

ВИСНОВКИ

У випускній кваліфікаційній роботі представлено результати теоретичних і прикладних досліджень, що полягають у розробці інформаційної технології визначення споживчих пріоритетів.. Результати прикладних досліджень стали основою для створення автоматизованої системи визначення споживчих пріоритетців в електронній торгівлі. В результаті проведених досліджень були отримані такі висновки:

1. Ринок електронної торгівлі є динамічним по своїй суті і регулярно поповнюється новими видами взаємодії та розширеним спектром для кожної категорії учасників ринку. Виникає закономірна проблема вибору конкретної послуги для конкретного клієнта, або клієнтського кластера серед ринкового різноманіття.

2. Представлена в роботі інформаційно-логічна модель визначення клієнтських вподобань дозволяє удосконалити бізнес-процес вибору раціонального набору товарів і послуг залежно від клієнтських характеристик, сформувані актуальну систему ринкових відносин та забезпечити раціональність і своєчасність надходження товарів і послуг на ринок електронної торгівлі та забезпечить їхню диверсифікованість і диференціацію.

3. Продавцю в електронній торгівлі необхідно розробляти спеціальні набори послуг, що ґрунтуються на характеристиках споживчого профілю всередині кластера і наборі соціально-демографічних характеристик клієнта. Кожному елементу кластера споживачів ринку електронної торгівлі ставиться у відповідність один з спеціально створених наборів товарів і послуг із загального числа наборів n . Якщо змінюється споживчий профіль і набір особистісних характеристик центру кластера, або його елемента - відповідно змінюється і набір товарів та послуг.

4. Для реалізації процесу визначення споживчих пріоритетів було використано алгоритм кластеризації k-means неконтрольованого машинного навчання. Побудова та тестування алгоритму здійснювались за допомогою інструменту – Microsoft Azure Machine Learning Studio. Експерименти з кластеризації клієнтської інформації виявило цілий ряд залежностей пов'язаних з отриманими клієнтськими профілями і використаними в процесі кластеризації кількістю кластерів. Для проведення експериментів, взято базу з 5000 клієнтів (інтернет-магазин Розетка) і випадковим чином обрано 1000 клієнтів. Виявлено, що для застосованого алгоритму k-means оптимальне число кластерів рівняється п'яти ($k = 5$).

5. Проведені дослідження показали, що для підвищення конкурентоспроможності компаній у сфері електронної торгівлі потрібно впроваджувати сучасні інформаційні технології визначення споживчих пріоритетів на основі бази знань, яка забезпечує найбільш персоналізований підхід до взаємодії з клієнтами та можуть використовувати будь-яку базу даних для своєї роботи, завдяки чому є універсальним і підходить для впровадження в діяльність підприємств електронної торгівлі..

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Економіка України: стратегія і політика довгострокового розвитку / За редакцією акад.. НАН України В.М.Гейця. - К.: Інститут економічного прогнозування; 2003. – 1008 с.
2. Ramazanov S. Dynamic model of the competition in regional market of e-commerce enterprises / S. Ramazanov, L. Istomin, A. Dyubanov // Polish Academy of sciences branch in Lublin. ТЕКА: Comission of motorization and power industry in agriculture, Volume 12 No 4, - Lublin - Lugansk. – 2012. – р. 248-252.
3. Вітлінський В.В. Моделювання збутової електронної логістичної підсистеми з використанням нечіткої мережі Петрі / В. В. Вітлінський, Г. В. Мельник, В. І. Скіцько // Бізнес Інформ. - 2014. - № 8. - С. 82-87.
4. Лысенко Ю. Г. Электронная коммерция / Ю.Г. Лысенко, В.Н. Андриенко, И.И. Иванов. - Донецк: ООО "Юго-Восток, Лтд", 2004. - 187 с.
5. Базилевич В. Д. Формування ринку електронної комерції в Україні / В. Д. Базилевич // Вісн. Ін-ту екон. прогнозування. - 2002. - № 2. - С. 32-36.
6. Внутрішня торгівля в Україні: економічні умови ефективного розвитку: монографія / [А. А. Мазаракі, І. О. Бланк, Л. О. Лігоненко, Н. М. Гуляєва та ін.]; за ред. Мазаракі А. А. — К. : Київ. нац. торг.-екон. ун-т, 2006. — 195 с.
7. Порохня В. М. Моделювання економіки / В.М. Порохня. – Запоріжжя: Запоріз. держ. інж. акад. 2001. - 382 с.
8. Ситник В.Ф. Телекомунікації в бізнесі: Навч.-метод. посіб. для самот. вивч. дисц. / В.Ф. Ситник, І.А. Козак. – К.: КНЕУ, 1999. – 204 с.
9. Хейг М. Основы электронного бизнеса / М. Хейг; пер. с англ. С. Косихина. – М.: ФАИР-ПРЕСС, 2002. – 208 с.
10. Копитко Б.І., Юрчук Г.В. Поняття електронного бізнесу та класифікація його основних бізнес-моделей // Соціально-економічні

