

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

КВАЛІФІКАЦІЙНА РОБОТА

на тему: **«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ СПОЖИВЧОГО ПОПИТУ
ДЛЯ ЗАБЕЗПЕЧЕННЯ МАКСИМАЛЬНОГО ПРИБУТКУ
ПІДПРИЄМСТВА»**

на здобуття освітнього ступеня магістра
зі спеціальності 124 Системний аналіз
освітньо-професійної програми Інтелектуальні системи управління

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

_____ Анастасія КОСТРУБІЦЬКА
(підпис) Ім'я, ПРІЗВИЩЕ здобувача

Виконав: здобувачка вищої освіт
групи САДМ-61
_____ Анастасія КОСТРУБІЦЬКА

Керівник: _____ Сергій КОЗАЧЕКО

Рецензент:

Київ 2024

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ**

Кафедра Системного аналізу

Ступінь вищої освіти Магістр

Спеціальність 124 Системний аналіз

Освітньо-професійна програма Інтелектуальні системи управління

ЗАТВЕРДЖУЮ

Завідувач кафедри СА

_____ Ровіл НАФССВ

«_____» _____ 20__р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

Кострубіцька Анастасія В'ячеславівна

1. Тема кваліфікаційної роботи: Інтелектуальний аналіз споживчого попиту для забезпечення максимального прибутку підприємства

керівник кваліфікаційної роботи Сергій КОЗАЧЕНКО к.е.н. ,

затверджені наказом Державного університету інформаційно-комунікаційних технологій від

«_____» _____ 20__р. № _____

2. Строк подання кваліфікаційної роботи

«15» січня 2024р.

3. Вихідні дані до кваліфікаційної роботи:

Споживчий попит та його еластичність

Методи Data Mining

Сегментація клієнтської бази

Науково-технічна література

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

1. Дослідити методи та засоби інтелектуального аналізу даних.

2. Здійснити інтелектуальний аналіз споживчого попиту підприємства

5. Перелік ілюстративного матеріалу: *презентація*

6. Дата видачі завдання «13» жовтня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Підбір науково-технічної літератури	01.11.2023	Викон.
2	Формування та аналіз споживчого попиту	25.11.2023	Викон.
3	Методи Інтелектуального аналізу даних	12.12.2023	Викон.
4	Інтелектуальний аналіз споживчого попиту підприємства	27.12.2023	Викон.
5	Висновки, вступ, реферат	02.01.2024	Викон.
6	Розробка презентації	05.01.2024	Викон.
7	Дата здачі дипломної роботи	15.01.2024	Викон.

Здобувач(ка) вищої освіти

(підпис)

Анастасія КОСТРУБИЦЬКА

(Ім'я, ПРІЗВИЩЕ)

Керівник
кваліфікаційної роботи

(підпис)

Сергій КОЗАЧЕКО

(Ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Текстова частина дипломної роботи: 76 с., 11 рис., 15 джерел

Актуальність роботи. Хто володіє інформацією, той володіє світом. А той, хто розуміє покупців та вміє працювати з попитом, отримує високий прибуток. Тому аналіз споживчого попиту є важливою складовою будь-якого бізнесу. Підприємець, який не приділяє йому достатньої уваги, згодом ризикує не просто зазнати збитків через товари, що лежать на складі, а й виявитися вимушеним закрити своє підприємство.

Мета і завдання дослідження. Метою випускної кваліфікаційної роботи є інтелектуальний аналіз споживчого попиту для забезпечення максимального прибутку підприємства.

Для досягнення поставленої мети необхідно виконати такі завдання:

- Розглянути поняття споживчого попиту, чинники, що впливають на споживчий попит та поведінку споживача, еластичність споживчого попиту шляхом його моделювання.
- Дослідити методи та засоби інтелектуального аналізу даних, щоб обрати відповідну модель інтелектуальних обчислень.
- Здійснити інтелектуальний аналіз споживчого попиту підприємства.

Об'єктом дослідження є методи та засоби інтелектуального аналізу даних.

Предметом дослідження є ТОВ «Здорові продукти»

Методи дослідження. В роботі виконуються дослідження, які базуються на використанні статистичних методів аналізу, маркетингових досліджень, методи узагальнення, системний аналіз.

