

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ
КАФЕДРА СИСТЕМНОГО АНАЛІЗУ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Системний аналіз факторів, що впливають на
лояльність клієнтів у роздрібній торгівлі»

на здобуття освітнього ступеня бакалавра
зі спеціальності 124 Системний аналіз
(код, найменування спеціальності)
освітньо-професійної програми Системний аналіз
(назва)

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

_____	Ольга УДОДИЧ
(підпис)	Ім'я, ПРІЗВИЩЕ здобувача

Виконав:	здобувачка вищої освіти гр. САД-41
	Ольга УДОДИЧ
	Ім'я, ПРІЗВИЩЕ
Керівник:	Михайло КУЗМІЧ
ст викладач каф. СА.	Ім'я, ПРІЗВИЩЕ

Рецензент:	_____
науковий ступінь, вчене звання	Ім'я, ПРІЗВИЩЕ

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**
Навчально-науковий інститут телекомунікацій

Кафедра Системного аналізу

Ступінь вищої освіти бакалавр

Спеціальність 124 Системний аналіз

Освітньо-професійна програма Системний аналіз

ЗАТВЕРДЖУЮ

Завідувач кафедри Ровіл НАФЄЄВ

_____ Ім'я, ПРІЗВИЩЕ
«_____» _____ 2024р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

_____ Удодич Ольга Іванівна

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: Системний аналіз факторів, що впливають на лояльність клієнтів у роздрібній торгівлі

керівник кваліфікаційної роботи ст. викладач каф. СА Михайло КУЗМІЧ

(Ім'я, ПРІЗВИЩЕ, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологій від «_____» _____ 2024р. № _____

2. Строк подання кваліфікаційної роботи «_____» _____ 2024р.

3. Вихідні дані до кваліфікаційної роботи: Інформаційно-аналітичний метод, інтернет-ресурси, хмарні сервіси, .

Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Аналіз підходів та методів управління

2. Розробка комплексної динамічної системи управління

4. Перелік ілюстративного матеріалу: презентація

5. Дата видачі завдання «_____» _____ 2024р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Збір даних		Виконано
2	Обробка матеріалу		Виконано
3	Аналіз підходів та методів управління		Виконано
4	Розробка комплексної динамічної системи управління		Виконано
5	Висновки за результатами аналізу		Виконано
6	Розробка презентації		Виконано
7	Передзахист		Виконано
8	Здача в деканат		Виконано

Здобувачка вищої освіти

(підпис)

Ольга УДОДИЧ
(Ім'я, ПРІЗВИЩЕ)

Керівник кваліфікаційної роботи

(підпис)

Михайло КУЗМІЧ
(Ім'я, ПРІЗВИЩЕ)

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**

Навчально-науковий інститут телекомунікацій

**ПОДАННЯ
ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ
ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
на здобуття освітнього ступеня бакалавра**

Направляється здобувачка Удодич О.І. до захисту кваліфікаційної роботи
(*прізвище та ініціали*)
за спеціальністю 124 Системний аналіз
(*код, найменування спеціальності*)
освітньо-професійної програми Системний аналіз
(*назва*)
на тему: «Системний аналіз факторів, що впливають на лояльність клієнтів у роздрібній торгівлі».

Кваліфікаційна робота і рецензія додаються.

Директор ННІ

(*підпис*)

Владислав КРАВЧЕНКО

(*Ім'я, ПРІЗВИЩЕ*)

Висновок керівника кваліфікаційної роботи

Здобувачка Удодич Ольга Іванівна під час написання кваліфікаційної роботи показала гарну теоретичну підготовку, володіння необхідними знаннями з системного аналізу, користуватися навчальною, довідковою і науково-технічною літературою. Працюючи над завданнями, які доручались керівником, проявила ініціативність, сумлінність та хист до інженерної роботи.

Все це дозволяє оцінити виконану кваліфікаційну роботу здобувачки Удодич О.І. на оцінку «відмінно» та присвоїти їй кваліфікацію «Бакалавр з системного аналізу».

Керівник кваліфікаційної роботи

(*підпис*)

Михайло КУЗМІЧ

(*Ім'я, ПРІЗВИЩЕ*)

«____» _____ 2024 року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувачка Удодич О.І. допускається до захисту даної роботи в Експертній комісії.

Завідувач кафедру Системний аналіз

(*назва*)

(*підпис*)

Ровіл НАФЄЄВ

(*Ім'я, ПРІЗВИЩЕ*)

ВІДГУК РЕЦЕНЗЕНТА **на кваліфікаційну бакалаврську роботу**

здобувачка вищої освіти Удодич Ольга Іванівна

(прізвище, ім'я, по батькові)

на тему «Системний аналіз факторів, що впливають на лояльність клієнтів у роздрібній торгівлі»

Актуальність.

Сучасна економічна ситуація характеризується зростаючим рівнем конкуренції серед українських та світових компаній і високою волатильністю купівельних переваг. А саме в умовах, коли на ринку існує велика кількість пропозицій, йде гостра конкурентна боротьба за кожного покупця між підприємствами у будь-якій галузі, тому компанії вимушені шукати нові ефективні способи управління.

Сьогодні більшість торгових компаній зіткнулося з падінням попиту і уповільненням темпів збуту товарів. Особливо явно ця тенденція почала проявлятися після початку пандемії, що спричинила переорієнтацією кінцевих споживачів на економію і зростання заощаджень.

Позитивні сторони.

1. Розроблено методичний підхід і інструментарій прогнозування клієнтської поведінки для максимально ефективного стимулювання попиту та клієнтоорієнтованості підприємства на основі підходу оцінки CLV клієнта..
2. За допомогою стохастичних Pareto/NBD та Gamma-Gamma моделей був здійснений прогноз майбутньої кількості та цінності покупок клієнтів компанії, був ідентифікований ймовірний відтік клієнтів..

Недоліки.

1. В роботі мають місце деякі неточності в перекладі технічних термінів.
2. В текстовій частині пояснювальної записки дипломної роботи надано не всі посилання на джерела інформації.

Висновок: *кваліфікаційна бакалаврська робота заслуговує оцінку "відмінно", а здобувачка Удодич Ольга Іванівна заслуговує присвоєння кваліфікації: «Бакалавр з системного аналізу».*

Рецензент:

науковий ступінь, вчене звання

підпис

Ім'я, ПРІЗВИЩЕ

РЕФЕРАТ

Робота об'ємом 63 сторінок, яка містить 15 ілюстрацій, 6 таблиць, 20 джерел за переліком посилань.

У ході дипломної роботи було досліджено теоретичні аспекти аналізу та управління клієнтською базою підприємства. Розроблено методичний підхід і інструментарій прогнозування клієнтської поведінки для максимально ефективного стимулювання попиту та клієнтоорієнтованості підприємства на основі підходу оцінки CLV клієнта.

За допомогою стохастичних Pareto/NBD та Gamma-Gamma моделей був здійснений прогноз майбутньої кількості та цінності покупок клієнтів компанії, був ідентифікований ймовірний відтік клієнтів.

Для зручного аналізу і управління результатів був створений динамічний дашборд.

Ключові слова: КЛІЄНТООРІЄНТОВАНІСТЬ, CRM, ЖИТТЄВА ЦІННІСТЬ, КЛІЄНТСЬКА БАЗА, ЗАЛУЧЕННЯ ТА УТРИМАННЯ КЛІЄНТІВ, PARETO/NBD МОДЕЛЮВАННЯ, ГАММА-ГАММА МОДЕЛЮВАННЯ.

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1. АНАЛІЗ ПІДХОДІВ І МЕТОДІВ УПРАВЛІННЯ КЛІЄНТСЬКОЮ БАЗОЮ КОМПАНІЇ	11
1.1 Економічне обґрунтування необхідності управління клієнтською базою.....	11
1.2 Підходи до управління клієнтською базою компанії на основі показника лояльності клієнта	14
1.3 Підходи до управління клієнтською базою компанії на основі показника довгострокової вартості клієнта.....	20
1.4. Порівняльна характеристика підходів до управління клієнтською базою компанії	25
1.5. Метод кластеризації клієнтської бази.....	29
РОЗДІЛ 2. РОЗРОБКА КОМПЛЕКСНОЇ ДИНАМІЧНОЇ МОДЕЛІ УПРАВЛІННЯ КЛІЄНТСЬКОЮ БАЗОЮ КОМПАНІЇ	33
2.1 Постановка задачі управління клієнтською базою компанії.....	33
2.2 Характеристика інформаційної бази дослідження.....	35
2.3 Прогнозування кількості майбутніх покупок клієнтів, їх CLV та ідентифікація «відтоку» за допомогою Парето/NBD моделювання	36
2.4 Програмна реалізація моделі прогнозування майбутньої життєвої цінності клієнта	44
ВИСНОВКИ.....	56
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	57
Додаток А. ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ (Презентація).....	59

ВСТУП

Сучасна економічна ситуація характеризується зростаючим рівнем конкуренції серед українських та світових компаній і високою волатильністю купівельних переваг. А саме в умовах, коли на ринку існує велика кількість пропозицій, йде гостра конкурентна боротьба за кожного покупця між підприємствами у будь-якій галузі, тому компанії вимушені шукати нові ефективні способи управління.

Сьогодні більшість торгових компаній зіткнулося з падінням попиту і уповільненням темпів збуту товарів. Особливо явно ця тенденція почала проявлятися після початку пандемії, що спричинила переорієнтацією кінцевих споживачів на економію і зростання заощаджень.

Остання світова криза показала, що в умовах нестабільного фінансового становища компанії, які мають значний рівень лояльності клієнтів, змогли утримати лідируючі позиції на ринку.

Це в черговий раз доводить, що методи управління, засновані на масовому знеособленому виробництві, знову поступаються місцем клієнтоорієнтованому веденню бізнесу або CRM - Customer Relationship Management, спрямованому на розуміння потреб своїх клієнтів і підвищення ефективності роботи з ними [1].

Перехід підприємств до клієнтоорієнтованого ведення бізнесу дозволяє компанії збільшити свій прибуток і ефективність, підвищити свою прибутковість за рахунок збільшення виручки від існуючої клієнтської бази та оптимізації операційних витрат.

Важливість якісного управління клієнтською базовою компанією в процесі ефективного управління компаніями відзначається у своїх роботах і таких авторів, як Майер [2], Пепперс і Роджерс [3], Сьюелл [4].

Предметом дипломної роботи є розробка моделі для прогнозування ймовірності повторної покупки клієнта, кількості та цінності його майбутніх транзакцій задля ефективного управління.

Об'єктом дипломної роботи є клієнтська база підприємства та її структура залежно від цінності для компанії.

Мета дипломної роботи – це побудова та реалізація універсальної динамічної моделі оптимізації управління клієнтською базою компанії із врахуванням поведінкових характеристик та цінності груп клієнтів.

Для досягнення цієї мети слід поетапно виконати завдання:

- Розробити підхід до кластеризації клієнтської бази компанії на основі поведінкових характеристик груп клієнтів;
- побудувати динамічну модель оцінки довгострокової вартості клієнтської бази компанії;
- виявити фактори, що впливають на динаміку поведінкових характеристик груп клієнтів;
- сформулювати критерії та інструменти оптимального управління клієнтською базою з точки зору компанії на основі купівельних характеристик груп клієнтів;
- провести апробацію розробленої моделі;
- розробити програмний інструментарій для визначення оптимального плану проведення заходів по управлінню клієнтською базою компанії.

Під час написання дипломної роботи були використані наступні методи та моделі. Перш за все, цінність клієнтів була оцінена та спрогнозована на основі стохастичних моделей Pareto/NBD та Gamma-Gamma. Результати даних моделей були використані для сегментації клієнтської бази на групи за подібними характеристиками за допомогою методу k-середніх.

Інформаційною базою дипломної роботи є транзакційна історія та дані про клієнтів мережі АЗК «WOG». Але, так як дані підприємства

конфіденційні, то у роботі продемонстрована реалізація моделі на основі відкритих тестових даних про транзакції інтернет-магазину.

У результаті була спрогнозована оцінка доходу, який отримає компанія в майбутні 30 днів від груп клієнтів з подібними характеристиками. Були ідентифіковані клієнти, які з високою ймовірністю не здійзнять повторну покупку в компанії для подальшої персональної пропозиції від компанії у вигляді підвищеної знижки з метою їх повернення і утримання.

Такий підхід був реалізований на реальній клієнтській базі мережі АЗК «WOG»: клієнтам з високою ймовірністю «відтоку» була запропонована вища знижка.

Ефект оцінювався на основі А/В-тестування, в ході якого вся вибірка клієнтів була розділена на контрольну та тестову групу, і рівень їх повернення порівнювався.

Таким чином був виділений чистий ефект у розмірі 8% клієнтів, які здійснили повторну покупку в компанії, що збільшило прибуток компанії на 2%.

Отже, отримані результати дають змогу оптимально управляти базою клієнтів і обирати маркетингові плани та стратегії для цільових груп.

РОЗДІЛ 1 АНАЛІЗ ПІДХОДІВ І МЕТОДІВ УПРАВЛІННЯ КЛІЄНТСЬКОЮ БАЗОЮ КОМПАНІЇ

1.1 Економічне обґрунтування необхідності управління клієнтською базою

У західній науковій літературі проблема управління клієнтською базою широко розглядалася різними фахівцями як у галузі маркетингу, так і в галузі управління бізнесом. Проблема про ефективне управління базою клієнтів з'явилася ще з популяризацією перших програм лояльності у світі у 1980-х роках та досліджується дотепер.

Згідно з дослідженням Фредеріка Рейххельда, консультанта Bain & Company, 5%-ве зростання рівня утримання клієнтів може призвести до збільшення прибутку від 25% до 95% [5].

Збільшення коефіцієнту лояльності впливає на такі ключові параметри росту прибутку підприємства:

1. Динаміка чисельності споживачів.

Між рівнем лояльності і коефіцієнтом вибуття клієнтської бази існує статистично показана зворотна залежність - чим вище коефіцієнт лояльності, тим нижче коефіцієнт вибуття клієнтів.

2. Зростання прибутку на одного клієнта (середнього чеку).

Однією з переваг довготривалих відносин з клієнтом є те, що в більшості галузей, в тому числі в нафтопродуктовій, обсяги продажів мають тенденцію до зростання з часом. Основними причинами цього є:

- Збільшення кількості перехресних продажів;

Адже з часом клієнт краще дізнається асортимент компанії та, як правило, розширює діапазон товарів, що купує;

- Зростання потреб клієнта з часом.

Наприклад, зміна соціального становища, поява дітей і т.д. призводить до появи нових потреб, задоволення яких з більшою часткою ймовірності буде відбуватися в компанії, з якою клієнт вже співпрацює і ставиться до неї лояльно.

3. Зниження витрат на залучення клієнтів.

До витрат на залучення клієнтів відносяться: реклама, цінові знижки, відсоток комісійних для торгових агентів і т.д. Очевидно, що підвищення індексу лояльності веде до зниження темпів зростання вибуття клієнтів. Таким чином, для підтримки певного рівня клієнтської бази потрібно залучати менше нових клієнтів, що призводить до зменшення величини витрат компанії на залучення споживачів.