дослідження в перехідний період. Механізми регулювання регіонального ринку праці. Зб. наук. пр. Випуск 3 (XXXIV) / НАН України. Інститут регіональних досліджень. – Львів, 2002. – С. 349-364.

11. Плєскач В.Л. Електронна комерція / В.Л. Плєскач, Т.Г. Затонацька. – К.: Знання, 2007. – 535с.

12. Дюбанов О.С., Рамазанов С.К. Моделі поведінки суб'єктів ринку електронної торгівлі в сучасних умовах // Журнал «Бізнес Інформ» (Харківський національний економічний університет МОН України). – Харків, 2011 - №6 – с. 104-105.

13. Laudon К.С. E-commerce: business, technology, society / К.С. Laudon, С.Г. Traver. — USA : Addison Wesley, 2002. — 762 p.

14. Козье Д. Электронная коммерция / Д. Козье. - М.: ИТД "Русская редакция", 1999. - 228 с.

15. Хартман А. Стратегии успеха в Интернет-экономике / А. Хартман. - М.: ЛОРИ, 2001. - 272 с.

16. Global Perspective On Retail: Online Retailing [Електронний ресурс]: Cushman&Wakefield. – Режим доступу: <http://www.cushmanwakefield.us/en/research-and-insight/2013/global-perspectives-on-retail>.

17. Global Retail E-Commerce Index [Електронний ресурс]: А.Т. Kearney. – Режим доступу: <http://www.atkearney.com.au/consumer-products-retail/e-commerce-index>.

18. E-commerce worldwide - important statistics [Електронний ресурс]: Statista – Режим доступу: <https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide>.

19. Ілляшенко С.М. Перспективи і основні проблеми розвитку Інтернет-торгівлі в Україні / С. М. Ілляшенко, Т.Є. Іванова // Механізм регулювання економіки. – 2014. – № 3. – С. 72-81.