Наукова новизна одержаних результатів полягає у доведенні необхідності застосування інтелектуального аналізу споживчого попиту для формування економічних стратегій підприємств.

Практичне значення одержаних результатів полягає у можливості використання результатів дослідження при обґрунтуванні цінових концепцій підприємств.

Ключові слова: СПОЖИВЧИЙ ПОПИТ, ЕЛАСТИЧНІСТЬ СПОЖИВЧОГО ПОПИТУ, ПОВЕДІНКА СПОЖИВАЧАБ МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.

ABSTRACT

The text part of the thesis: 76 pp., 11 figures, 15 sources

Relevance of work. He who owns information owns the world. And the one who understands buyers and knows how to work with demand gets a high profit. Therefore, the analysis of consumer demand is an important component of any business. An entrepreneur who does not pay enough attention to it, later runs the risk not only of suffering losses due to the goods lying in the warehouse, but also of being forced to close his enterprise.

The purpose and tasks of the work. The goal of the final qualification work is an intellectual analysis of consumer demand to ensure maximum profit of the enterprise.

To achieve the goal, the following tasks must be completed:

- Consider the concept of consumer demand, factors influencing consumer demand and consumer behavior, elasticity of consumer demand by modeling it.
- Investigate the methods and means of intelligent data analysis in order to choose the appropriate model of intelligent computing.
- To carry out an intellectual analysis of the enterprise's consumer demand.

The object of research is the methods and means of intellectual data analysis.

The subject of the study is «Healthy Products»

Research methods. The work includes research based on the use of statistical analysis methods, marketing research, generalization methods, and system analysis.

The scientific novelty of the obtained results lies in the proof of the need to apply intellectual analysis of consumer demand for the formation of economic strategies of enterprises.

The practical significance of the obtained results lies in the possibility of using the research results in justifying the price concepts of enterprises.

Keywords: CONSUMER DEMAND, ELASTICITY OF CONSUMER DEMAND, CONSUMER BEHAVIOR, METHODS OF DATA MINING.

ЗМІСТ

ВСТУП.....	9
РОЗДІЛ ПЕРШИЙ.....	11
ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ АСПЕКТИ ФОРМУВАННЯ ТА АНАЛІЗ СПОЖИВЧОГО ПОПИТУ	11
1.1. Поняття «Споживчий попит» та його класифікація	11
1.1. Моделі поведінки споживачів	12
1.2.1. Чинники, які впливають на споживчий попит	18
1.3. Важливість проведення аналізу споживчого попиту.....	20
1.4. Моделювання споживчого попиту, як спосіб вивчення його еластичності ..	22
1.5. Співвідношення сумарних витрат та доходу, як фактор попиту.....	31
ВИСНОВКИ ДО ПЕРШОГО РОЗДІЛУ	34
РОЗДІЛ ДРУГИЙ	35
ТЕОРІЯ ТА МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ	35
2.1. Поняття «Data Mining»	35
2.2. Застосування інтелектуального аналізу даних.....	36
2.2. Інтелектуальний аналіз даних в бізнесі та торгівлі.....	39
2.3. Методи та засоби інтелектуального аналізу даних	40
2.3.1. Етапи проведення інтелектуального аналізу даних.....	43
2.3.2. Попередня обробка даних	45
2.6. Моделі інтелектуального аналізу	50
2.6.1. Архітектура моделі інтелектуального аналізу даних	51
2.6.2. Вибір правильного алгоритму	53
2.6.3. Стовпці моделі інтелектуального аналізу даних	54
2.6.4. Обробка моделей інтелектуального аналізу даних.....	56
2.6.5. Властивості моделі інтелектуального аналізу даних	56
ВИСНОВКИ ДО ДРУГОГО РОЗДІЛУ	58
ТРЕТІЙ РОЗДІЛ	59
ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ СПОЖИВЧОГО ПОПИТУ НА ПРИКЛАДІ ПРИВАТНОГО ПІДПРИЄМСТВА «ЗДОРОВИЙ ПРОДУКТ».....	59
3.1. Аналіз та сегментування клієнтської бази	59
3.1. Методологія дослідження.....	60
ВИСНОВКИ ДО ТРЕТЬОГО РОЗДІЛУ	66

ВИСНОВКИ.....	67
ПЕРЕЛІК ПОСИЛАНЬ	69
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ	71

ВСТУП

Ми живемо у світі даних. Обсяг даних, які створюються, використовуються та зберігаються, зростає з кожним днем. Велика кількість компаній мають дуже багато даних, але використати їх в повному обсязі не можуть, бо бракує знань.