4. Зниження операційних витрат.

Як правило, величина операційних витрат на обслуговування клієнта з часом скорочується. Персонал витрачає менше часу на обслуговування старих клієнтів, так як вони володіють достатніми знаннями про цю компанію, і компанія, в свою чергу, багато знає про звички клієнта.

Вплив рівня лояльності споживачів на операційні витрати особливо виражений в роздрібній торгівлі та дистриб'юторській діяльності. Так підприємство, яке має клієнтську базу з високим рівнем плинності, повинен тримати на складі більш значні запаси товарів, ніж підприємство, яке обслуговує постійних покупців. Постійна клієнтська база допомагає організувати управління запасами, мінімізувати змушені знижки для зменшення величини надлишків продукції на складі і спростити прогнозування попиту на товар в майбутньому.

5. Збільшення кількості рекомендацій.

Споживачі, що з'являються після особистих рекомендацій, зазвичай більш привабливі як клієнти, тобто забезпечують велику прибутковість і залишаються взаємодіяти з компанією довше, ніж клієнти, які реагують на рекламу конкурентів, розпродажі і цінові просування.

Клієнтоорієнтованої ведення бізнесу, в свою чергу, вимагає реорганізації існуючих, заснованих на продуктоорієнтованої стратегії, бізнес-процесів компанії, зміни принципів планування і розробку нових підходів до управління організацією.

Новий клієнтоорієнтований підхід до управління бізнесом тягне за собою зміну методів розрахунку ключових показників ефективності ведення бізнесу. Зокрема, прибуток компанії повинен розраховуватися в залежності від розміру і якості клієнтської бази, а не продукту. Витрати компанії повинні враховуватися в розрахунку на клієнта, а не одиницю товару. Крім модифікації існуючих показників для ефективного управління клієнтською базою потрібна розробка нових показників, що відображають якість клієнтів компанії і ефективність взаємодії з ними.

На поточний момент розроблений ряд окремих показників, які дозволяють оцінювати вартість і якість наявної клієнтської бази. Але все більше компаній, як українських, так і зарубіжних, розуміють важливість комплексного підходу до управління групами клієнтів. У наслідок чого виникає необхідність розробки комплексної моделі управління, що дозволяє не тільки прогнозувати прибутковість груп клієнтів в майбутньому, але також оцінювати ефективність заходів, пов'язаних з управлінням клієнтською базою, і гнучко налаштовувати інструменти управління клієнтами.

Так, в рамках CRM стратегії управління компанією, можна виділити два ключові підходи до оцінки ефективності управління клієнтами і вартості клієнтської бази компанії. Один з напрямків, якому віддають перевагу західні компанії, - підхід на основі показника лояльності клієнтів. Другий підхід - оцінка довгострокової вартості клієнта.

1.2 Підходи до управління клієнтською базою компанії на основі показника лояльності клієнта

Моделі управління на основі показника лояльності клієнтів представлені в роботах Ж.-Ж. Ламбена [6], Ф. Рачхелда [7], Д. Аакера [8], Я. Хофмайера і Б.Райса [9].

Поняття «лояльність», як і більшість понять в економіці та інших галузях, має багато варіантів визначення. Одне із перших визначень лояльності до бренду було сформоване у 1923 році. Згідно з даним визначенням, «Споживач, лояльний до бренду, – це людина, яка купує ваш бренд у 100% випадків». [10]

Сучасне трактування терміну «лояльність» не настільки категоричне. Так, згідно з загальновизнаною визначення лояльними є ті споживачі, які позитивно ставляться до діяльності компанії, її продуктів і послуг, персоналу тощо. Це позитивне ставлення виражається перевагою, що віддається продуктам даної компанії в порівнянні з конкурентами, причому ця перевага стійка в часі і характеризується здійсненням повторних покупок.

Споживач здійснює повторні покупки, тому що саме ця торгова марка добре задовольняє його потреби, або тому що у нього формується особиста прихильність марці. Прихильність марці також може бути наслідком її емоційного впливу на споживача або її впливу на самооцінку споживача.

Також існує визначення лояльності як міри прихильності споживача бренду. Тоді лояльність показує, який ступінь ймовірності, що споживач змінить бренд на бренд іншої компанії, особливо, коли він зазнає змін за ціновими або будь-яким іншим показникам. При зростанні лояльності знижується схильність споживачів до сприйняття дій конкурентів.

Отже, деякі фахівці розуміють лояльність як ступінь нечутливості поведінки покупців товару або послуги до дій конкурентів - таким як зміни

цін, товарів, послуг, супроводжувана емоційної прихильністю до товару чи послуги.

Крім емоційної складової, також виділяють і раціоналізм споживачів. Саме наявність чи відсутність раціональної компоненти і представляє різницю між поведінковою лояльністю і лояльністю, пов'язаною з прив'язаністю.

Я.Хофмайер і Б.Райс [9], з огляду на раціональний компонент в поведінці клієнтів, виділяють три типи лояльності:

1. Прихильність.

Лояльність, пов'язана зі ставленням, проявляється як зацікавленість споживача в покупці саме у даній компанії, передбачає емоційну залученість і прихильність споживача до компанії, повне задоволення взаємовідносинами з цією організацією, взаємодія з компанією протягом необмеженого часу. Споживач високо оцінює компанію, задоволений співпрацею з нею, відчуває зацікавленість в ній і емоційну прихильність, але не має можливості часто купувати в цій організації (економічні фактори або відсутність компанії на ринку). При появі такої можливості він набуває саме у неї.

2. Поведінкова лояльність.

Поведінкова лояльність проявляється при здійсненні покупок в даній компанії на постійній основі, але при відсутності прихильності. Споживач або не задоволений співпрацею з компанією, у якої він купує товар/послуги, або ставиться до неї з байдужістю. Незважаючи на це, споживач змушений взаємодіяти з даною компанією через відсутність на ринку «улюбленого» бренду або з причин економічного характеру. При першій же нагоді споживач переходить до компанії, до якої він відчуває емоційну прихильність. Зустрічаються випадки, коли у споживача взагалі немає брендів, до яких він відчуває прихильність.

3. Змішаний тип.

Споживач постійно здійснює покупки у даній організації і відчуває при цьому емоційну прихильність і глибоку задоволеність. Таким чином, в даному випадку можна говорити про поєднання прихильності і лояльності.

Отже, можна зробити висновок, що лояльність – це показник, що характеризує відношення клієнта до компанії, до її товарів та послуг. Включає в себе як емоційний, так і раціональний компонент. Ступінь лояльності знаходить своє відображення в купівельній поведінці клієнта, його намір продовжувати свої взаємини з компанією і готовність до співпраці з нею.

Можна виділити два основних напрямки оцінки рівня лояльності клієнта в залежності від типу лояльності. Перший із них для оцінки раціональної лояльності: використовуються кількісні оцінки, засновані на аналізі купівельної поведінки клієнта:

1. Кількість повторних покупок.

Одним із основних показників, що характеризують лояльність клієнта до компанії, є коефіцієнт повторних покупок. Даний показник дає досить об'єктивну оцінку поведінкової лояльності клієнта для компаній, характерною особливістю яких є короткий споживчий цикл здійснення покупок (менше року). У такому випадку показник повторних покупок може бути використаний для прийняття оперативних рішень.

2. «Доля гаманця» клієнта.

Згідно з даним методом оцінюється доля компанії в бюджеті споживача, тобто рівень його лояльності розраховується як відношення кількості покупок у даної фірми до всіх аналогічних покупок клієнта. Але даний метод досить складно реалізувати у реальних бізнес умовах, адже обсяги затрат клієнта на товари галузі у конкурентів найчастіше є невідомими для компанії.

3. Індекс Entis-Paul.

Спроба врахувати можливість вибору клієнтом компанії була зроблена в 1970 році Б.Еніс і Г.Паулі [11]. Модель оцінки лояльності споживачів отримала назву Enis-Paul Index і включає в себе такі параметри, як:

- Доля бюджету на продукти даної категорії, яку клієнт витрачає у компанії;
- Кількість зміни нашої компанії і іншими компаніями галузі за визначений період;

- Кількість можливостей для зміни компаній;
- Кількість компаній у галузі, у яких клієнт здійснював покупки за період;
- Кількість компаній, доступних для клієнта;
- Загальна кількість відвідувань всіх компаній за визначений термін.

Дана модель є найскладнішою в реалізації, адже потребує багато вхідних даних про клієнта, але дозволяє достатньо широко оцінити рівень лояльності клієнтів.

Другим напрямком є оцінка емоційної складової лояльності (або прихильності) з використанням показників на основі балів, які розраховуються, спираючись на проведення опитувань клієнтів. До таких моделей належать:

1. Конверсійна модель.

Дана модель для вимірювання ступеня лояльності клієнта була запропонована Я.Хофмайером і Б.Райсом [9] і включає в себе чотири основні показники:

- Задоволеність торговою маркою.

Даний показник є важливою характеристикою взаємовідносин клієнта і компанії, але носить односторонній характер. Низька задоволеності клієнта, як правило, говорить і про низьку частоту покупок клієнта, в той час як високий ступінь задоволеності не обов'язково призводить до високої купівельної активності. Проте, даний показник необхідний для комплексної оцінки ступеня лояльності.

- Альтернативи.

Для оцінки ступеня лояльності необхідно враховувати також такий показник, як наявність альтернатив (конкуруючих компаній/брендів). У ряді випадків високий ступінь лояльності може бути пов'язаний з відсутністю товарів-субститутів, і в разі появи більш якісної продукції/послуги клієнт може перейти до конкуруючої фірми.

2. Методика Ламбена.

У своїй роботі Ж.-Ж. Ламбен [6] наводить опис процедури оцінки ступеня задоволеності споживачів. В основу пропонованої методики покладено концепцію мультиатрибутивної моделі відносини. Відповідно до цієї концепції, практична корисність товару для споживача визначається набором атрибутів (істотних невід'ємних властивостей даного предмета).

За класифікацією Ламбена товар має наступні три групи атрибутів:

- Ядерні (характеризують основну, функціональну корисність товару);
- Периферійні (додаткові, но пов'язані з основною функцією товару);
- Додані послуги (не пов'язані з основною функцією, але розширюючі можливості задоволення потреб споживачів при покупці даного товару).

Процедура оцінки ступеня задоволення включає в себе три етапи:

1. Оцінка загальної (комплексної) задоволеності продуктом.
2. Оцінка задоволеності по кожній характеристиці по 10-бальній шкалі і ступені важливості даної характеристики для клієнта.
3. Оцінка намірів здійснити повторну покупку.

За результатами отриманих оцінок розраховуються показники задоволеності товаром по найбільш важливих характеристиках і ступінь розкиду думок респондентів, а також виявляються характеристики, які грають вирішальну роль при виборі споживача.

Підводячи підсумок аналізу напрямку до розробки показників ефективного управління клієнтською базою компанії на основі показника лояльності, можна відзначити, що в даний час немає єдиної думки про можливість і об'єктивність виміру рівня лояльності клієнтів. Так, згідно з дослідженням, проведеним Ф.Райхелдом в своїй книзі «Ефект лояльності» [7], від 60 до 80% покупців, які відмовилися від послуг компанії, під час нещодавно проведеного опитування показали, що вони задоволені або дуже

задоволені покупкою. Але при цьому частка повторних покупок цих клієнтів була не така висока. Підходи до управління клієнтською базою компанії на основі показника лояльності можна віднести до двох категорій:

1. Стратегія заохочення найбільш лояльних клієнтів компанії.

В рамках даної моделі відбувається сегментація клієнтської бази на групи в залежності від значення показника лояльності клієнта і відбувається перерозподіл маркетингових ресурсів на заохочення найбільш лояльних з них.

2. Стратегія підвищення рівня лояльності клієнтів компанії.

Дана стратегія націлена на довгострокові ефекти від проведених маркетингових заходів. При реалізації даної стратегії компанії потрібно виявити причини низького рівня задоволеності своїх клієнтів і розробити методи для їх усунення. В рамках даного підходу бюджет маркетингових заходів витрачається, в першу чергу, на найменш лояльні групи клієнтів з метою підвищити їх ступінь задоволеності роботою компанії.

Слід зазначити, що моделі управління клієнтською базою на основі показника лояльності досить прості і інтуїтивно зрозумілі, але несуть скоріше експертний, ніж об'єктивний характер. Винятком є модель повторних покупок, але дана модель розраховує результат тільки по минулим даними (тобто постфактний аналіз), і не дозволяє прогнозувати майбутню купівельну поведінку. Іншими словами, область застосування даних оцінок – аналіз поточного стану клієнтської бази. Дані показники не дозволяють прогнозувати стан клієнтської бази на майбутньому інтервалі часу, тому не можуть бути використані для розробки довгострокової стратегії підприємства.

Крім того, дані моделі не враховують витрати компанії на залучення і обслуговування клієнтів, що обмежує можливості компанії при виборі оптимального розподілу фінансових ресурсів на управління клієнтами. Тому компанії, які при розробці стратегії управління клієнтами ґрунтуються тільки на підвищенні рівня лояльності та задоволеності споживачів, можуть виявитися заручниками своєї лояльності: з одного боку, максимальний рівень

задоволеності клієнтів і високі оцінки за якістю обслуговування, з іншого, низькі показники прибутковості і рентабельності вкладених коштів.

1.3 Підходи до управління клієнтською базою компанії на основі показника довгострокової вартості клієнта

Для подолання обмежень моделей управління клієнтською базою компанії, заснованих на показнику лояльності клієнтів, розглянутих вище, використовується показник довгострокової вартості клієнтської бази (CLV – customer lifetime value).

Довгострокова вартість клієнта – це сукупність чистого доходу, очікуваного від клієнта в майбутньому. В роботі Бергера і Насра [12] наводиться визначення показника CLV як чистого прибутку (збитку) фірми, що розраховується за всіма транзакціями клієнта в даній компанії.

Таким чином, показник цінності клієнта для фірми представляє собою дохід, отриманий від клієнта протягом всього періоду взаємодії клієнта і компанії, за вирахуванням вартості залучення, продажу і обслуговування цього клієнта з урахуванням тимчасової вартості грошей.

Для розрахунку показника CLV використовується наступна базова формула (1.1):

$$CLV = \sum_{i=1}^t \frac{D_i - Z_i}{(1 + d)^{i-1}} \quad (1.1)$$

де i – номер періоду, в який здійснюється розрахунок надходжень від клієнтів;

D_i – дохід від клієнта в період i ;

Z_i – загальні витрати на отримання доходу D_i в період i ;

t – загальна кількість періодів протягом життєвого циклу клієнта;

d – ставка дисконтування.

Також часто зустрічаються модифікації наведеної вище формули розрахунку, які детально описані в роботах Бергера і Насра [12].