20. Шалева О.І. Електронна комерція: Навч. посіб. – К.: Центр учбової літератури, 2011. – 216 с.
21. European Ecommerce Report 2017 [Электронный ресурс]: Ecommerce Foundation. - Режим доступа: <http://www.ecommercefoundation.org/download-free-reports>.
22. Kim K. Electronic and Algorithmic Trading Technology: The Complete Guide (Complete Technology Guides for Financial Services) / К. Kim. – Academic Press, 2007. – 224 p.
23. Davis G. E-Commerce Handbook: Boosting Your Profits with eCommerce/ G. Davis. – CreateSpace Independent Publishing Platform, 2015 – 40 p.
24. Крастинь О.П. Разработка и интерпретация моделей корреляционных связей в экономике / О.П. Крастинь. - Рига: Зинатне, 1983. – 302 с.
25. Крылов Э. И., Журавкова И.В. Анализ эффективности инвестиционной и инновационной деятельности предприятия: учебное пособие / Э.И. Крылов, И.В. Журавкова. - М. : Финансы и статистика, 2001. - 384 с.
26. Курс экономической теории: Учебное пособие / Под общ. ред. М. Н. Чепурина, Е. А. Киселёвой. — Киров : АСА, 2000. – 752 с.
27. Лукинов И.И. Вопросы совершенствования методов экономического управления / И.И. Лукинов // Экономика и математические методы – 1985 - т. XXI. - вып. 2. – с.211-223.
28. Малыхин, В.И. Математика в экономике: учеб. Пособие / В.И. Малыхин. — М.: ИНФРА-М, 2002. — 352 с.
29. 43. Цзян Ю., Шан Ж. і Лю Ю. Оптимізація графіків платних доставок, щоб максимізувати прибутки від електронних платежів / Ю. Цзян, Ж. Шан, Ю. Лю // Міжнародний журнал економіки виробництва. 2003. - №146. - с. 634-645.
30. Мескон М.Х., Альберт М., Хедоури Ф. Основы менеджмента / М.Х. Мескон, М. Альберт, Ф Хедоури. - М.: Вильямс, 2006. – 672 с.

31. Многомерный статистический анализ / Под ред. С.А. Айвазяна и С.Е. Кузнецова. - М.: ЦЭМИ, 1974. – 416 с.
32. Українські онлайн-покупці стали активніше купувати через соціальні мережі [Електронний ресурс]: GfK Ukraine. – Режим доступу: <http://www.gfk.com>.
33. Сравнение рынков электронной коммерции в Украине, России, США и Европе [Електронний ресурс]: e-Commerce. – Режим доступу: <http://e-commerce.com.ua>.
34. E-commerce в Україні: розвиток попри кризу [Електронний ресурс]: Економічна правда – Режим доступу: <https://www.epravda.com.ua/publications/2015/10/27/564857/http://www.prostobiz.ua>.
35. Електронний торговий майданчик як ефективний метод оптимізації закупівельного процесу [Електронний ресурс]: Українська електронна торгова площадка – Режим доступу: <http://www.uetp.com.ua>.
36. Исследование рынка электронной комерции Украины [Електронний ресурс]: Ukrainian direct marketing association – Режим доступу: <http://www.uadm.com.ua>.
37. About Us [Електронний ресурс]: Horizon Capital. – Режим доступу: <http://horizoncapital.com.ua/uk/company-overview>.
38. Все в Мережу. Український ринок електронної комерції стоїть на порозі масштабних змін [Електронний ресурс]: Новое время. – Режим доступу: <http://nv.ua/ukr/publications/vse-v-merezhu-ukrajinskij-rinok-elektronnoji-komertsiji-stojit-na-porozi-masshtabnih-zmin-78270.html>.
39. Пурський О.І. Аналіз стану і перспективи розвитку електронної торгівлі в Україні / О.І. Пурський, Б.В. Гринюк, І.О. Жарій // Економіка Фінанси Право. – 2015. – № 6. – С. 14–17.

40. Пурський О.І. Моделювання трансакційних витрат споживачів в електронній торгівлі / О.І. Пурський, Б.В. Гринюк, Д.П. Мазоха // Проблеми Економіки – 2014. – №4. – с. 466-473.
41. Україна оказалась в середине списка стран по развитию e-commerce [Електронний ресурс]: e-Commerce – Режим доступу: <http://e-commerce.com.ua>.
42. Пурський О.І. Моделювання рекламних витрат на електронному торговельному ринку / О.І. Пурський, Б.В. Гринюк, Д.А. Шестопап // Бізнес Інформ. – 2015. – №2. – С. 113-118.
43. Електронна комерція в Україні 2016: цифри і факти. [Електронний ресурс]: Webexpert – Режим доступу: <https://webexpert.com.ua/ua/elektronna-komertsiya-v-ukrayini-2016>.
44. IT industry needs less rules, more red tape cut [Електронний ресурс]: Ukraine digital news – Режим доступу: <http://uadn.net>.
45. Три фактора, которые помогут e-commerce справиться с кризисом [Електронний ресурс]: e-Commerce – Режим доступу: <http://e-commerce.com.ua>.
46. Макдональд М. Планы маркетинга: Как их составлять и использовать / М. Макдональд. - М.: Технология, 2004. – 651 с.
47. Ансофф И. Стратегическое управление. Классическое издание / И. Ансофф. - М.: Экономика, 2009. – 344 с.
48. Андрианов Д.Л. Имитационное моделирование и сценарный подход в системах принятия решений / Д.Л. Андрианов, Г.К. Полушкина, М.Н. Балаш // Проблемы теории и практики управления. - 2002. - №5. – С. 74-75.
49. Atkinson A.B. Lectures on Public Economics / A.B. Atkinson, J.E. Stiglitz, New York, McGraw-Hill, 1980. – 568 p.