Нині ті, хто не вміє отримувати з океану зібраної інформації цінні практичні знання, програють у конкурентній боротьбі. Вже неважливо, чим займається компанія, data mining зможе вивести її прибуток на новий рівень.

Наступні десятиліття важливість аналітики лише зростатиме. Інструменти і далі будуть розвиватися, щоб розкрити ще більший потенціал інформації. Це означає, що у data mining велике майбутнє.

Інтелектуальний аналіз даних – це процес обробки великих наборів даних для виявлення відповідної інформації. Підприємства використовують інтелектуальний аналіз даних для бізнес-аналітики та для виявлення конкретних даних, які можуть допомогти компаніям приймати більш ефективні управлінські рішення.

Концепція data mining походить ще до 1700-х років. Теорема Байєса та регресійний аналіз – ранні приклади того, як люди виявляли закономірності в морі даних.

Термін «data mining» вперше з'явився в 1983 році, а до 90-х років, з народженням інтернету, він став по-справжньому популярним.

1995 року мережа гіпермаркетів Tesco передала багато своїх даних геніальному стартапу – Dunnhumby. Ця аналітична компанія вміла перетворювати необроблені дані спочатку в ідеї, а потім на корисні для бізнесу дії, використовуючи лише аналіз даних.

Dunnhumby розписали для Tesco цікаві ідеї для впровадження у бізнес, після чого голова ради директорів Tesco заявив, що Dunnhumby дізналися про бізнес за три місяці більше, ніж він за 30 років. Це був лише початок.

В даний час компанії збирають набагато більше даних, ніж можуть переробити: 90% даних було створено за останні 2-3 роки.

Зараз щоразу, коли хтось натискає лайк у Фейсбуці, кладе товар у кошик, проводить кредитною карткою в магазині, заходить на сайт або просто спить зі смарт-годинником на руці, якась компанія отримує дані.

Ринок систем Data Mining зростає. Цьому сприяє діяльність великих корпорацій: SAS, IBM, Microsoft, Oracle та інших. Очікується, що до 2027 року обсяг глобального ринку розширеної аналітики зросте на 23,1% і досягне позначки \$56,2 млрд.

Останні тенденції в Data Mining включають розвиток методів аналізу з елементами віртуальної та доповненої реальності, їх інтеграцію із системами баз даних, видобуток біологічних даних для інновацій у медицині, веб-майнінг (аналіз даних в інтернеті), аналіз даних у реальному часі, а також заходи захисту конфіденційності при видобуванні даних. Лідери галузі вважають, що в майбутньому майнінг даних застосовуватиметься в інтелектуальних додатках, які будуть вбудовані у корпоративні сховища даних.

Головною проблемою виявлення закономірностей даних є час, який потрібно для перебору інформаційних масивів. Відомі методи або штучно обмежують такий перебір, або будують цілі дерева рішень, які знижують ефективність пошуку. Вирішення цієї проблеми залишається головною метою розробників продуктів для Data Mining.

РОЗДІЛ ПЕРШИЙ

ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ АСПЕКТИ ФОРМУВАННЯ ТА АНАЛІЗ СПОЖИВЧОГО ПОПИТУ

1.1. Поняття «Споживчий попит» та його класифікація

Попит (англ. demand) – кількість товару, яку споживачі бажають і можуть придбати за додаткові ціни протягом певного періоду часу [3]. Споживчим він називається тому, що товар купується покупцем для особистого використання, а не для подальшого перепродажу.

Попитом не можна назвати ситуацію, якщо покупець має бажання мати цю річ, але немає реальних фінансових можливостей її купити. Аналогічно і в тому випадку, якщо у споживача є гроші на покупку, але немає бажання купувати її.