Основними перевагами використання методу розрахунку CLV для управління клієнтською базою компанії є те, що:

- Розрахунок CLV дозволяє компаніям створити систему ранжування всіх клієнтів залежно від їх цінності для компанії. Це, в свою чергу, дозволяє виділяти більше маркетингових засобів і часу на обслуговування та збереження найцінніших з них. Наприклад, якщо керівництво має намір провести кампанію по утриманню клієнтів, то йому необхідно мати уявлення про їх цінність. Як правило, більшу частину доходів тій чи іншій організації забезпечує відносно невелика кількість клієнтів (близько 20% всієї клієнтської бази). Вони є для компанії найціннішими, і, отже, саме на цю групу клієнтів необхідно зробити акцент у майбутній маркетинговій компанії для отримання максимальної ефективності.

- Аналіз показника CLV дозволяє співставити прибуток, який приносить клієнт компанії, і ті сили і засоби, які були витрачені на пошук і утримання цього клієнта. Іншими словами, це дозволяє оцінити максимальну планку витрат для залучення і утримання цих клієнтів.

Таким чином, показник CLV дозволяє надати об'єктивну оцінку ефективності управління клієнтської бази компанії, чисельно вимірну і порівнянну між собою в часі і в сукупності з іншими компаніями, як того ж сегмента економіки, так і різних галузей.

Аналіз існуючих моделей управління клієнтською базою компанії на основі показника CLV описаний в основному в роботах зарубіжних авторів. За основу для дослідження були взяті роботи Діпака і Сідтхарта [13], а також праця Фадера і Харді [14].

Регресійні моделі. Один з підходів до управління клієнтською базою компанії спирається на побудову регресійних моделей для оцінки майбутніх доходів від клієнтів.

Загальна формула для оцінки CLV має наступний вигляд (1.2):

$$g(y_i) = f(x_i) + e_i \quad (2) \quad (1.2)$$

де $g(y_i)$ – функція вартості клієнтської бази компанії;

y_i – параметр для розрахунку CLV клієнта;

$f(x_i)$ – регресійна функція незалежних показників x_i ;

e_i – параметр помилки моделі;

Даний метод був описаний у роботі Малтхауса і Блаттберга [15], де для оцінки виду регресійної функції використовували 3 методи:

- лінійну регресію, побудовану з використанням методу найменших квадратів (OLS);
- лінійну регресію, побудовану з використанням модифікованого методу найменших квадратів (IRLS);
- метод нейронних ланцюгів.

Апробація моделі проводилася за даними компаній, що працюють в 5 різних секторах економіки. За критерій вибору оптимальної моделі використовувалися ймовірність помилки незаохочень дійсно цінного клієнта і ймовірність помилки заохочення звичайного (нецінного) клієнта. Завдання управління клієнтською базою зводиться до задачі визначення кількості клієнтів, що володіють максимальної CLV, з метою їх подальшого заохочення за лояльну поведінку. На жаль, незважаючи на гнучкість і універсальність регресійних моделей, вони мають ряд недоліків:

- модель вимагає адаптації і перерахунку основних коефіцієнтів для кожної конкретної організації;
- модель чутлива до вихідних даних і може видавати помилкові результати при незначних відхиленнях в даних;
- точність прогнозування різко знижується зі збільшенням кількості періодів прогнозування, тобто модель не придатна для прогнозування купівельної поведінки на тривалому горизонті часу.

Стохастичні моделі (NBD модель та її модифікації). У роботі розглянута група стохастичних моделей управління клієнтською базою (NBD

та її модифікації), які в першу чергу, орієнтовані на вирішення завдання визначення активних/неактивних клієнтів, тобто на визначення ймовірності того, що клієнт активний і зробить покупку в наступний період часу з урахуванням історії його транзакцій в минулому. В якості критерію оптимального управління клієнтською базою виступає розподіл фінансових ресурсів компанії в залежності від ймовірності скоєння клієнтом покупки в наступний період часу. У разі якщо клієнт неактивний, тобто ймовірність того, що він зробить покупку невелика, то є сенс витратити фінансові ресурси на взаємодію з цим клієнтом з метою його повернення і утримання.

Перша імовірнісна модель моделювання купівельної поведінки клієнтів компанії була розроблена в 1972 р. Еренбергом [16].

Основні припущення NBD - моделі (Negative Binomial Distribution) полягають в наступному:

- клієнт здійснює покупки в довільний момент часу, тобто випадково в межах середньої частоти здійснення транзакцій (стаціонарного значення). Даний процес може бути описаний з допомогою розподілу Пуассона;
- частота здійснення транзакцій у кожного клієнта своя і може бути описана за допомогою Гамма-розподілу.

Таким чином, загальний вигляд NBD-моделі представлений наступним чином (1.3):

$$Pr_{NDB}(X = x; r; \alpha) = \left(\frac{x + r - 1}{x} \right) \left(\frac{\alpha}{\alpha + 1} \right)^r \left(\frac{1}{\alpha + 1} \right)^x \quad (1.3)$$
$$x = 0, 1, 2, \dots; \quad r, \alpha > 0$$

де Pr_{NDB} – ймовірність здійснення клієнтом заданої кількості покупок x ;

x – кількість покупок;

r, α – параметри розподілу.

Нехай M і S^2 – вибіркове середнє та вибіркова дисперсія відповідно, t – кількість інтервалів часу спостереження. Тоді оцінки параметрів розподілу r, α мають такий вигляд (1.4):

$$\hat{\alpha} = \frac{Mt}{(S^2 - M)}; \quad \hat{r} = \frac{\hat{\alpha}M}{t} \quad (1.4)$$

Шмітляйн, Моррісон і Коломбо [17] у своїй роботі модифікували завдання управління клієнтською базою і зробили спробу виявити активних/неактивних клієнтів. Тобто клієнтів, які з високою часткою ймовірності можуть зробити покупку в наступний період часу, і тих, хто припинив з тих чи інших причин свою співпрацю з компанією.

Основне припущення моделі: в випадковий період часу клієнт може перестати здійснювати покупки і стає неактивним. Причин цьому може бути багато: від зміни смакових переваг, фінансової неспроможності та зміни географічного розташування до фізичної смерті клієнта. Тобто відносини клієнта з фірмою складається з двох фаз: клієнт активний («живий») протягом невизначеного часу, а потім переходить в стан «постійно неактивний». Купівельна поведінка клієнта в активний період може бути описано за допомогою NBD-моделі. Період часу, після якого клієнт переходить в неактивний стан, характеризується як випадкова складова за допомогою експоненціального розподілу, а неоднорідність моментів відтоку різних клієнтів характеризується гамма-розподілом. Таким чином, ймовірність того, що досліджуваний клієнт "активний", визначається за такою формулою (1.5):

$$\begin{aligned} \alpha &> \beta, \\ P(\text{Alive} | r, s, \alpha, \beta, x, t) &= \quad (1.5) \\ &= \left\{ 1 + \frac{s}{r+x+s} \times \left[\left(\frac{\alpha+T}{\alpha+t} \right)^{r+x} \left(\frac{\beta+T}{\alpha+t} \right)^x F(a_1, b_1, c_1, z_1(t)) - \left(\frac{\beta+T}{\alpha+t} \right)^x F(a_1, b_1, c_1, z_1(T)) \right] \right\}^{-1}, \\ a_1 &= r + x + s, \\ b_1 &= s + 1, \\ c_1 &= r + x + s + 1, \end{aligned}$$

де r, s, α, β – параметри моделі;

t – момент часу останньої транзакції клієнта;

T – час з моменту останньої транзакції клієнта;

$F(a_1, b_1, c_1, z_1(t))$ – функція гіпергеометричного (гаусівського)

розподілу;

x – кількість покупок клієнта в період $(0, T]$, при чому час останньої покупки клієнта $t \leq T$.

У момент часу 0 клієнт вважається «активним».

Оскільки поєднання гамма розподілу з експоненціальним відомо як Парето-розподіл то дана модель отримала назву Парето/NBD-моделі.

На жаль, запропоновані імовірнісні моделі мають ряд обмежень, визнаних самими авторами:

- обмеження на вимоги до вихідних даних. Зокрема, модель може видавати помилкові результати в разі використання в якості вхідних даних моделі дані про транзакції клієнта на інтервалі більше 2 років. Тому в разі, якщо клієнт має тривалий інтервал між покупками, то дана модель може потенційно прибуткового клієнта оцінити як неприбуткового;
- модель може бути використана для розрахунку CLV тільки для окремого клієнта, що ускладнює її використання в компаніях з великим числом клієнтів;
- у порівнянні з регресійними дана модель містить набагато складніший принцип розрахунку параметрів, що перешкоджає її широкому використанню для вирішення завдань оперативного управління.

1.4 Порівняльна характеристика підходів до управління клієнтською базою компанії

На жаль, запропоновані імовірнісні моделі мають ряд обмежень, визнаних самими авторами:

Для проведення порівняльного аналізу розглянутих в перших двох розділах моделей до управління клієнтською базою компанії необхідно

виявити сильні і слабкі сторони кожної з моделей, в тому числі, з точки зору відповідності вимог бізнесу до таких моделей.

Узагальнені результати проведеного аналізу існуючих підходів до управління клієнтською базою компанії і схема існуючих напрямків дослідження завдання управління клієнтською базою і підходів до її вирішення наведені нижче (рис. 1.1).



Рисунок 0.1 Схема існуючих напрямків дослідження задачі управління клієнтською базою

Результати порівняльного аналізу двох підходів по таким показниками, як ефективність використання, складність розрахунку та практика застосування в компаніях наведені в таблиці нижче (Таблиця 1.1).

Так, в частині ефективності використання для вирішення поставлених

завдань показник лояльності клієнта має слабші показники в порівнянні з показником CLV за рахунок того, що він не враховує витрати компанії на управління клієнтською базою і не має єдиної методики оцінки.

При цьому за складністю розрахунку показник довгострокової вартості клієнта вимагає більш значних обчислювальних витрат, так як необхідне прогнозування грошового потоку від клієнтів і детальний облік витрат на обслуговування клієнта. Показник лояльності заснований на оцінці минулого поведінки.

За рахунок останнього показника – простота розрахунку, по частоті використання в компаніях моделі управління клієнтською базою на основі показника лояльності випереджають моделі на основі показника довгострокової вартості клієнтською базою.

Таблиця 1.1 – Порівняльна характеристика двох підходів до управління базою клієнтів

Характеристика	Показник лояльності	Показник CLV
Ефективність використання	Середня (не відображає витрати на управління клієнтською базою)	Висока (не відображає якісну сторону управління клієнтською базою)
Складність реалізації	Низька	Висока (потребує прогнозування грошового потоку клієнтів)
Практика застосування	Часто	Рідко

Але необхідно відзначити, що без введення додаткового параметра, що характеризує співвідношення витрат на клієнта і його реального доходу, може виникнути ситуація, при якій витрати на утримання та подальше збільшення лояльності клієнта перевищать сукупний дохід від цього клієнта.

Таким чином, для вирішення завдань компанії щодо максимізації прибутку від клієнтської бази при скороченні витрат на маркетингові заходи найбільш ефективними є моделі, що використовують значення показника CLV в якості критерію оптимальності.

Моделі на основі показника лояльності зможуть відобразити тільки

якісні характеристики клієнтської бази, тому їх доцільно застосовувати як додатковий критерій для прийняття управлінських рішень.

Результати порівняння моделей управління клієнтською базою компанії на основі показника довгострокової вартості клієнта за такими параметрам як складність розрахунку, точність прогнозування, чутливість до вихідних даних і гнучкість настройки параметрів моделі наведені в таблиці (Таблиця 1.2).

Таблиця 1.2 – Порівняльна характеристика підходів до управління клієнтською базою на основі показника CLV

Параметр порівняння	Регресійна модель	Парето/NBD модель
Складність розрахунку	Середня	Висока
Точність прогнозування	Середня	Висока
Чутливість до даних	Висока	Висока
Гнучкість	Середня	Низька

За параметром складності розрахунку менших обчислювальних витрат потребує регресійна модель, ніж Парето/NBD-моделі

При цьому найбільшою точністю прогнозування володіють Парето/NBD-моделі, і менш точні показники на довгостроковому інтервалі часу у регресійних моделях.

Низькі показники точності регресійних моделей пояснюються також високою чутливістю до вихідних даних. Парето/NBD- моделі також вимогливі до якості вхідної інформації.

За параметром гнучкості і часу адаптації для використання в компанії більшу перевагу отримують регресійні моделі, а імовірнісні Парето/NBD моделі є менш гнучкими і більш вимогливими до впровадження.

Зважаючи на всі переваги та недоліки, описані вище, для реалізації поставленої у роботі задачі була обрана стохастична Парето/NBD модель та підхід управління клієнтською базою компанії на основі показника CLV.

1.5 Метод кластеризації клієнтської бази

Кластеризація — це об'єднання об'єктів в однорідні групи. Іншими словами, це пошук схожих об'єктів, які відрізняються від інших спостережень за своїми характеристиками.

При якісному розбитті розкид значень всередині груп повинен бути мінімальним, а груповий розкид максимальним. Виділені групи повинні бути стійкими і відтворюваними.

Одним із найпопулярніших і найпростіших методів кластеризації є метод k -середніх.

Даний метод відноситься до категорії «Навчання без учителя».

Метод кластеризації k -середніх був винайдений в 1950-х роках математиком Гуго Штайнгаузом [18] і майже одночасно Стюартом Ллойдом [19]. Особливу популярність отримав після виходу роботи МакКвіна.

Навчання без вчителя — один зі способів машинного навчання, при вирішенні яких випробовувана система спонтанно навчається виконувати поставлене завдання, без втручання з боку експериментатора. З точки зору кібернетики, є одним з видів кібернетичного експерименту [20].

Мета методу — розділити n спостережень на k кластерів, так щоб кожне спостереження належало до кластера з найближчим до нього середнім значенням.

Принцип алгоритму полягає в пошуку таких центрів кластерів та наборів елементів кожного кластера при наявності деякої функції $f(x)$, що виражає якість поточного розбиття множини на k кластерів, коли сумарне квадратичне відхилення елементів кластерів від центрів цих кластерів буде найменшим.

Іншими словами, обчислюється похибка кожної точки даних, тобто її евклідова відстань до найближчого центроїда, а потім обчислюється загальна сума квадратних помилок. Маючи два різні набори кластерів, які утворюються двома різними прогонами K -середніх, ми віддаємо перевагу тому, що має

найменшу похибку в квадраті, оскільки це означає, що прототипи (центроїди) цієї кластеризації є кращим представленням точок у їх кластері.

Отже цільовою функцією є мінімізація суми квадратичних помилок моделі (1.6):

$$I = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \rightarrow \min \quad (1.6)$$

де k – кількість кластерів,

n – кількість спостережень,

x_i – спостереження i ,

c_j – центроїд кластера j .

Головною перевагою методу є його простота та швидкість виконання.

Метод k -середніх більш зручний для кластеризації великої кількості спостережень, ніж інші методи.