50. Канеман Д. Карты ограниченной рациональности: психология для поведенческой экономики / Д. Канеман // Психологический журнал. 2006. - №4. – С. 5-28.
51. Данько Т.П. Управление маркетингом: Учебник. 2-е изд. / Т.П. Данько. - М.: Инфра-М, 2001. – 334 с.
52. Канеман Д. Рациональный выбор, ценности, фреймы / Д. Канеман, А. Тверски // Психологический журнал. 2003. - № 4. – С. 31-42.
53. Пурський О.І. Інтелектуальний аналіз даних як частина концепції управління взаємодією з клієнтами в електронній торгівлі // О.І. Пурський, І.О. Мороз, Д.А. Шестопап // Математичні методи, моделі та інформаційні технології в економіці: Матеріали V Міжнародної науково-методичної конференції. – Чернівці: Друк Арт, 2017. - С. 126-127.
54. Sankar R. Customer data clustering using data mining technique / R. Sankar //International Journal of Database Management Systems (IJDMS) – 2011. – Vol.3, No.4. – P. 1-11.
55. Woo J.Y. Visualization method for customer targeting using customer map / J.Y. Woo, S.M. Bae, S.C. Park // Expert Systems with Applications – 2005. – Vol 28. – P. 763–772.
56. McDonald M. Market Segmentation: How to do it, how to Profit from it / M. McDonald, J. Dunbar. – 3 ed. – Butterworth-Heinemann, 2004. – 492 p.
57. Jadczková V. Review of Segmentation Process in Consumer Markets / V. Jadczková // Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensi – 2013. – №4. – P. 1215-1224.
58. Ahola J. Data mining case studies in customer profiling. Research report TTE1-2001-29 / J. Ahola, E. Rinta-Runsala. – VTT Information Technology, 2001. – 22 p.
59. Pursky O.I. Analysis of consumer’s activity at the electric trade market / O.I. Pursky, D.A. Shestopal // Public and private partnership in condition of

innovative development of the economy: Collective monograph. – Edizioni Magi: Rome, Italy, 2017. – P. 62-73.

60. Electronic trade company "Rosetka" [Electronic resource]. - Access mode: <http://www.rosetka.ua>.

61. Official site of the e-trade company "Mobilluck" [Electronic resource]. - Access mode: http://www.mobilluck.com.ua/mob_cont.php.

62. Online store "27.ua" [Electronic resource]. - Access mode: <https://27.ua>.

63. Amat J.L. Using reporting and data mining techniques to improve knowledge of subscribers; applications to customer profiling and fraud management / J.L. Amat // J. Telecom. Inform. Technol. – 2002. – №3. – P. 1-5.

64. Thierauf R.J. Knowledge Management Systems for Business / R.J. Thierauf. – Praeger, 1999. – 376 p.

65. Cox D.R. Principles of Applied Statistics/ D.R. Cox, C.A. Donnelly. – 1 ed. – Cambridge University Press, 2011. – 212 p.

66. Mali K. Clustering and its validation in a symbolic framework / K. Mali, S. Mitra // Pattern Recognition Lett. – 2003. – Vol. 24. – P. 2367-2376.