Споживчий попит буває індивідуальним чи ринковим [8]:

- у першому випадку вивчають, наскільки буде цікавий конкретний товар певному покупцю за різних цін;
- у другому – попит всіх споживачів на цей же товар за різних цін.

Наприклад, підприємець Михайло займається продажем футболок із зображенням модних персонажів інтернет-мемов. Якість його товару висока, тому Михайло продає їх за ціною 1200 грн за одну штуку. Але за цією ціною футболку зможе купити лише одна людина.

Якщо б ціна була 800 грн, то футболку вже куплять дві людини, до яких додасться той перший, що готовий був здійснити покупку і за вищу ціну. Якщо ціна буде 600 грн/штуку – три людини, а також ті троє, що готові були купувати за вищою ціною, тобто сумарно Михайло вже матиме 6 клієнтів.

Тобто, величина попиту – це кількість товару, яка користуватиметься увагою покупців. Основний фактор, який впливає на неї, – ціна. Між цими двома величинами є зворотний зв'язок: чим вища вартість товару, тим менший попит вона матиме.

1.1. Моделі поведінки споживачів

Моделі поведінки споживачів є спрощеним уявленням їхньої реальної поведінки, створені для того, щоб показати відносини між різними елементами системи або процесу після попереднього розгляду цієї проблеми. Відносини охоплюють систематичний вибір рис, які проявляються у взаємозалежностях та теоретично обґрунтовують досягнення результату цієї дії.

Існує велика кількість різних моделей, які з'ясовують поведінку споживачів на ринку.

Залежно від прийнятого критерію можна виділити такі моделі [1]:

- прості та складні;
- на макро– та мікрорівні;
- теоретичні та емпіричні;
- описові (історичні або поточні), нормативні та прогностичні;
- динамічні та дихотомічні;
- функціональні та інтелектуальні;
- біхевіоральні та статистичні;
- кількісні та якісні;
- математичні та типологічні;
- загальні та ad hoc.

Враховуючи необхідність комплексного аналізу моделі поведінки споживачів на ринку, можна їх класифікувати за рівнем складності, виділивши прості та складні моделі.

Прості моделі характеризують та з'ясовують поведінку споживачів загалом. У протиставленні складним моделям представляють лише основні чинники, що впливають на поведінку споживачів. Серед них можна виділити чотири види моделі [1,8]:

1. «чорна скринька» – це збірна назва моделі, яка не містить аналізу щодо внутрішніх процесів прийняття рішень споживачами, модель є результатом

безпосереднього спостереження факторів, що впливають на споживача, та реакцій, які вони викликають;

2. процес прийняття рішень – моделі знаходять широке застосування у маркетинговій діяльності фірм, ілюструючи різні рівні (етапи) прийняття рішень активним споживачем, наприклад модель AIDA;

3. «особистих змінних» – ці моделі, на противагу моделі «чорна скринька», концентруються на внутрішніх чинниках, що формують поведінки споживачів, вони повинні з'ясувати конкретні внутрішні процеси прийняття рішень, наприклад модель поведінки Fishbeina;

4. гібридні – моделі цього типу повинні поєднати модель процесу прийняття рішення та модель особистих змінних, наприклад модель Rice'a.

Найпростішою моделлю поведінки споживачів є модель «стимул-реакція», що зазвичай використовується як зображення експерименту Павлова. На ринку таку модель можна спостерігати, коли, наприклад, споживач бачить на вітрині магазину інформацію про зниження цін на товари, заходить у магазин і робить покупку.

Більш складною моделлю типу «чорного ящика», що містить більше змінних, які впливають на поведінку споживача, є модель «середовища споживчої поведінки». Модель є стимулом, який може впливати на купівельне рішення споживача. У центрі такої моделі є сам споживач. Слід підкреслити, деякі зі стимулів (змінних) можуть лише впливати на поведінку споживачів, але й можуть перебувати під впливом споживача та її поведінки. І тут має місце явище двостороннього впливу елементів цієї моделі.