В початковий момент роботи алгоритму довільним чином обираються центри кластерів, далі для кожного елемента множини ітеративно обраховується відстань від центрів з приєднанням кожного елемента до кластера з найближчим центром. Для кожного з отриманих кластерів обчислюються нові значення центрів, намагаючись при цьому мінімізувати функцію I , після чого повторюється процедура перерозподілу елементів між кластерами.

Алгоритм методу k -середніх заключається в таких етапах (рис. 2.7):

1. Вибір k точок як центрів кластерів поки не завершиться процес зміни центрів кластерів;
2. Зіставити кожну інформаційну точку з кластером, відстань до центра якого мінімальна;
3. Переконатися, що в кожному кластері міститься хоча б одна точка. Для цього кожний порожній кластер потрібно доповнити довільною точкою, що розташована «далеко» від центра кластера;
4. Центр кожного кластера замінити середнім від елементів кластера.

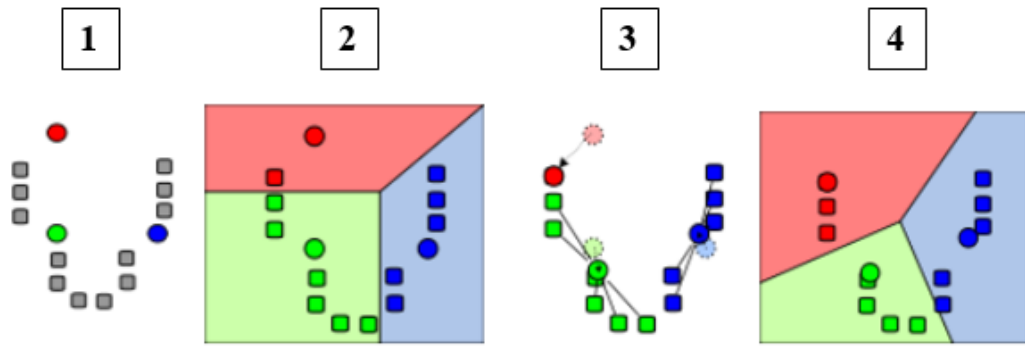


Рисунок 2.7 –Принципу алгоритму кластеризації методом к-середніх

Одним із методів вибору значення k є метод ліктя. У цьому методі ми будемо запускати кластеризацію k -Means для діапазону значень k , скажімо ($k=1$ до 10) і обчислювати відстань між точками всередині кластера (1.7):

$$W_k = \sum_{r=1}^k \frac{1}{n_r} D_r \quad (1.7)$$

Де k – кількість кластерів,

n_r – кількість точок в кластері,

D_r – сума евклідових відстаней між точками в кластері (1.8):

$$D_r = \sum_{i=1}^{n_r-1} \sum_{j=1}^{n_r} \|d_i - d_j\|^2 \quad (1.8)$$

Потім будується лінійна діаграма для значень W_k , і точка, у якій графік починає суттєво вирівнюватися, – це шуканий «лікоть» (рис. 2.8).

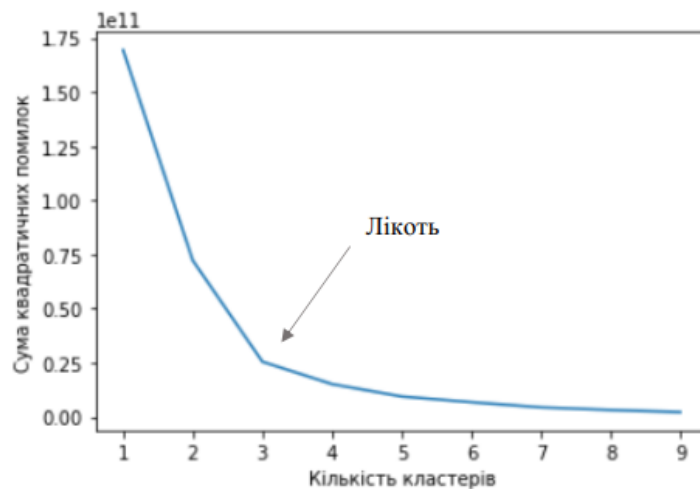


Рисунок 2.8 – Графік зміни W_k залежно від кількості кластерів

Незважаючи на очевидні переваги методу, він має суттєві недоліки:

- Результат класифікації сильно залежить від випадкових початкових позицій кластерних центрів;
- Алгоритм чутливий до викидів, які можуть викривлювати середнє;
- Кількість кластерів повинна бути заздалегідь визначена дослідником.

Отже, за методологічну основу для створення комплексної моделі управління клієнтською базою та прогнозування майбутньої кількості покупок клієнта було взято стохастичні моделі Pareto/NBD та Гамма-Гамма. За результатами даних прогнозів буде здійснена кластеризація клієнтів методом к-середніх.

РОЗДІЛ 2 РОЗРОБКА КОМПЛЕКСНОЇ ДИНАМІЧНОЇ МОДЕЛІ УПРАВЛІННЯ КЛІЄНТСЬКОЮ БАЗОЮ КОМПАНІЇ

2.1 Постановка задачі управління клієнтською базою компанії

Незважаючи на досить широкую лінійку розроблених моделей для управління клієнтською базою, практика впровадження повноцінних моделей управління клієнтською базою компанії досить невелика. Це пов'язано, в першу чергу, з трудомісткістю і високими витратами на реалізацію комплексної моделі управління клієнтською базою компанії. Також проблема криється і в обмеженнях самих моделей.

Як правило, потреба в управлінні своєю клієнтською базою компанії виникає у компаній з розміром клієнтської бази від декількох тисяч до мільйонів клієнтів. В цьому випадку впровадження методів і підходів, спрямованих на управління одним клієнтом, не раціональне, і потрібні моделі, адаптовані на роботу з групами клієнтів.

Важливим обмеженням в існуючих дослідженнях виступають і параметри кластеризації клієнтської бази. Більшість авторів пропонує в якості критерію поділу на групи клієнтів розглядати кількість покупок і давність скоєння останньої покупки клієнтом. При цьому не враховуються такі параметри купівельної поведінки клієнтів, як сума здійснюваних покупок, цінність клієнта для компанії і час взаємодії клієнта і компанії.

Ще одним важливим критерієм для компанії при виборі моделі управління кластерами клієнтів, є можливість динамічного управління клієнтськими кластерами. Останнім часом збільшилася кількість досліджень, спрямованих на спроби динамічного прогнозування чисельності клієнтської бази, але комплексного рішення, що враховує динаміку зміни всіх показників розрахунку довгострокової вартості клієнтської бази (доходи від клієнтських сегментів і чисельності клієнтської бази), так і не було представлено.

Таким чином, компанії потребують комплексної динамічної моделі управління групами клієнтів. Дана модель повинна дозволяти:

- Прогнозувати майбутнє значення CLV для клієнта;
- Ідентифікувати найлояльніших клієнтів з максимальним CLV;
- Ідентифікувати відтік шляхом розрахунку прогнозованої ймовірності здійснення повторної покупки в компанії;
- Здійснювати кластеризацію клієнтів за їх цінністю для компанії з метою заохочення і утримання найцінніших та повернення відтоку за допомогою цільових персоналізованих маркетингових заходів;
- Розрахунок прогнозованого прибутку компанії за період.

Постановка задачі управління клієнтською базою компанії базується на таких основних припущеннях:

Припущення 1: Цінність клієнта можна спрогнозувати за допомогою стохастичних Pareto/NBD та Gamma-Gamma моделей.

Припущення 2: За допомогою стохастичних Pareto/NBD та Gamma-Gamma моделей можна розрахувати ймовірність повторної покупки клієнта і таким чином виявити відтік з високою ймовірністю для подальших маркетингових заходів з метою його повернення і утримання.

Припущення 3: Всіх клієнтів можна розділити на K кластерів, що не перетинаються, залежно від їх цінності для компанії.

Припущення 4. Зміна купівельної поведінки клієнтів: інтенсивність переходу між кластерами, частота покупок і величина середнього чека, відбувається під впливом наступних факторів:

- зовнішнього впливу ринку;
- внутрішнього впливу маркетингових заходів компанії;
- зміни власних потреб клієнтів незалежно від зовнішнього світу.

Критерій оптимального управління клієнтською базою – максимізація показника довгострокової цінності (CLV) клієнтської бази компанії за інтервал часу T .

Інструментами управління клієнтської бази є маркетингові заходи, які

спрямовані на стимулювання попиту груп клієнтів з метою збільшення їх прибутковості на довгостроковому інтервалі часу. У даній роботі за приклад взяте надання підвищеної знижки для ймовірного відтоку.

2.2 Характеристика інформаційної бази дослідження

Для апробації теоретичних результатів роботи була розроблена динамічна модель управління клієнтською базою для компанії. Модель була реалізована засобами програмування мовою Python з використанням вбудованих бібліотек на прикладі тестового транзакційного набору даних інтернет магазину з вільного доступу мережі Інтернет. [21]

Цей набір даних “Online Retail II” містить усі транзакції, здійснені в умовному Інтернет-магазині з 01.12.2009 по 09.12.2011. Компанія в основному продає унікальні сувенірні вироби. Складається з 1 067 371 записів, кожен з яких представляє одну транзакцію. Приклад записів з даного набору представлено на рис. 2.1.

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

Рисунок 2.1 – Приклад записів з набору даних для апробації моделі

Де представлені такі необхідні поля :

- «Invoice» – номер замовлення (чеку);
- «StockCode» – код товару;
- «Description» – короткий опис товару;
- «Quantity» – кількість товарів в чеку;
- «InvoiceDate» – дата і час замовлення;

- «Price» – ціна одиниці товару;
- «Customer ID» – ідентифікаційний номер клієнта.
- «Country» – країна.

Був проведений EDA аналіз (розвідувальний аналіз даних). Після чого дані були очищені від шуму, пустих та аномальних значень.

У результаті EDA аналізу були зроблені такі висновки про набір даних:

- Більшість клієнтів з Великобританії, далі йдуть Німеччина, Ірландія та Франція;
- Середня ціна продажів (середній чек) у Сінгапурі найвища, далі йдуть Норвегія та Мальта;
- Дані більшості країн є незбалансованими, коли йдеться про параметр «Ціна», тому що спостерігається велика кількість екстремальних значень;
- Всього придбано 44 876 унікальних товарів;
- Більшість транзакцій відбулись у листопаді;
- Найбільше транзакцій було здійснено у 2010 році;
- Наприкінці 1-го тижня та на початку 3-го тижня люди, як правило, купують більше;
- Найпопулярнішим днем для покупок є четвер, потім вівторок та середа.

У додатку 1 представлені детальні дослідження набору даних, графіки розподілів та діаграми.

2.3 Прогнозування кількості майбутніх покупок клієнтів, їх CLV та ідентифікація «відтоку» за допомогою Парето/NBD моделювання

Зважаючи на всі переваги та недоліки моделей для прогнозування CLV клієнтів, описані у розділі 1, для реалізації поставленої задачі була обрана

Парето/NBD модель. Вибір був здійснений на основі таких факторів, як висока ефективність використання та висока точність прогнозу у порівнянні з іншими моделями.

Парето/NBD модель базується на 5 припущеннях [22]:

- Кількість транзакцій клієнта за період t описується розподілом Пуассона з параметром транзакцій λ (рис.2.2). Це еквівалентно припущенню, що час між транзакціями розподіляється експоненційно зі швидкістю транзакції λ , тобто

$$f(t_j | t_{j-1}; \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, \quad (2.1)$$

$$t_j > t_{j-1} \geq 0$$

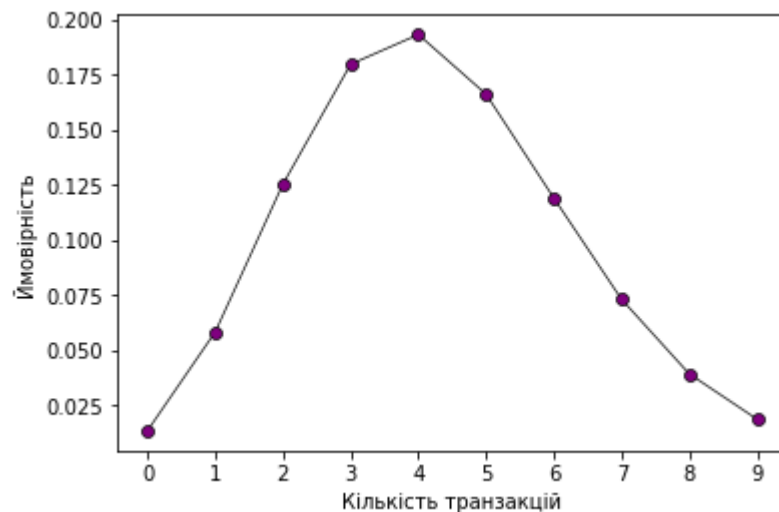


Рисунок 2.2 – Візуалізація розподілу Пуассона з параметром $\lambda = 4,3$

Тут ми маємо припустити, що наш випадковий клієнт має швидкість транзакцій $\lambda = 4,3$.

Тоді ймовірність здійснення 4 покупок у випадковому 12-місячному періоді становить 19%, 4% – ймовірність, що буде здійснено 8 покупок, і тд.

- Гетерогенність λ описується Гамма розподілом з параметрами r (форма) і α (масштаб) (рис. 2.3):

$$f(\lambda | r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda \alpha}}{\Gamma(r)} \quad (2.2)$$

$$\lambda > 0$$

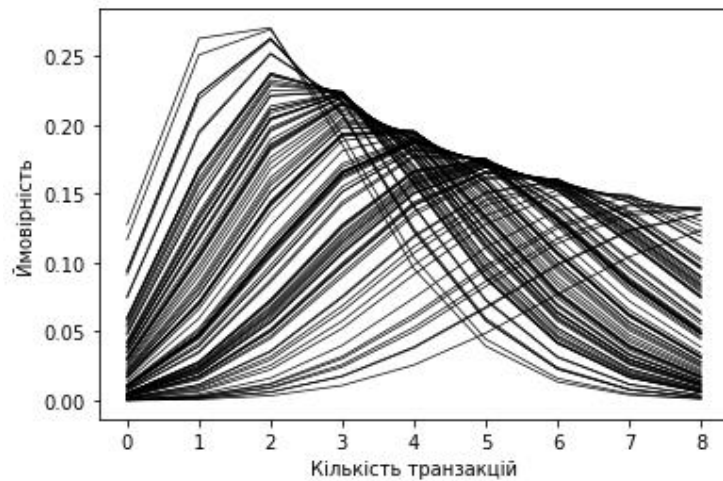


Рисунок 2.3 – Візуалізація розподілу Пуассона для 100 клієнтів, де λ кожного клієнта залежить від гамма-розподілу з $r = 9$ і $\alpha = 0.5$

Як вказано у припущеннях, кожен клієнт має власну ймовірність покупок x раз за даний інтервал часу.