67. Amat J.L. Using reporting and data mining techniques to improve knowledge of subscribers; applications to customer profiling and fraud management / J.L. Amat // J. Telecomm. Inf. Technol. – 2002. – №3 – P. 11-16.

68. Sato-Ilic M. Innovations in Fuzzy Clustering: Theory and Applications (Studies in Fuzziness and Soft Computing) / M. Sato-Ilic. – Springer; Softcover reprint of hardcover 1st ed., 2006. – 151 p.

69. Feldman R. Knowledge discovery in textual databases (KDT) / R. Feldman, I. Dagan / In Proc. 1st Int. Conf. Knowledge Discovery and Data Mining (KDD), 1995. – 112-117 p.

70. Mansuo Zhao M. Image Thresholding Techniques: Based On Fuzzy Partition And Entropy Maximization / M. Zhao. – VDM Verlag, 2009. – 156 p.

71. Заде Л. Размытые множества и их применение в распознавании образов и кластер-анализе. В сб.: Классификация и кластер /Л. Заде. – М: Мир, 1980. – С. 208–247.
72. Kassambara A. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning: Volume 1 (Multivariate Analysis) / A. Kassambara. – CreateSpace Independent Publishing Platform; 1st ed., 2017. – 188 p.
73. Xu Z. Intuitionistic Fuzzy Aggregation and Clustering (Studies in Fuzziness and Soft Computing) / Z. Xu. – Springer; 2012. – 278 p.
74. Bezdek J.C. Optimal fuzzy partition: A heuristic for estimating the parameters in a mixture of normal distributions / J.C. Bezdek, J.C. Dunn // IEEE Trans. Comput. – 1975. – C-24. – P. 835-838.
75. Gath I. Fuzzy clustering for the estimation of the parameters of the components of mixtures of normal distributions / I., A.B. Geva // Pattern Recognition Letters – 1988. – 9(2). – P. 77-86.
76. Жуковская В.М. Факторный анализ в социально-экономических исследованиях / В.М. Жуковская. — М.: Статистика, 1976. – 152 с.
77. Virvou M., Savvopoulos A., Tsihrintzis G.A. and Sotiropoulos D.N. Constructing Stereotypes for an Adaptive e-Shop Using AIN-Based Clustering / M. Virvou, A. Savvopoulos, G.A. Tsihrintzis and D.N. Sotiropoulos // International Conference on Adaptive and Natural Computing Algorithms: Adaptive and Natural Computing Algorithms – 2007. – P. 837-845/
78. Dibike Y.B. Model Induction with Support Vector Machines: Introduction and Applications / Y.B. Dibike, S. Velickov, D. Solomatine and M.B. Abbott // J. Comp. In Civ. Engrg. – 2001. – 15(3) – P. 208-216.
79. Noble W.S. What is a support vector machine? / W.S. Noble // Nature Biotechnology – 2006. – 24(12) – P. 1565-1567.
80. Zoeller E.A. Pattern Recognition: Theory and Application / E.A. Zoeller – Nova Science Pub Inc., 2008. – 403 p.