Наступною моделлю типу «чорний ящик» є модель «процесу покупки» Ф. Котлера. Ця модель ідентифікує фактори на вході та на виході цього процесу. Вона показує складнішу модель поведінки споживача над ринком проти попередніх. Більше того, відображає ширшу гаму факторів, що стимулюють купівлю та результати рішень, прийнятих споживачами.

Модель «ухвалення рішень» представляє окремі етапи процесу прийняття рішень споживачем. Так як моделі типу «чорна скринька» концентруються на

поданні дії зовнішніх факторів, вони можуть бути корисні лише тоді, коли внутрішні фактори є незначними або відіграють незначну роль у формуванні споживчої поведінки. Крім того, необхідно взяти до уваги значення використаних у цих моделях зовнішніх факторів.

Представлені прості моделі поведінки споживачів зосереджені на ідентифікації стимулів, які впливають на споживачів, а ще на визначенні результатів дії цих стимулів (чинників), чи ідентифікації певного поведінки споживачів над ринком.

З метою пояснення різноманітної поведінки споживачів над ринком слід використовувати складні моделі.

До цієї групи належать моделі:

- Структурні – це моделі, що відображають цілісний процес прийняття рішень про купівлю товару. Вони представляють істотні чинники та їх взаємовплив, завдяки яким з'ясовуються психічні процеси, що пояснюють суть споживчої поведінки [2];
- стохастичні, які здійснюють спробу передбачити поведінку споживача на ринку за допомогою можливого розрахунку;
- поведінка споживачів, яка об'єднана в процес, що складається з вхідних та вихідних змінних.

Ці моделі, однак, не відображають фактори, що знаходяться усередині; ці чинники неможливо простежити, оскільки вони утворюються усередині організму.

До моделей, що найчастіше зустрічаються та які описують поведінку споживача на ринку, відносяться

- стохастичні (імовірнісні) моделі, такі, як: моделі Говарда-Шета (Howarda-Shetha), модель Нікосії (Nicosii), а також моделі Енгеля (Engela), Коллата (Kollata), Блеквелла (Blackwella).

Модель Говарда-Шета (Howard-Sheth's madel). Найбільш ранньою та відомою моделлю, що представляє поведінку споживачів на ринку, є модель Говарда-Шета (Howarda-Shetha) [4]. Вона здійснює спробу відображення

цілісного процесу прийняття рішення споживачем над ринком. Ця модель виділяє три рівні знань споживача, що визначають прийняття рішень споживачем [5]:

- широке поле діяльності має призвести до вирішення проблеми, коли знання споживача про даний продукт або марку дуже обмежені або відсутні; тоді у споживача немає окремих переваг марки продукту, і він активно шукає інформацію, пов'язану з альтернативними марками продуктів;
- обмежене поле діяльності, пов'язане з вирішенням проблеми, коли знання споживача та його переваги марки частково визначені. Споживач не повністю може побачити різницю між окремими марками, завдяки яким встановлюються переваги. Споживач шукає більше інформації, яка здатна встановити критерії прийняття рішення;
- рутинна поведінка, коли знання споживача, його переваги марки, а також доступні на ринку варіанти вибору йому добре знайомі; споживач, переконаний у покупці однієї конкретної марки.

У моделі Говарда-Шета (Howarda-Shetha) виділяється чотири основні групи змінних: вхідні, психологічні (наприклад, усвідомлення, процес навчання), вихідні та зовнішні. Центральну частину моделі Говарда-Шета (Howarda-Shetha) становлять психологічні змінні, які зумовлюють дії споживача під час ухвалення рішення. Однак слід підкреслити, що ця конструкція сприймається як абстрактна, невизначена і безпосередньо не вимірювана. Деякі змінні сприймаються в натуральній формі і є одночасно пов'язаними з процесом отримання та присвоєння інформації, що встановлює вхідні стимули. Прикладом може бути ситуація, коли цей стимул є двозначно зрозумілим. Це відбувається тоді, коли інформація, отримана споживачем із оточення, неясна. Більше того, споживач може спотворити отриману інформацію, щоб пристосувати її до певних власних потреб і досвіду.