- Клієнт може стати неактивним після будь-якої транзакції з ймовірністю P , і точка їх відтоку розподілена між покупками по геометричному закону (рис. 2.4):

$$P(\text{відтік після } j \text{ транзакції}) = p(1 - p)^{j-1},$$

$$j = 1, 2, 3, \dots$$

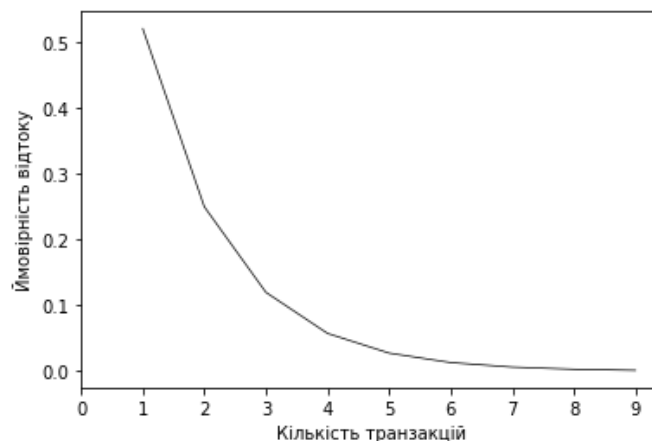


Рисунок 2.4 – Візуалізація геометричного розподілу при $p = 0.52$

Якщо припустити, що наш клієнт стає неактивним із ймовірністю $p=0,52$, то ймовірність того, що він стане неактивним після 2-ої транзакції, становить 25%, ймовірність того, що він стане неактивним після 3-ї транзакції, складає 12 %

Отже, чим більше клієнт купує, тим вище його ймовірність бути живим.

- Гетерогенність P має Бета розподіл з параметрами форми α і β (рис. 2.5):

$$f(p | a, b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a, b)} \quad (2.3),$$

$$0 \leq p \leq 1,$$

де $B(a, b)$ – бета функція, яка може бути виражена через гамма функції:

$$B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$$

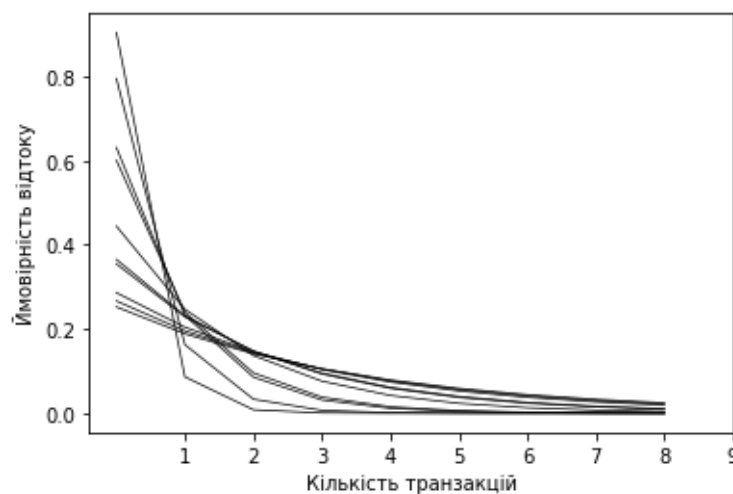


Рисунок 2.5 – Візуалізація геометричного розподілу, де p – бета розподіл з параметрами $a = 2$ і $b = 3$

- Швидкість транзакції λ і ймовірність відтоку p відрізняються залежно від клієнта.

Постановка моделі на рівні індивідуального клієнта

Припустимо, що клієнт здійснив x транзакцій за період $(0, T]$, транзакції позначимо, як $t_1, t_2, \dots, t_x$ (рис. 2.6).

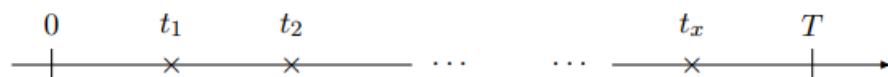


Рисунок 2.6 – Візуалізація x транзакцій клієнта за період $(0, T]$

Тоді функція правдоподібності індивідуального рівня виводиться наступним чином:

- Ймовірність того, що перша транзакція відбулася в t_1 – це стандартна експоненціальна компонента, що дорівнює

$$\lambda e^{-\lambda t_1}$$

- Ймовірність того, що друга транзакція відбулася в t_2 – це ймовірність, що клієнт залишився активним після t_1 , що дорівнює

$$(1 - p)\lambda e^{-\lambda(t_2 - t_1)}$$

...

- Ймовірність того, що транзакція x відбулася в t_x – це ймовірність, що клієнт залишився активним після t_{x-1} , що дорівнює

$$(1 - p)\lambda e^{-\lambda(t_x - t_{x-1})}$$

- Ймовірність того, що спостерігатимемо 0 покупок в період $(t_x, T]$ дорівнює сумі ймовірності того, що клієнт стане неактивним в t_x і ймовірності того, що він залишився активним, але не зробив покупок у цьому інтервалі:

$$p + (1 - p)\lambda e^{-\lambda(T - t_x)}$$

Отже функція правдоподібності описується формулою (2.4):

$$\begin{aligned} L(\lambda, p \mid t_1, t_2, \dots, t_x, T) &= \lambda e^{-\lambda t_1} (1 - p) \lambda e^{-\lambda(t_2 - t_1)} \dots \times \\ &\times (1 - p) \lambda e^{-\lambda(t_x - t_{x-1})} \{ p + (1 - p) \lambda e^{-\lambda(T - t_x)} \} = \\ &= p(1 - p)^{x-1} \lambda^x e^{-\lambda t_x} + (1 - p)^x \lambda^x e^{-\lambda T} \end{aligned} \quad (2.4)$$

Як зазначалося раніше, для моделі Парето/NBD не потрібна інформація про терміни проведення x транзакцій, достатньо підсумку на основі історії покупок клієнта ($X = x, t_x, T$).

Ми припускаємо, що всі споживачі активні на початку періоду спостереження; отже, функція правдоподібності для клієнта, який робить 0 покупок в інтервалі $(0, T]$ має вигляд (2.5):

$$L(\lambda, p \mid X = 0, T) = e^{-\lambda T} \quad (2.5)$$

Тому можемо переписати функцію правдоподібності як (2.6):

$$L(\lambda, p \mid X = x, T) = (1 - p)^x \lambda^x e^{-\lambda T} + \delta_{x>0} p(1 - p)^{x-1} \lambda^x e^{-\lambda t_x} \quad (2.6)$$

де $\delta_{x>0} = 1$ якщо $x > 0$, в іншому випадку 0.

Нехай випадкова величина $X(t)$ позначає кількість транзакцій, що відбуваються за певний період часу t (з початком 0). Для того щоб отримати формулу для розрахунку $P(X(t) = x)$, необхідно врахувати фундаментальну залежність між часом між транзакціями та кількістю транзакцій:

$X(t) \geq x \leftrightarrow T_x \leq t$, де T_x випадкова величина, що позначає час транзакції x .

Тоді формулу для розрахунку $P(X(t) = x)$ матиме вигляд (2.7):

$$P(X(t) = x) = P(\text{активний після покупки } x) \times P(T_x \leq t \text{ і } T_{x+1} > t) + \delta_{x>0} \times P(\text{неактивний після покупки } x) \times P(T_x \leq t) \quad (2.7)$$

Враховуючи, що час між транзакціями описується експоненціальним розподілом, $P(T_x \leq t \text{ і } T_{x+1} > t)$ Пуассоновська ймовірність того, що $X(t) = x$, і $P(T_x \leq t)$ – розподіл Erlang- x , отримуємо (2.8):

$$\begin{aligned} P(X(t) = x | \lambda, p) &= \\ &= (1 - p)^x \frac{(\lambda t)^x e^{-\lambda t}}{x!} \\ &+ \delta_{x>0} p (1 - p)^{x-1} \left[1 - e^{-\lambda t} \sum_{j=0}^{x-1} \frac{(\lambda t)^j}{j!} \right] \end{aligned} \quad (2.8)$$

Перейдемо до виведення формули $E[X(t)]$ для розрахунку очікуваної кількості покупок за період часу t . Так як кількість транзакцій клієнта описується Пуассоновським випадковим розподілом, то $E[X(t)] = \lambda t$, якщо клієнт активний в t . Для клієнтів, які стають неактивними в $\tau < t$, очікувана кількість транзакцій за період $(0, \tau]$ дорівнює $\lambda \tau$.

Але ймовірність, що клієнт стає неактивним в τ , спираючись на λ та p , представлена у наступному вигляді (2.9):

$$P(\tau > t) = P(\text{активний в } t | \lambda, p) = \sum_{j=0}^{\infty} (1 - p)^j \frac{(\lambda t)^j e^{-\lambda t}}{j!} = e^{-\lambda p t} \quad (2.9)$$

Це означає, що щільність часу, коли клієнт відтікає, виражається як

$$g(\tau | \lambda, p) = \lambda p e^{-\lambda p \tau}$$

Отже, очікувана кількість транзакцій за період довжини t розраховується за формулою (2.10):

$$E(X(t) | \lambda, p) = \lambda t \times P(\tau > t) + \int_0^t \lambda \tau g(\tau | \lambda, p) d\tau = \frac{1}{p} - \frac{1}{p} e^{-\lambda p t} \quad (2.10)$$

Постановка моделі для випадково обраного індивіда

Усі розроблені вище вирази залежать від швидкості транзакції λ та ймовірності відсіву p , обидва з яких не очевидні. Щоб вивести еквівалентні вирази для випадково необхідно перетворити функцію правдоподібності та сформулювати вираження ймовірності та очікуваної кількості наступним чином:

- Виразивши очікування (3) через розподіл λ та p перетворимо функцію правдоподібності для випадково обраного індивіда з історією транзакцій $(X(t) = x, t_x, T)$ (2.11):

$$L(r, \alpha, a, b | X = x, t_x, T) = \frac{B(a, b + x)}{B(a, b)} \frac{\Gamma(r + x) \alpha^r}{\Gamma(r)(\alpha + T)^{r+x}} + \\ + \delta_{x>0} \frac{B(a + 1, b + x - 1)}{B(a, b)} \frac{\Gamma(r + x) \alpha^r}{\Gamma(r)(\alpha + T)^{r+x}} \quad (2.11)$$

Парметри (r, α, a, b) можуть бути оцінені за допомогою методу максимальної правдоподібності наступним чином. Припустимо, маємо набір з N клієнтів, де клієнт i має $X_i = x_i$ транзакцій за період $(0, T_i]$ з останньою транзакцією, яка відбулася в t_{xi} . Функція правдоподібності матиме вигляд (2.12):

$$LL(r, \alpha, a, b) = \sum_{i=1}^N \ln[L(r, \alpha, a, b | X_i = x_i, t_{xi}, T_i)] \quad (2.12)$$

Дана функція максимізується стандартними процедурами числової оптимізації.

- Виразивши очікування (2.10) через розподіл λ та p ймовірність здійснення x покупок в часовому періоді t матиме вигляд:

$$P(X(t) = x | r, \alpha, a, b) = \\ = \frac{a + b - 1}{a - 1} \left[1 - \left(\frac{\alpha}{\alpha + t} \right)^r {}_2F_1(r, b; a + b - a; \frac{t}{(\alpha + t)}) \right] \quad (2.13)$$

Де ${}_2F_1()$ – гіпергеометрична функція Гауса.

Тоді функція очікуваної кількості транзакцій матиме вигляд (2.14):

$$E(Y(t)|X = x, t_x, T, r, \alpha, a, b)$$

$$= \frac{\frac{a + b + x - 1}{a - 1} \left[1 - \left(\frac{\alpha + T}{\alpha + T + t} \right)^{r+x} {}_2F_1(r + x; b + x; a + b + x - 1; \frac{t}{\alpha + T + t}) \right]}{1 + \delta_{x>0} \frac{a}{b + x - 1} \left(\frac{\alpha + T}{\alpha + t_x} \right)^{r+x}} \quad (2.14)$$

Гамма-гамма модель для оцінки CLV. До цього моменту, ми використовували тільки частоту і давність покупок клієнта. Але на додаток до цього, ми можемо застосувати грошову складову його транзакцій. Гамма-гамма модель допомагає передбачити найбільш ймовірну цінність транзакції в майбутньому.

Підсумовуючи все, маємо всі елементи, щоб визначити LTV клієнта:

$$CLV = \text{Очікувана кількість транзакцій} \times \text{Цінність транзакції}$$

Модель базується на таких загальних припущеннях [23]:

- Грошова вартість транзакції клієнта випадкова і знаходиться в межах його середньої вартості транзакції.
- Середня вартість транзакції варіюється серед клієнтів, але не варіюється з часом в межах одного клієнта.
- Середня цінність транзакції має Гамма розподіл серед клієнтів

Припустимо, що клієнт здійснив x транзакцій, грошова цінність яких $z_1, z_2, \dots, z_x$. Тоді середня цінність транзакцій матиме вигляд:

$$\bar{z} = \sum_{i=1}^x z_i / x$$

Тоді очікувана грошова вартість клієнта виражається формулою (2.15):

$$\begin{aligned} E(M|p, q, \gamma, m_x, x) &= \frac{(\gamma + m_x x)p}{px + q - 1} = \\ &= \left(\frac{q - 1}{px + q - 1} \right) \frac{\gamma p}{q - 1} + \left(\frac{px}{px + q - 1} \right) m_x \quad (2.15) \end{aligned}$$

де p та q – параметри форми,

γ – параметр масштабу гамма розподілу для транзакцій z_i ,

Як і раніше, ми можемо використовувати метод максимуму правдоподібності, щоб знайти параметри моделі.

2.4. Програмна реалізація моделі прогнозування майбутньої життєвої цінності клієнта

Для реалізації поставлених у роботі завдань були використані програмні засоби мови програмування Python.

Згідно з математичною постановкою моделей Парето/NBD та Гамма-Гамма, представленою у п. 2.3, необхідно розрахувати показники, які описуватимуть характер поведінки клієнтів.

Тому були розраховані і використані показники, які лежать в основі RFM-аналізу:

- Frequency – історична кількість покупок клієнта;
- Recency – період активності клієнта, що розраховується за формулою:

$$R = \text{Дата останньої покупки} - \text{Дата першої покупки};$$

- T – «вік» клієнта, період існування клієнта в базі, що розраховується за формулою:

$$T = \text{Дата аналізу} - \text{Дата першої покупки}$$

- monetary_value – сума вартості всіх транзакцій клієнта;

Для розрахунку показників RFM була використана вбудована бібліотека Lifetimes та функція `lifetimes.utils.summary_data_from_transaction_data()`.