ДОДАТКИ

Додаток А

Програмний код формування особистого профілю клієнта

```
.dialog('sign up', [  
  function (session) {  
    session.send("Let's create an account!");  
    builder.Prompts.text(session, "Please, enter your first name");  
    function (session, results) {MongoClient.connect(url, function(err, db) {  
      if (err) throw err;  
      var myobj = { FirstName: results.response, LastName: ".", UserName: "."};  
      var dbase = db.db("users");  
      dbase.collection("profiles").insertOne(myobj, function(err, res) {  
        if (err) throw err;  
        console.log("1 document inserted");  
        db.close();});});  
      builder.Prompts.text(session, "... and your last name");  
      function (session, results) {  
        MongoClient.connect(url, function(err, db) {  
          if (err) throw err;  
          var dbase = db.db("users");  
          var myquery = { LastName: "." };  
          var newvalues = { $set: { LastName: results.response } };  
          dbase.collection("profiles").updateOne(myquery, newvalues,  
          function(err, res) {  
            if (err) throw err;  
            console.log("1 document updated");
```

```

db.close();});});
session.replaceDialog('addUserName');}
]).triggerAction({ matches: /^(sign up)/i });
bot.dialog('addUserName', [
function (session, args) {
if (args && args.reprompt) {
builder.Prompts.text(session, "Oops, this login is already taken :( I know, it is annoying, but please, choose another one");}
else {session.send("Thanks! Almost done! I need a login, it must be unique and more than five characters");
builder.Prompts.text(session, "Please, enter a login");}},
function (session, results) {session.userData.userName = results.response;
MongoClient.connect(url, function (err, db) {if (err) throw err;
var dbase = db.db("users");
var query = { UserName: session.userData.userName };
dbase.collection("profiles").find(query).toArray(function (err, result)
{if (err) throw err;
console.log(result);
if (result.length !== 0)
session.replaceDialog('addUserName', { reprompt: true });
else {var myquery = {UserName: "."};
var newvalues = {$set: {UserName: results.response}};
dbase.collection("profiles").updateOne(myquery, newvalues,
function (err, res) {
if (err) throw err;
console.log("1 document updated");
db.close();});
session.send(session.userData.userName + ", welcome to our

```

online-shop! I will help you to choose all what you need (and even more)!

```
"); session.replaceDialog('start');});});
```

```
session.send().endDialog();}]);
```

```
bot.dialog('log in', [
```

```
function (session, args) {
```

```
if (args && args.reprompt) {
```

```
builder.Prompts.text(session, "I do not find this login :( Please, try again or say 'sign up' to create new one");
```

```
} else {builder.Prompts.text(session, "Please, enter your login");}}
```

```
function (session, results) {
```

```
MongoClient.connect(url, function (err, db) {
```

```
if (err) throw err;
```

```
var dbase = db.db("users");
```

```
session.userData.userName = results.response;
```

```
var query = { UserName: session.userData.userName };
```

```
dbase.collection("profiles").find(query).toArray(function (err, result)
```

```
{if (err) throw err;
```

```
console.log(result);
```

```
if (result.length === 0)
```

```
session.replaceDialog('log in', { reprompt: true });
```

```
else {session.send(session.userData.userName + ", welcome back to our
```

```
online-shop! I will help you to choose all what you need (and even more)!
```

```
:)");
```

```
session.replaceDialog('start');});});
```

```
session.send().endDialog();}]).triggerAction({ matches: /^(log in)/i });
```


Програмний код виділення кластерів

```
function findPrices(session, callback) {  
  MongoClient.connect(url, function (err, db) {  
    if (err) throw err;  
    var dbase = db.db("items");  
    for (var i = 0; i < 5; i++) {  
      var query = {cluster: String(i)};  
      dbase.collection("clusteringSorted").find(query).toArray(function (err,  
        result) {  
        if (err) throw err;  
        // console.log(result);  
        var maxMin = {cluster: result[0].cluster, max: 0, min: 500000};  
        for (var j = 0; j < result.length; j++) {  
          if (parseInt(result[j].price) < maxMin.min) {  
            maxMin.min = parseInt(result[j].price);  
            // maxMin.cluster = parseInt(result[j].cluster); }  
          if (parseInt(result[j].price) > maxMin.max) {  
            maxMin.max = parseInt(result[j].price);  
            // maxMin.cluster = parseInt(result[j].cluster); } }  
          session.conversationData.prices.push(maxMin); }); } } );  
        callback(null, session.conversationData.prices); }  
    bot.dialog('Choose price', [  
      function (session) {  
        console.log(session.conversationData.prices);  
        var clusters = [];  
        var index;
```

```

for (var i = 0; i < session.conversationData.prices.length; i++) {
  var min = 500000;
  for (var j = 0; j < session.conversationData.prices.length; j++) {
    if (clusters.length === 0) {
      if (session.conversationData.prices[j].min < min) {
        index = session.conversationData.prices[j].cluster;
        min = session.conversationData.prices[j].min; } }
      else { var number = parseInt(clusters[i - 1]);
        for (var k = 0; k < session.conversationData.prices.length;
          k++) { if (parseInt(session.conversationData.prices[k].cluster)
            === number)
              break; }
          if (session.conversationData.prices[j].min < min &&
            session.conversationData.prices[j].min >
            session.conversationData.prices[k].min) {
              index = session.conversationData.prices[j].cluster;
              min = session.conversationData.prices[j].min; } } }
          clusters.push(index);
          session.conversationData.clusters = clusters;
          console.log(session.conversationData.clusters);
          var msg = new builder.Message(session)
            .text("Okay, let's choose the price!")
            .suggestedActions(
              builder.SuggestedActions.create(
                session, [builder.CardAction.imBack(session, "cluster 1",
                  session.conversationData.prices[session.conversationData.clusters[0]].min +
                    " - " + session.conversationData.prices[session.conversationData.
                    clusters[0]].max + " UAH"),