Крім функції сприйняття стимулів із довкілля психологічні змінні виконують так само функцію навчання. Це концептуальна функція. У ній враховуються цілі споживача, інформація про продукти, критерії оцінки альтернатив, переваги та наміри щодо покупки. Численні інтеракції і зв'язки між

змінними, що виконують функції сприйняття і навчання споживача, надають особливого характеру моделі Говарда-Шета (Howard-Shetha). Модель також представляє серію вихідних змінних, які відносяться до видимих ефектів внутрішніх процесів, наприклад, рішення про здійснення покупки або відмову від неї, уточнення інших намірів, виявлення власних характеристик та інтересів.

Істотно, хоч і безпосередню, частину процесу прийняття рішень становлять зовнішні змінні, тобто, значення купівлі, особисті чинники, фінансовий статус споживача, а також вплив часу.

Модель Енгеля (J.F. Engel), Коллат (D.T. Kollat), Блеквелла (R.D. Blackwell) була спроектована як цілісна модель з повним розв'язанням проблеми і відображає процес прийняття рішення споживача про нову покупку [4]. Як і попередня модель, вона представляє різні змінні, які впливають на процес прийняття рішень споживачем, одночасно пробуючи з'ясувати початкові відносини, встановлювані між окремими компонентами і субкомпонентами цього процесу. Модель складається з чотирьох частин: процесу прийняття рішень, інформації на вході, процесу переробки інформації та змінних (індивідуальних та зовнішніх), що впливають на процес прийняття рішення. Ця модель враховує процес перетворення інформації споживачем (не розглядається в інших моделях), найкраще представляє специфіку процесів, що відбуваються в пам'яті, і розглядає однаково як позитивні, так і негативні результати покупок.

Модель Нікосії (Nicosia) – концентрується на відносинах, що встановлюються між фірмою та її потенційними клієнтами [5]. У широкому значенні фірма спілкується зі споживачами у вигляді виробничого процесу, споживачі ж спілкуються з фірмою у вигляді покупок. Модель Нікосії (Nicosia) представляє, інтерактивні зв'язки, що встановлюються між двома головними учасниками ринкової гри: підприємством та споживачем. Підприємство намагається управляти певною поведінкою споживача над ринком, а споживач – з допомогою своєї дії (чи відмови від дії) – впливає фірму.

Модель Нікосії знаходить особливе застосування в ідентифікації та оцінці реакції на певні стимули. Ця модель охоплює поведінку споживачів більшою

мірою як процес, що робить його більш правильним та корисним у дослідженні поведінки споживачів. Представлені моделі концентруються на процесі прийняття рішення особистістю, саме на тому, як індивідуальний споживач приймає рішення про вибір даного продукту чи послуги.

Модель Шета (J.N. Shetha) – «ухвалення рішення у сім'ї» – з'ясовує процеси, які у сім'ї, та його впливом прийняття рішення про купівлю. Ця модель об'єднує елементи прийняття рішень про купівлю індивідуальними споживачами (коли рішення про купівлю приймає один член сім'ї), і взаємодія між членами сім'ї (коли рішення про купівлю приймають кілька чи всі члени сім'ї) [5].

Модель відносин «середовище – домашнє господарство» показує відносини домашнього господарства (модифікації його ринкової поведінки) зміни у його ринковому середовищі. Ідентифікація та сприйняття змін у середовищі, а також оцінка тих з них, які впливають на поточний та майбутній стан домашнього господарства, є основою процесу реагування на зміни. Ця модель показує, як члени домашнього господарства оцінюють зміни, визначають їх напрями та здійснюють вибір можливих реакцій на них.

Щоб повною мірою здійснити аналіз поведінки споживачів на ринку, необхідно розглядати її крізь призму моделі реакції «середовище – домашнє господарство» щодо змін у середовищі. Модель виявляє як пасивну поведінку, засновану на регулюванні потреб до рівня наявних можливостей та забезпечення задоволення первинних потреб, так і активну внутрішню та зовнішню поведінку домашніх господарств (у сфері споживання та сфері ринку).