Дана функція перетворює вхідний набір у таблицю з розрахованими RFM-показниками (рис. 2.7):

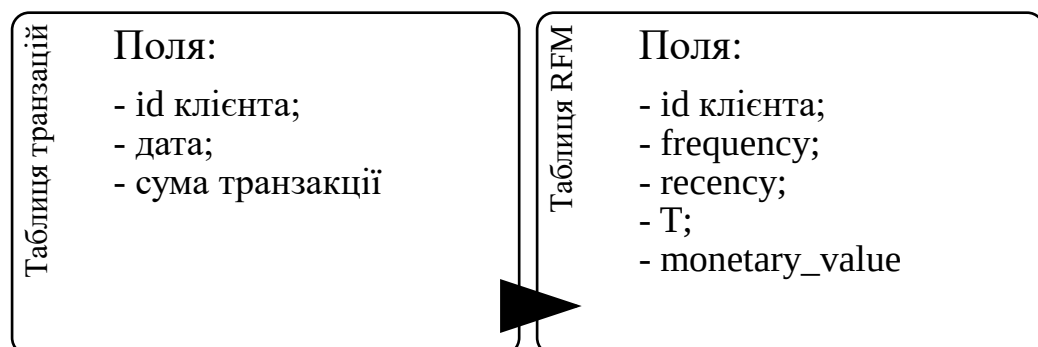


Рисунок 2.7 – Необхідне перетворення вхідної таблиці даних

Вихідні дані функції представлені на рисунку 2.8:

Customer ID	frequency	recency	T	monetary_value
12346.0	10.0	400.0	725.0	-15.468000
12347.0	7.0	402.0	404.0	717.398571
12348.0	4.0	363.0	438.0	449.310000
12349.0	4.0	717.0	735.0	1107.172500
12350.0	0.0	0.0	310.0	0.000000

Рисунок 2.8 – Приклад вихідних даних RFM-аналізу клієнтської бази

Так як основою моделей Pareto/NBD та Гамма-Гамма є припущення про випадкові розподіли вхідних показників, то необхідною складовою є їх попередній аналіз.

Для цього були побудовані діаграми та функції розподілу кожного із RFM-показників (рис. 2.9):

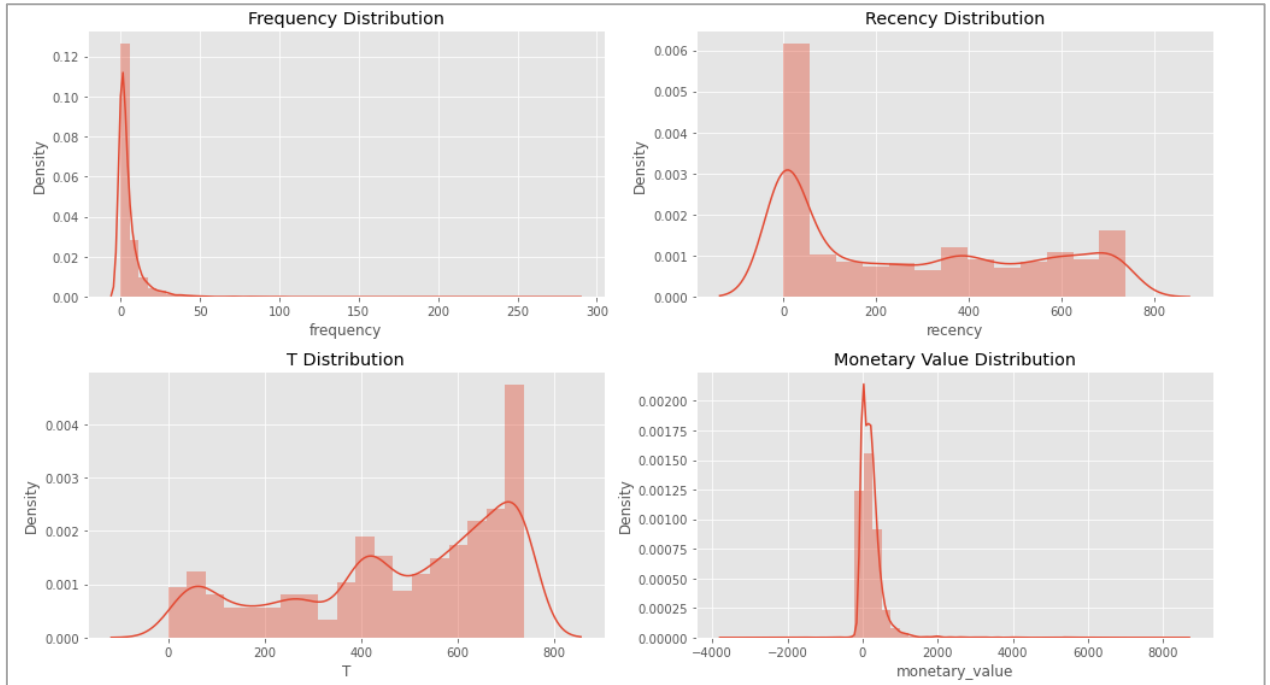


Рисунок 2.9 – Графіки розподілу показників frequency, recency, T та monetary

Отже, можемо зробити висновки, що показник frequency та monetary value мають певну кількість екстремальних значень, на відміну від recency та

T, де розподіл більш рівномірний. Аналогічний висновок можна зробити з статистичного опису показників: їх медіани та процентилів (табл. 2.1).

Таблиця 2.0.1 – Статистичний опис показників frequency, recency, T та monetary

	Customer ID	frequency	recency	T	monetary_value
count	5942.000000	5942.000000	5942.000000	5942.000000	5942.000000
mean	15316.500000	5.479636	275.772299	478.229384	228.814496
std	1715.451981	11.293673	259.830840	223.879537	363.067124
min	12346.000000	0.000000	0.000000	0.000000	-3610.500000
1%	12405.410000	0.000000	0.000000	15.000000	-40.623900
10%	12940.100000	0.000000	0.000000	89.000000	0.000000
25%	13831.250000	0.000000	0.000000	320.500000	0.000000
50%	15316.500000	2.000000	225.000000	536.000000	174.900625
75%	16801.750000	6.000000	518.000000	674.000000	314.594375
90%	17692.900000	13.000000	672.900000	731.000000	502.095625
99%	18227.590000	44.590000	734.000000	738.000000	1328.480453
max	18287.000000	284.000000	738.000000	738.000000	8513.271143

З таблиці 2.1 бачимо, що значення 90% та 99% процентилів значно відрізняються у показників frequency та monetarty_value, що підтверджує наявність екстремальних значень.

На рис. 2.10 зображені блокові діаграми, які демонструють діапазон і розподіл числових даних.

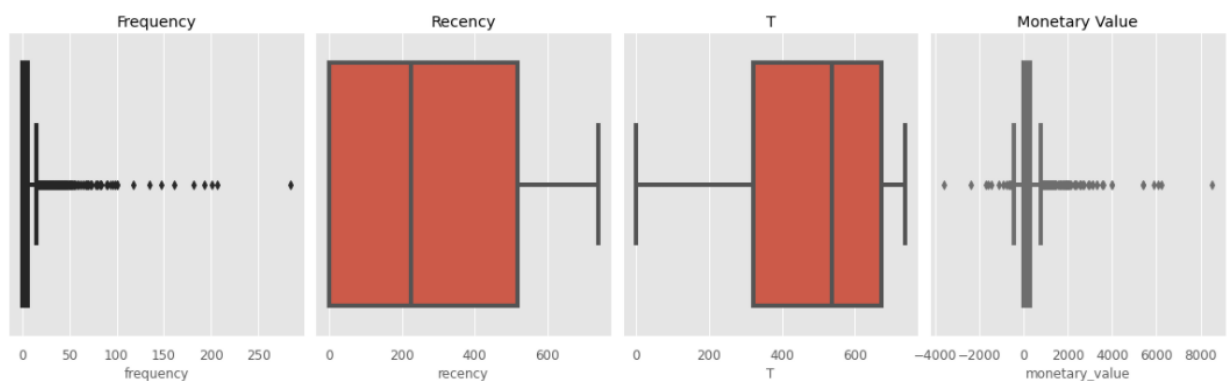


Рисунок 2.10 Блокова діаграма показників frequency, recency, T та monetary

Отже, знову переконуємося в наявності екстремальних значень у показників frequency та monetary.

Після перетворення вхідної таблиці транзакцій в таблицю з розрахованими RFM-показниками дані підготовлені до застосування прогнозних моделей Pareto/NBD та Гамма-Гамма.

Pareto/NBD моделювання майбутніх покупок клієнтів.

Для програмної реалізації Pareto/NBD моделі була використана вбудована функція lifetimes.ParetoNBDFitter() з бібліотеки lifetimes.

За допомогою функції lifetimes.ParetoNBDFitter() відбулося тренування моделі і підбір параметрів випадкового розподілу для опису процесу покупок, на основі моделювання якого отримали такі показники:

- p_alive (ймовірність активності клієнта за період);
- predicted_purchases (прогнозована кількість покупок клієнта).

Вхідними даними для моделі слугувала таблиця результатів RFM-аналізу (рис. 2.8).

Вихідні дані мають вигляд, представлений на рис. 2.11.

	Customer ID	frequency	recency	T	monetary_value	p_not_alive	p_alive	predicted_purchases
0	12346.0	10.0	400.0	725.0	-15.468000	0.819902	0.180098	0.073993
1	12347.0	7.0	402.0	404.0	717.398571	0.000634	0.999366	0.499740
2	12348.0	4.0	363.0	438.0	449.310000	0.034898	0.965102	0.277712
3	12349.0	4.0	717.0	735.0	1107.172500	0.003674	0.996326	0.180377
4	12350.0	0.0	0.0	310.0	0.000000	0.334801	0.665199	0.044307
...	...	...	...	...	...	...	...	...
5937	18283.0	18.0	655.0	658.0	146.405556	0.000657	0.999343	0.779789
5938	18284.0	1.0	2.0	431.0	-25.000000	0.730145	0.269855	0.029883
5939	18285.0	0.0	0.0	660.0	0.000000	0.523655	0.476345	0.016428
5940	18286.0	2.0	247.0	723.0	362.740000	0.417577	0.582423	0.062764
5941	18287.0	6.0	696.0	738.0	697.165000	0.009871	0.990129	0.252462

Рисунок 2.11 – Вихідна таблиця даних у результаті тренування моделі Pareto/NBD

Залежність очікуваної кількості майбутніх транзакцій та ймовірності,

що клієнт буде активним за період 30 днів візуалізовано на теплових картах (рис. 2.12 – 2.13).

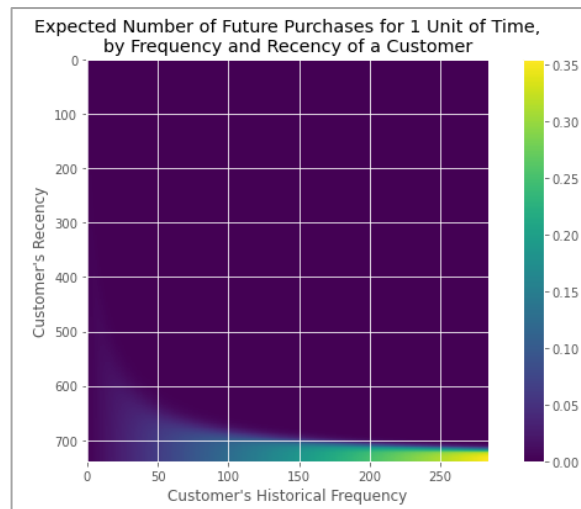


Рисунок 2.12 - Очікувана кількість майбутніх покупок залежно від Recency та Frequency

Отже, найбільша кількість прогнозних покупок за одиницю часу припадає на клієнтів з $Frequency > 250$, що співпадає з теоретичним очікуванням на основі здорового глузду. Адже, чим більше покупок здійснив клієнт за той же період часу, що і інші, то тим більше покупок він має здійснити у майбутньому, спираючись на його нормальну частоту здійснення транзакцій.

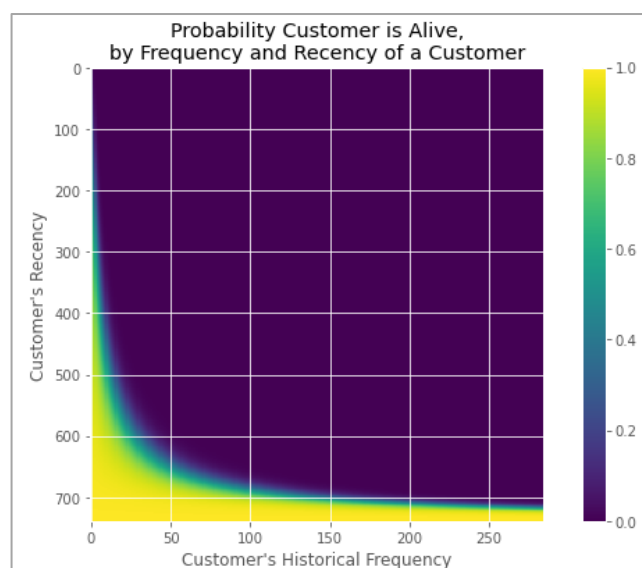


Рисунок 2.13 - Очікувана ймовірність, що клієнт буде активним

Можемо зробити висновок, що найвища ймовірність того, що клієнт активний спостерігається у тих, хто має показник Recency > 400.

Яскраво зображено принцип моделювання ймовірності повторної покупки на прикладі історії одного клієнта на рис. 2.14.

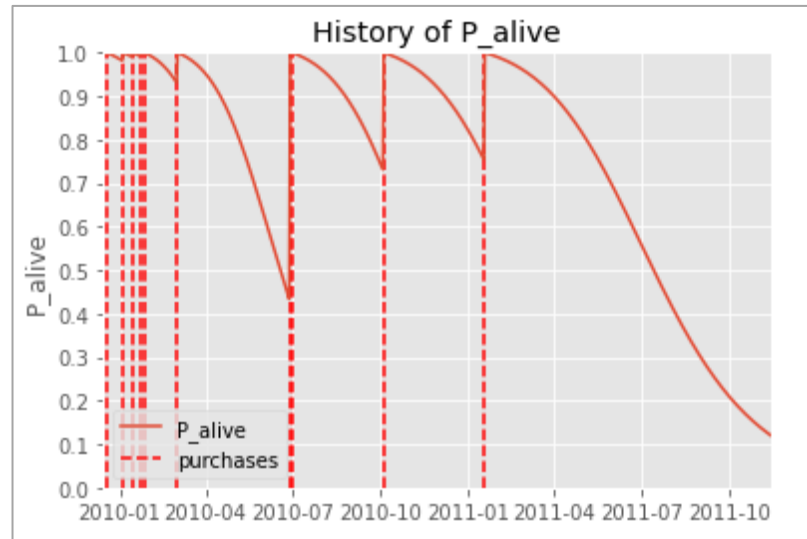


Рисунок 2.14 - Динаміка зміни ймовірності наступної покупки на прикладі одного клієнта (ID 12346.0)

У момент часу 0 ймовірність активності клієнта дорівнює 1. Потім вона спадає згідно з описаним вище розподілом до тих пір, поки клієнт не здійснить повторну покупку. Отже, клієнт з ID 12346.0 з ймовірністю 12% здійснить повторну покупку протягом наступного періоду, тобто з ймовірністю 88% можна вважати його відтоком.

Модель була відкорегована за допомогою розбивки всієї історії на тренувальний та тестовий період функцією `calibration_and_holdout_data()`.

Результати корегування і оцінка точності моделі зображена на рис. 2.15 та на рис. 2.16.