```

```
builder.CardAction.imBack(session, "cluster 2",
session.conversationData.prices[session.conversationData.clusters[1]].min +
" - " +
session.conversationData.prices[session.conversationData.clusters[1]].max +
" UAH"),
builder.CardAction.imBack(session, "cluster 3",
session.conversationData.prices[session.conversationData.clusters[2]].min +
" - " +
session.conversationData.prices[session.conversationData.clusters[2]].max +
" UAH"),
builder.CardAction.imBack(session, "cluster 4",
session.conversationData.prices[session.conversationData.clusters[3]].min +
" - " +
session.conversationData.prices[session.conversationData.clusters[3]].max +
" UAH"),
builder.CardAction.imBack(session, "cluster 5",
session.conversationData.prices[session.conversationData.clusters[4]].min +
" - " +
session.conversationData.prices[session.conversationData.clusters[4]].max +
" UAH") ]]);
builder.Prompts.choice(session, msg, [
session.conversationData.prices[session.conversationData.clusters[0]].min +
" - " +
session.conversationData.prices[session.conversationData.clusters[0]].max +
" UAH",
session.conversationData.prices[session.conversationData.clusters[1]].min +
" - " +
session.conversationData.prices[session.conversationData.clusters[1]].max +
```



```

    " UAH",
    session.conversationData.prices[session.conversationData.clusters[2]].min +
    " - " +
    session.conversationData.prices[session.conversationData.clusters[2]].max +
    " UAH",
    session.conversationData.prices[session.conversationData.clusters[3]].min +
    " - " +
    session.conversationData.prices[session.conversationData.clusters[3]].max +
    " UAH",
    session.conversationData.prices[session.conversationData.clusters[4]].min +
    " - " +
    session.conversationData.prices[session.conversationData.clusters[4]].max +
    " UAH" ] },
    function(session, results) {
        session.replaceDialog('cluster'); }
    ]).triggerAction({ matches: /(choose price)$/i });
    function findCluster(query, callback) {
        MongoClient.connect(url, function (err, db) {
            if (err) throw err;
            var dbase = db.db("items");
            dbase.collection("clusteringSorted").find(query).toArray(function (err,
            myDoc) {if (err) {callback(err, null);} else {
            callback(null, myDoc);}});});
            bot.dialog('cluster', [function (session) {
            session.send("Okay, that's what I found");
            console.log(session.conversationData.clusters);
            var query = { cluster: String(session.conversationData.clusters[0])};
            console.log(query);