Модель поведінки споживача Маркова належить до стохастичних моделей [5]. Вона стверджує, що поведінка споживачів на ринку – це безперервний процес прийняття рішень, в якому виділяються певні стани в послідовних одиницях часу, і в якому досягнення будь-якого стану в певному відрізку часу залежить від досягнутого певного стану в попередньому відрізку часу $(t - 1)$. Модель Маркова визначає умовну можливість досягнення окремих станів. Якщо число ймовірності вибору становить n , умовна ймовірність, що процес у періоді t досягне певного способу вибір, утворюючи матрицю з вимірами $n \times n$.

1.2.1. Чинники, які впливають на споживчий попит

Є кілька чинників, які впливають на споживчий попит. Вони можуть бути [14]:

- типовими, тобто такими, якими можна прогнозувати поведінку покупця на ринку;
- нетипові, коли покупець діє нестандартно;
- неціновими – коли вибір товару не залежить від того, яку він має вартість.

До чинників типових у поведінці покупця можна віднести такі фактори:

1. Здоровий глузд – чим нижча ціна, то більша кількість продукту реалізовуватиметься. Товар, який коштує дорожче, менше привертає увагу клієнтів.

2. Гранична корисність – це «плюшки», які отримає покупець, якщо придбає додаткову одиницю товару. Саме на цьому принципі і ґрунтуються багато рекламних акцій.

Наприклад, коли взуттєвий магазин пропонує купити другу пару взуття за півціни. Або піцерія, де при замовленні на певну суму клієнт отримує «комплімент» від закладу у вигляді додаткової страви чи напою.

Такі акції допомагають підвищити споживчий попит

3. Ефект прибутку. Він проявляється у тому, скільки товару може купити людина, не відмовляючи собі у щоденних витратах. Наприклад, сік 1 л Galicia 51 грн. На тиждень сім'я готова витратити на нього лише 200 грн, яких вистачить на покупку 4 штук. Якщо сік подешевшає до 40 грн за 1 л, то сім'я придбає вже 5 штук.

4. Ефект заміщення. Виявляється у тому, що з нижчої вартості покупець прагне придбати більше звичного продукту, замість шукати йому заміну серед аналогічного товару.

На прикладі того ж соку: зниження ціни призведе до того, що споживач купуватиме саме його, а не замінники в іншого виробника. Останні два фактори

діють лише в сукупності: якщо підприємець реалізує товар хорошої якості, то обидва ефекти діють в одному напрямку.

До чинників нетипової поведінки споживача відноиться випадки, коли підвищення вартості призводить до того, що люди прагнуть закупити ще більше цього продукту. Але таке відбувається в основному тільки з товарами, які складають більшу частину споживчого кошика: молоко, крупи, яйця. Аналогічне відбувалося в період пандемії, коли ціна на гречку, що піднялася, призвела до ще більшого ажіотажу.

Однак результатом такого стає те, що люди перестають купувати інші товари. У них банально не вистачає грошей.

Також до нетипової поведінки покупця на ринку [1] можна віднести:

1. Бажання бути «як усі» чи «не гірше за інших». Коли споживач купує товар лише тому, що це відповідає загальноприйнятим нормам. Наприклад, джинси є в гардеробі практично у кожної людини. Це зручний одяг, в якому можна вийти на прогулянку, сходити на шопінг і навіть на роботу (якщо там немає жорсткого дрес-коду).

2. Бажання відрізнятись від інших. У цьому випадку попит змінюється через те, що більшість людей купує певний товар, а той, хто хоче виділитися, принципово не набуває його. Наприклад, таке спостерігається у підприємців, які спеціалізуються на продажах весільних суконь: обов'язково знайдеться клієнтка, яка не захоче нічого із представленого актуального асортименту.

3. Бажання бути престижним. Покупець обирає товар, спираючись не на його ціну чи якісь характеристики, але на престиж. Наприклад, айфони це круто, але це дорого. Однак їх все одно купують.

І останній чинник, який впливає на споживчий попит – нецінові фактори, оскільки не лише ціна впливає на попит. Є й інші моменти, які підкажуть підприємцю, за рахунок чого можна стати цікавим для покупця:

- мода. Товар, який популярний, матиме попит. Але тут потрібно вчасно вловити момент, коли інтерес до продукту почне пропадати. Найчастіше таке відбувається тоді, коли виробник випускає «нову версію» девайсу чи речі.