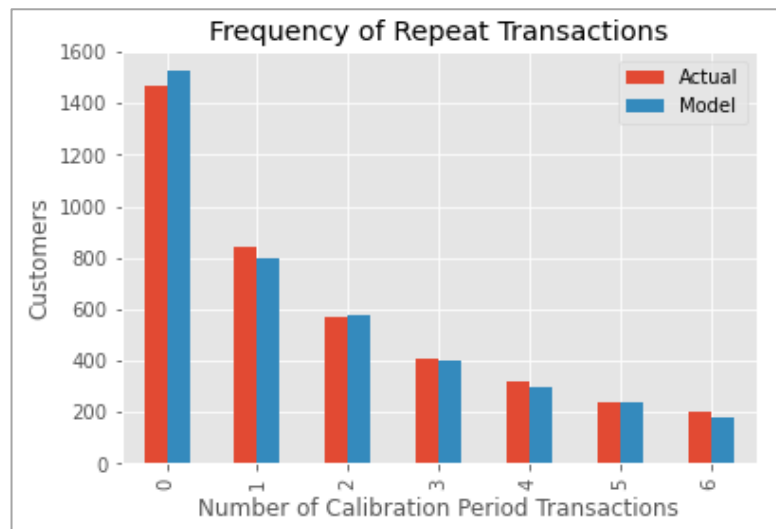


Рисунок 2.15 - Порівняння кількості прогностичних повторних транзакцій та фактичних

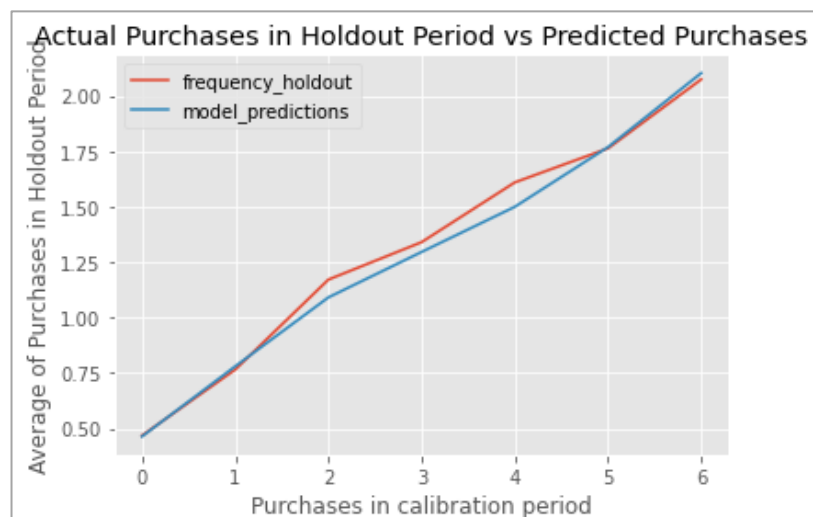


Рисунок 2.16 - Порівняння кількості транзакцій в тренувальному та тестовому періодах

Отже, моделювання купівельної поведінки клієнтів за допомогою стохастичної моделі Pareto/NBD є досить точним, але присутня незначна похибка в прогнозуванні від 2 до 5 покупок клієнтів, що яскраво виражено на рис. 2.16.

Гамма-Гамма моделювання майбутньої грошової цінності клієнтів

За допомогою вбудованої функції `lifetimes.GammaGammaFitter()` відбулося тренування моделі і підбір параметрів випадкового розподілу для опису процесу покупок, на основі моделювання якого отримали такі показники:

- `expected_avg_sales_` (очікувана середня вартість покупок клієнта),

- predicted_clv (прогнозне значення показника CLV клієнта).

Вхідними даними для моделі слугувала таблиця результатів RFM-аналізу (рис. 2.8).

Адекватність моделі була оцінена шляхом розрахунку таких основних видів помилки:

- Mean Squared Error (Середньоквадратична похибка) = 85,69;
- R2 Score (Коефіцієнт детермінації) = 0,99

Модель є достатньо точною для прогнозування майбутньої грошової цінності клієнта.

Кластеризація клієнтів за їх цінністю для компанії

У ході роботи була поставлена задача розділити на групи залежно від цінності для компанії.

Кластеризація здійснюватиметься методом к-середніх на основі попередньо розрахованих прогнозних значень клієнтів Pareto/NBD та Гамма-Гамма моделями.

Скористувавшись вбудованою функцією sklearn.Kmeans() здійснили сегментацію клієнтської бази залежно від показників:

- predicted_purchases – очікувана кількість покупок (Pareto/NBD),
- expected_avg_sales_ – очікувана середня вартість покупок (Гамма-Гамма модель)
- predicted_clv – очікувана позиттєва цінність клієнта (агрегований показник).

Вхідна таблиця даних має вигляд:

Таблиця 2.2 – Вхідний набір даних для кластеризації

	predicted_purchases	expected_avg_sales_	predicted_clv	profit_margin
0	0.499744	811.751032	8469.941195	423.497060
1	0.277711	564.130885	3170.514944	158.525747
2	0.180375	1389.675790	5400.501716	270.025086
3	0.568671	242.976996	2834.533397	141.726670
4	0.101370	479.625691	920.386917	46.019346
...	...	...	...	...
4181	0.080078	436.006243	687.574282	34.378714
4182	0.441472	65.122255	572.747558	28.637378
4183	0.779796	153.381233	2573.356789	128.667839
4184	0.062765	611.710079	563.528986	28.176449
4185	0.252461	806.536805	4347.701565	217.385078

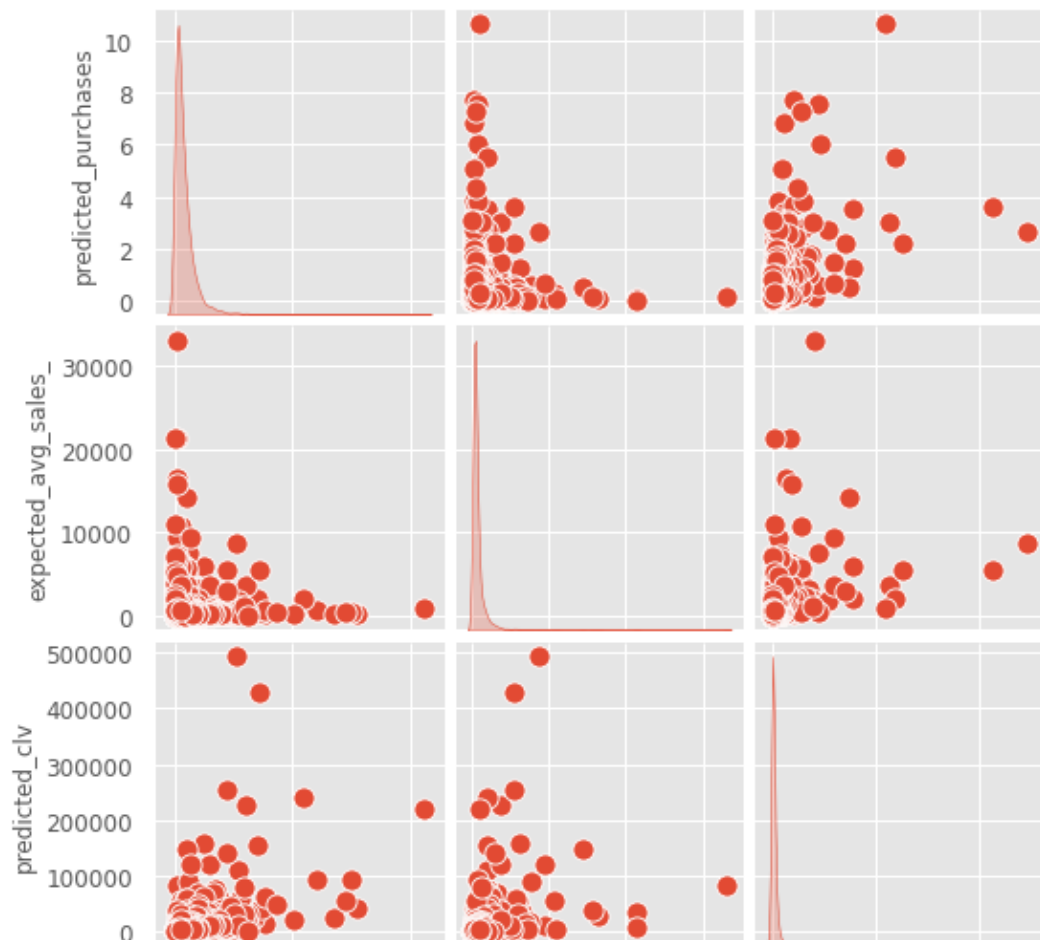


Рисунок 2.17 – Візуалізація залежності показників

На рис 2.17 зображений розподіл попарних залежностей між показниками, описаними вище.

Згідно підходу (1.7) була визначена оптимальна кількість кластерів та побудований графік суми відстаней між точками всередині кластерів (рис. 2.18).

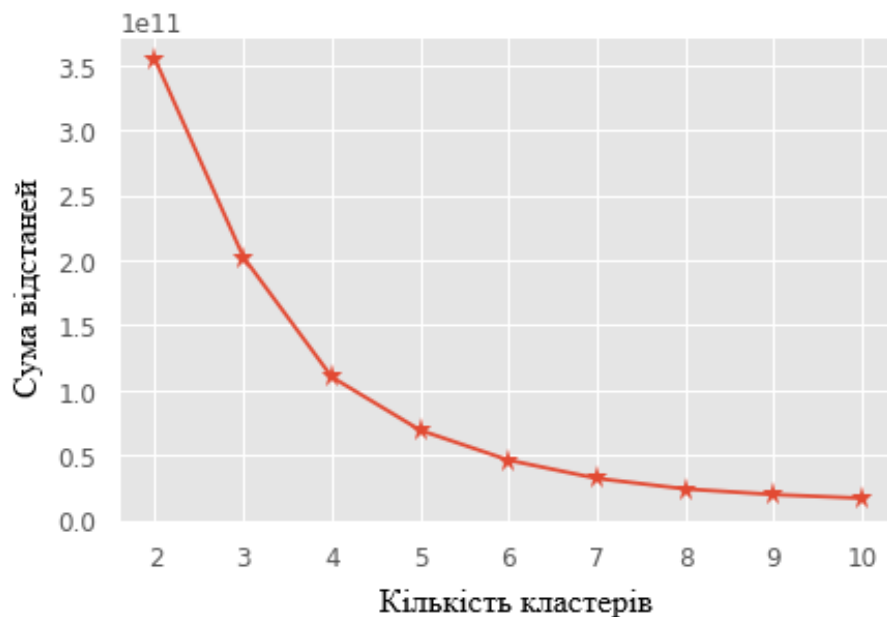


Рисунок 2.18 – Пошук оптимальної кількості кластерів

Отже, згідно з рис. 2.18 та експертною думкою було прийнято рішення обрати 4 кластери для групування всієї клієнтської бази.

У результаті отримали 4 групи:

- High (цінні),
- Low (малоцінні),
- Medium (помірно цінні),
- V_High (дуже цінні).

Опис середніх значень показників за групами зображено на рис. 2.19.

Таблиця 2.3 - Опис статистичного аналізу отриманих за кластерами даних

Labels	High	Low	Medium	V_High
Customer ID	15271.363636	15313.918428	14772.514563	16374.000000
frequency	70.181818	6.752826	37.349515	83.000000
recency	496.181818	386.509091	466.378641	737.000000
T	503.636364	526.279607	483.291262	737.500000
monetary_value	2965.939413	301.639238	1064.207404	6952.630519
p_not_alive	0.005292	0.159658	0.011983	0.000102
p_alive	0.994708	0.840342	0.988017	0.999898
predicted_purchases	3.062995	0.360167	1.637009	3.129489
actual30	6.203157	0.714849	2.063719	3.379286
error	3.140162	0.354682	0.426710	0.249796
expected_avg_sales_	4788.581355	584.471267	2688.549605	7025.639514
predicted_clv	171786.138264	3045.915288	35060.355304	460467.990272
profit_margin	8589.306913	152.295764	1753.017765	23023.399514

Кластеризація була здійснена рівномірно, групи не перетинаються та відповідають формалізованому опису їх цінності.

На рис. 2.20 продемонстровано розподіл клієнтів у кластерах залежно від їх прогнозованої цінності.

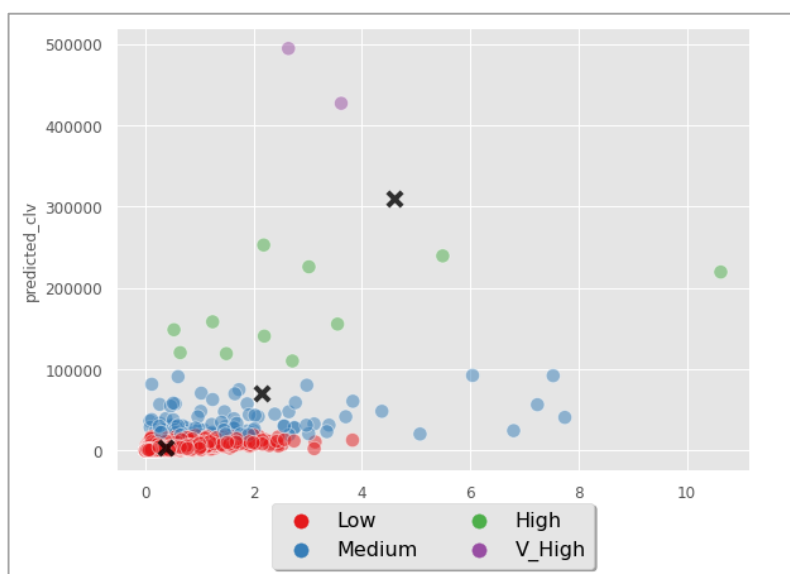


Рисунок 2.20 - Візуалізація розташування кластерів залежно від прогнозованої цінності

Розробка інструментарію управління клієнтською базою

Для зручного аналізу отриманих результатів і подальшого прийняття управлінських рішень було розроблено інформаційний динамічний дашборд засобами сервісу Google Data Studio (рис. 12). [Посилання на дашборд](#)

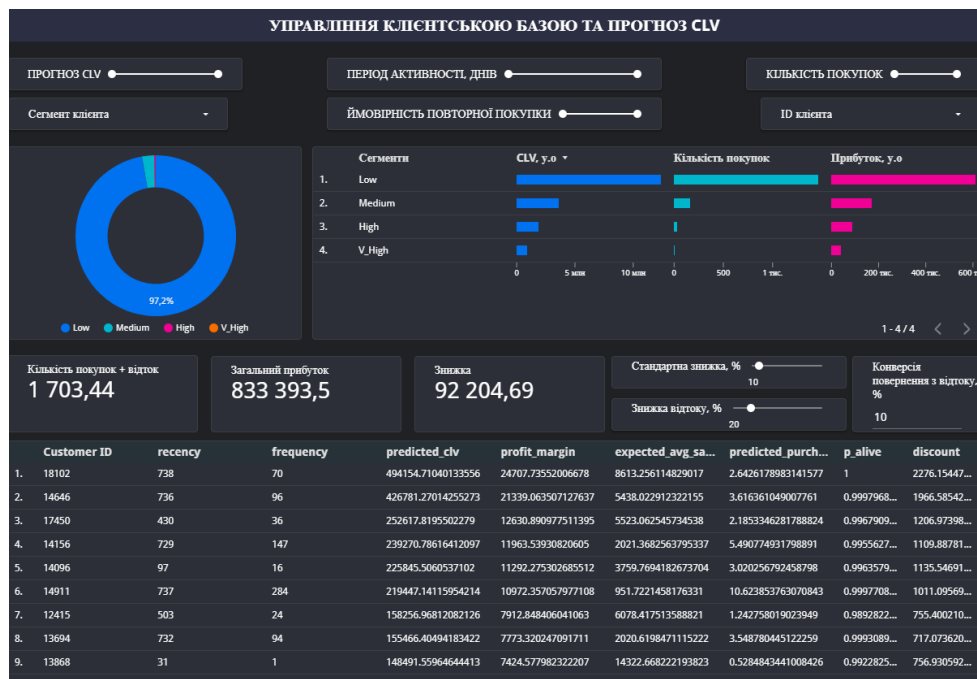


Рисунок 2.21 - Дашборд для управління клієнтською базою

Джерелом даних для дашборду виступає агрегована таблиця всіх результатів попереднього моделювання.