```

```

var myDoc;
findCluster(query, function (err, content) {
  if (err) {console.log(err);} else {myDoc = content;
  // console.log(myDoc);
  var msg = new builder.Message(session);
  msg.attachmentLayout(builder.AttachmentLayout.carousel);
  var arr = [];
  for (var i = 0; i < myDoc.length; i++) {
  // console.log(myDoc);
  var title = myDoc[i].producer + " " + myDoc[i].model;
  var subtitle = myDoc[i].display;
  var price = myDoc[i].price + " UAH";
  var link = myDoc[i].link;
  arr[i] =
  new builder.HeroCard(session)
  .title(title)
  .subtitle(subtitle)
  .text(price)
  .images([builder.CardImage.create(session, link)])
  .buttons([
  builder.CardAction.imBack(session, "buy " +
  myDoc[i].producer +
  " model: " + myDoc[i].model + " price: " +
  myDoc[i].price + " UAH", "Buy")])])
  msg.attachments(arr);
  session.send(msg);} }]).triggerAction({ matches: /^(cluster)$/i });

```

Програмний код виділення товарів за пріоритетами

```
dialog('buyButtonClick', [function (session, args) {// console.log(args);  
var collection = "phones";  
if (args.intent.matched.input.match(/(buy.*?model.*?price.*?UAH)/)) {  
var utterance = args.intent.matched.input;  
var model = utterance.slice((utterance.indexOf(":") + 2),  
(utterance.lastIndexOf("price") - 1));  
var price = utterance.slice((utterance.lastIndexOf(":") + 1),  
(utterance.length));  
var producer = utterance.split(" ")[1];  
if (model && price && producer) {  
// Initialize cart item  
if (producer === "Case")  
collection = "cases";  
else if (producer === "Screen Protector")  
collection = "accessories";  
var item = session.dialogData.item = {  
model: model,  
price: price,  
producer: producer,  
link: ""};  
if (!session.userData.cart) {  
session.userData.cart = [];}  
session.userData.cart.push(item);  
if (session.userData.cart[session.userData.cart.length -  
1].producer === "Case")
```



```

collection = "cases";
else if (session.userData.cart[session.userData.cart.length -
1].producer === "Screen Protector")
collection = "accessories";
else
collection = "phones";
findItem(0, collection,
session.userData.cart[session.userData.cart.length - 1].model, function (err,
content, j) {if (err) console.log(err);
else {session.userData.cart[session.userData.cart.length -
1].link = content[0].link;}});
// Send confirmation to users
session.send(item.producer + " " + item.model + " has been added to
your cart. You can say 'cart' to see +
" all products in your cart");
session.replaceDialog('recommend').endDialog();
} else {
// Invalid product
session.send("I'm sorry... That product wasn't
found.").endDialog();}
else {session.send("I'm sorry... That product wasn't found. Say 'buy' to
choose product").endDialog();}}].triggerAction({ matches: /(buy)/i });
dialog('cart', [
function (session) {
if (session.userData.cart.length !== 0) {
var msg = new builder.Message(session);
msg.attachmentLayout(builder.AttachmentLayout.carousel);
var arr = [];

```

```

for (var i = 0; i < session.userData.cart.length; i++) {
if (session.userData.cart[i].producer !== "Case" &&
session.userData.cart[i].producer !== "Screen Protector") {
var title = session.userData.cart[i].model;
var subtitle = session.userData.cart[i].producer;} else {subtitle =
session.userData.cart[i].model;
title = session.userData.cart[i].producer;}
var price = session.userData.cart[i].price;
var link = session.userData.cart[i].link;
arr[i] = new builder.HeroCard(session)
.title(title)
.subtitle(subtitle)
.text(price)
.images([builder.CardImage.create(session, link)])
.buttons([
builder.CardAction.imBack(session, "delete " +
session.userData.cart[i].producer +
" model: " + session.userData.cart[i].model,
"Delete"))])
msg.attachments(arr);
session.send(msg);
var total = 0;
for (i = 0; i < session.userData.cart.length; i++) {
total += parseInt(session.userData.cart[i].price);}
session.send("Total: " + total + " UAH");
session.send("You can make an order! Just say me 'order'").endDialog();}
else {session.send("Your cart is empty. Say 'buy' to choose
something.").endDialog();}}]).triggerAction({ matches: /(cart|show cart)/i });

```