Основними можливостями дашборду є:

- Візуалізація груп клієнтів, отриманих в результаті кластеризації на основі їх прогнозованої цінності,
- Розрахунок прогнозованої кількості продажів та прибутку залежно від змінних параметрів.

Змінними параметрами дашборду є (для управлінської оцінки):

- Стандартна знижка – розмір % знижки для всіх клієнтів, крім відтоку;
- Підвищена знижка – розмір % знижки для клієнтів відтоку;
- Конверсія повернення з відтоку: % клієнтів, що отримали підвищену знижку і здійснили покупку (план).

ВИСНОВКИ

У ході дипломної роботи було досліджено теоретичні аспекти аналізу та управління клієнтською базою підприємства. Розроблено методичний підхід і інструментарій прогнозування клієнтської поведінки для максимально ефективного стимулювання попиту та клієнтоорієнтованості підприємства на основі підходу оцінки CLV клієнта.

За допомогою стохастичних Pareto/NBD та Gamma-Gamma моделей був здійснений прогноз майбутньої кількості та цінності покупок клієнтів компанії, був ідентифікований ймовірний відтік клієнтів.

Основним недоліком даних моделей є висока чутливість до вхідних даних. Розподіл всіх вхідних показників має відповідати теоретичним припущенням моделі, що не завжди так в реальності. А у випадку невідповідності значно зростає похибка моделі.

Використовуючи підхід к-середніх було здійснено кластеризацію клієнтів на групи за їх цінністю для компанії. Метод к-середніх повністю задовольняє виконання поставленого завдання та є простим у реалізації.

У результаті моделювання поведінки клієнтів компанії WOG було виявлено клієнтів з низькою ймовірністю повторної покупки протягом наступних 30 днів, ця група була ідентифікована як «відтік». Після чого було прийняте управлінське рішення запуснути цільові маркетингові заходи щодо повернення даної групи клієнтів, а саме їм було надано підвищену знижку на товари компанії. На основі A/B аналізу було оцінено чистий ефект такої взаємодії: 8% клієнтів відреагували і здійснили повторну покупку. Тим самим компанія збільшила додатково свій прибуток.

Для зручного аналізу і управління результатів був створений динамічний дашборд.

Модель може бути допрацьована і покращена відповідно до можливих нових майбутніх потреб компанії у сфері управління клієнтською базою.

ПЕРЕЛІК ПОСИЛАНЬ

1. Гончарук І.В. Роль фінансового планування у розвитку малого і середнього бізнесу / І.В. Гончарук // Ефективна економіка. –2017. –№9. [Електронний ресурс]. – Режим доступу: <http://www.economy.nayka.com.ua/?op=1&z=5770>.
2. Дивеєв-Кириленко О. Р. Проблеми розвитку малого і середнього бізнесу в Україні: монографія /Дивеєв-Кириленко О. Р.; М-во освіти і науки України, Терноп. нац. екон. ун-т.— 2-ге вид. випр. і доп. — Т. :ТНЕУ, 2013. — 255 с.
3. Batat W (2019) The new luxury experience. Springer, Berlin.
4. Терещенко О. О. Фінансова діяльність суб'єктів господарювання: навч. посіб. / О. О. Терещенко. - К.: КНЕУ, 2013 – 547 с.
5. Chen Y, Fay S, Wang Q (2011) The role of marketing in social media: how online consumer reviews evolve. J Interact Mark 25(2):85–94.
6. Goncharuk I.V. (2017), “Rol finansovogo planuvannya u rozvitku malogo i serednogo biznesu”, Effektivnaya ekonomika, [Online], vol.9, available at: <http://www.economy.nayka.com.ua/?op=1&z=5770> (Accessed 10 October 2018).
7. Dyveiev-Kyrylenko O. R. (2013), Problemy rozvytku maloho i seredn'oho biznesu v Ukraini: monohrafiia [Problems of Small and Medium Business Development in Ukraine: Monograph]; 2nd ed, Ternop. nats. ekon. unt, Ternopil, Ukraine.
8. Cvijikj IP, Michahelles F (2013) Online engagement factors on Facebook brand pages. Soc Netw Anal 3(4):843–861.
9. State Fiscal Service of Ukraine (2018), Tax Code of Ukraine, available at: <http://sfs.gov.ua/nk/> (Accessed 15 October 2018).
10. The Verkhovna Rada of Ukraine (2018), The Law of Ukraine “On Collection and Registration of the Single Contribution for Mandatory State Social Insurance”, available at: <http://zakon.rada.gov.ua/laws/show/2464-17> (Accessed 15 October 2018).
11. Faia Pinto A.R., Nagano M.S, 2019. An approach for the solution to order batching and sequencing in picking systems, Production Engineering-Research and Development, 13(3–4), 325–341, <http://doi.org/10.1007/s11740-019-00904-4>

12. Gower J.C, 1971. A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4), 857, <http://doi.org/10.2307/2528823>
13. Grosse E.H., Glock C.H, 2015. The effect of worker learning on manual order picking processes, *International Journal of Production Economics*, 170, 882–890, <http://doi.org/10.1016/j.ijpe.2014.12.018>.
14. Gu J., Goetschalckx M., McGinnis L.F, 2007. Research on warehouse operation: A comprehensive review, *European Journal of Operational Research*, 177(1), 1–21, <http://doi.org/10.1016/j.ejor.2006.02.025>.
15. Guo X., Yu Y., De Koster R.B.M, 2016. Impact of required storage space on storage policy performance in a unit load warehouse, *International Journal of Production Research*, 54(8), 2405–2418, <http://doi.org/10.1080/00207543.2015.1083>
16. Hong S, 2019. A performance evaluation of bucket brigade order picking systems: Analytical and simulation approaches, *Computers & Industrial Engineering*, 135, 120–131, <http://doi.org/10.1016/j.cie.2019.05.037>.
17. Jemelka M., Chramcov B., Kriz P., Bata T, 2017. ABC analyses with recursive method for warehouse, 2017 4th International Conference on Control, Decision and Information Technologies (CoDIT), 0960–0963, <http://doi.org/10.1109/CoDIT.2017.810272>
18. Kaufman L., Rousseeuw P, 2005. Introduction. In *Finding Groups in Data: An Introduction to Cluster Analysis*, (pp. 1–67), John Wiley, Hoboken
19. Bottani E., Volpi A., Montanari R., 2019. Design and optimization of order picking systems: An integrated procedure and two case studies, *Computers & Industrial Engineering*, 137, <http://doi.org/10.1016/j.cie.2019.106035>.
20. Çelik M., Süral H, 2019. Order picking in parallel-aisle warehouses with multiple blocks: complexity and a graph theorybased heuristic, *International Journal of Production Research*, 57(3), 888–906

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

Системний аналіз факторів, що впливають на лояльність клієнтів у роздрібній торгівлі

УДОДІЧ ОЛЬГА ІВАНІВНА

КЕРІВНИК РОБОТИ - СТ. ВИКЛАДАЧ КАФ. СА КУЗМІЧ МИХАЙЛО ЮРІЙОВИЧ

Об'єкт, предмет, мета роботи

Об'єктом роботи є розробка моделі для прогнозування ймовірності повторної покупки клієнта, кількості та цінності його майбутніх транзакцій задля ефективного управління.

Предметом роботи є клієнтська база підприємства та її структура залежно від цінності для компанії.

Мета роботи – це побудова та реалізація універсальної динамічної моделі оптимізації управління клієнтською базою компанії із врахуванням поведінкових характеристик та цінності груп клієнтів.

Лояльність споживачів є ключовим фактором успішності ринкової діяльності компанії, а управління лояльністю споживачів стає першочерговою стратегічною задачею менеджменту підприємства.

Типи лояльності

1. **Поведінкова лояльність** (лояльність без прихильності). Проявляється при покупці товару певного бренду на постійній основі, але без прихильності до нього. У такій ситуації покупець байдужий до даного бренду або не задоволений ним. Він здійснює покупку певного бренду з-за відсутності на ринку «улюбленої» марки або з економічних причин, тому при першій можливості може переключитися на товар конкурента. При цьому можуть бути випадки, коли у споживача взагалі немає брендів, до яких він має прихильність. Тож покупка товарів певної марки обумовлена раціональними мотивами.

2. **Сприймана лояльність** (прихильність без лояльності). Це лояльність, пов'язана з відношенням. Вона проявляється як зацікавленість покупця у придбанні товару конкретного бренду. При цьому покупець повністю задоволений товаром даної марки, високо оцінює бренд, має до нього емоційну прихильність, тому буде придбати його протягом необмеженого проміжку часу (за умов наявності даного бренду на ринку та за сприятливих економічних умов).

3. **Змішана лояльність** (поєднання прихильності та лояльності). Покупець придбає бренд постійно, отримуючи при цьому як емоційну прихильність, так і глибоку задоволеність.

Схема існуючих напрямків дослідження задачі управління клієнтською базою



Підходи до управління клієнтською базою компанії на основі показника довгострокової вартості клієнта

Для подолання обмежень моделей управління клієнтською базою компанії, заснованих на показнику лояльності клієнтів, розглянутих вище, використовується показник довгострокової вартості клієнтської бази (CLV – customer lifetime value).

$$CLV = \sum_{i=1}^t \frac{D_i - Z_i}{(1 + d)^{i-1}} \quad (1.1)$$

де i – номер періоду, в який здійснюється розрахунок надходжень від клієнтів;

D_i – дохід від клієнта в період i ;

Z_i – загальні витрати на отримання доходу D_i в період i ;

t – загальна кількість періодів протягом життєвого циклу клієнта;

d – ставка дисконтування.

Порівняльна характеристика двох підходів до управління базою клієнтів

Характеристика	Показник лояльності	Показник CLV
Ефективність використання	Середня (не відображає витрати на управління клієнтською базою)	Висока (не відображає якісну сторону управління клієнтською базою)
Складність реалізації	Низька	Висока (потребує прогнозування грошового потоку клієнтів)
Практика застосування	Часто	Рідко

Порівняльна характеристика підходів до управління клієнтською базою на основі показника CLV

Параметр порівняння	Регресійна модель	Парето/NBD модель
Складність розрахунку	Середня	Висока
Точність прогнозування	Середня	Висока
Чутливість до даних	Висока	Висока
Гнучкість	Середня	Низька

За параметром складності розрахунку менших обчислювальних витрат потребує регресійна модель, ніж Парето/NBD-модель.

При цьому найбільшою точністю прогнозування володіють Парето/NBD-модель, і менш точні показники на довгостроковому інтервалі часу у регресійних моделях.

Низькі показники точності регресійних моделей пояснюються також високою чутливістю до вихідних даних. Парето/NBD-модель також вимоглива до якості вхідної інформації.

За параметром гнучкості і часу адаптації для використання в компанії більшу перевагу отримують регресійні моделі, а імовірнісні Парето/NBD-модель є менш гнучкими і більш вимогливими до впровадження.

Зважаючи на всі переваги та недоліки, для реалізації поставленої у роботі задачі була обрана стохастична Парето/NBD-модель та підхід управління клієнтською базою компанії на основі показника CLV.

Постановка задачі управління клієнтською базою компанії

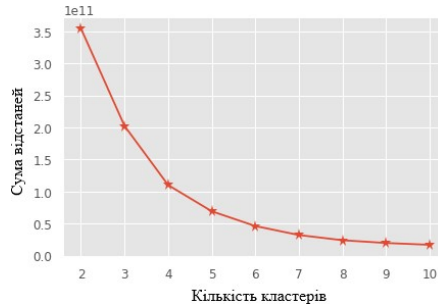
Компанії потребують комплексної динамічної моделі управління групами клієнтів.

Дана модель повинна дозволяти:

- Прогнозувати майбутнє значення CLV для клієнта;
- Ідентифікувати найлояльніших клієнтів з максимальним CLV;
- Ідентифікувати відтік шляхом розрахунку прогнозованої ймовірності здійснення повторної покупки в компанії;
- Здійснювати кластеризацію клієнтів за їх цінністю для компанії з метою заохочення і утримання найцінніших та повернення відтоку за допомогою цільових персоналізованих маркетингових заходів;
- Розрахунок прогнозованого прибутку компанії за період.

Кластеризація клієнтів за їх цінністю

У ході роботи була поставлена задача розділити на групи залежно від цінності для компанії. Кластеризація здійснювалася методом k -середніх на основі попередньо розрахованих прогнозних значень. Визначено оптимальну кількість кластерів та побудовано графік суми відстаней між точками всередині кластерів.



Згідно з рис. та експертною думкою було прийнято рішення обрати 4 кластери для групування всієї клієнтської бази.

У результаті отримали 4 групи

- High (цінні),
- Low (малоцінні),
- Medium (помірно цінні),
- V_High (дуже цінні).

Висновки

У ході роботи було досліджено теоретичні аспекти аналізу та управління клієнтською базою підприємства. Запропоновано методичний підхід і інструментарій прогнозування клієнтської поведінки для максимально ефективного стимулювання попиту та клієнтоорієнтованості підприємства на основі підходу оцінки CLV клієнта.

За допомогою стохастичних Pareto/NBD та Gamma-Gamma моделей був здійснений прогноз майбутньої кількості та цінності покупок клієнтів компанії, був ідентифікований ймовірний відтік клієнтів.

Основним недоліком даних моделей є висока чутливість до вхідних даних. Розподіл всіх вхідних показників має відповідати теоретичним припущенням моделі, що не завжди так в реальності. А у випадку невідповідності значно зростає похибка моделі.

Використовуючи підхід k -середніх було здійснено кластеризацію клієнтів на групи за їх цінністю для компанії. Метод k -середніх повністю задовольняє виконання поставленого завдання та є простим у реалізації.