- **Пора року.** Наприклад, інтернет-магазин спеціалізується лише на продажу зимових курток, то влітку їхній прибуток стане мінімальним. А от якщо власник бізнесу враховує сезонність, тобто навесні розширить асортимент плащами, легкими пальтами, а влітку – купальними костюмами та пляжними аксесуарами, то зможе розраховувати на отримання хорошого доходу.
- **Демографічна ситуація.** У період бебі-бумів прибуток зростатиме у магазинів, що спеціалізуються на дитячих товарах: візках, одязі, іграшках, сумішах.
- **Економічна ситуація.** Чим вищий рівень доходу потенційного покупця, то більше покупок здійснюється. Але якщо йдеться про товар невисокої якості, то зі зростанням доходу попит на такий продукт впаде.
- **Очікування споживачів щодо зміни ціни на товар.** Якщо покупець передбачає падіння вартості товару, то цей період попит на товар падатиме. І навпаки: коли є ймовірність, що ціна товару зросте, то в цей час кількість продажів зросте. Це особливо помітно на ринку мобільних телефонів: випуск кожної наступної моделі призводить до того, що вартість попередньої знижується.
- **Ціна на взаємозамінні або взаємодоповнювані товари.** На прикладі тих самих мобільних телефонів: зростання вартості певної моделі призведе не тільки до зниження попиту на неї, але й зменшення продажів аксесуарів для неї: чохлів, захисних стекол, бамперів, зарядних пристроїв та іншого.

1.3. Важливість проведення аналізу споживчого попиту

Аналіз споживчого попиту показує, наскільки покупець зацікавлений у продукті та допомагає побудувати стратегію продажів так, щоб зуміти втримати цільову аудиторію.

Якщо інтерес до товару знижується, потрібно переглядати стратегію продажів: змінювати асортимент, покращувати сервіс, «грати» з ціною на товар. Ще аналіз попиту допомагає:

- зрозуміти, наскільки буде потрібний товар ще на етапі відкриття бізнесу;
- з'ясувати переваги цільової аудиторії та скласти асортимент свого магазину з їх урахуванням;
- встановити оптимальні ціни товару;
- вчасно реагувати зміни ринку;
- створити контент, релевантний із запитами, де будуть використовуватися ключові запити та пошукові фрази, якими користується покупець при пошуку товару, що його цікавить;
- оцінити, наскільки продукт, що продається, ризиковий, тобто чи може він стати незатребуваним. Крім того, аналіз може допомогти виявити конкурентів, потенційних та реальних, дасть можливість вивчити стратегію їхньої роботи, її сильні та слабкі сторони.

Для проведення аналізу споживчого попиту необхідно вивчити всі показники, але немає постійного набору тих показників, оскільки все залежить від того, для якого магазину проводитиметься аналіз. Якщо магазин тільки планують відкривати, а для власника бізнесу важливо визначити, який товар матиме попит, то вивчаються:

- ціни та асортимент конкуруючих підприємств;
- загальні тенденції ринку;
- затребуваність конкретних товарів, категорій та окремих позицій;
- динаміку зміни попиту в різні пори року;
- реакції клієнтів, їх відгуки;

А якщо магазин вже працює, то досліджують:

- конверсію;
- розмір середнього чека;
- кількість повторних покупок;
- величину середнього прибутку з покупця;
- кількість кинутих кошиків;

- окупність реклами;
- рентабельність товару.

Ці дані допомагають зрозуміти, наскільки продукт цікавий покупцям (і як цей інтерес зробити ще вищим), дає можливість визначити, у чому сильні та слабкі сторони магазину.

Тому в останні роки стрімкий розвиток отримала область Data Mining

1.4. Моделювання споживчого попиту, як спосіб вивчення його еластичності

Ринкова кон'юнктура залежить від цілого комплексу факторів як соціально-економічних і демографічних, так і природно-природних. Ринок чуйно реагує на зовнішні дії. Сприятливий вплив змушує ринок пожвавлюватись і розширюватися, а негативний – навпаки. Найбільшого впливу зміну ринкової кон'юнктури схильні такі економічні категорії як попит і пропозицію, тобто основні елементи ринкового механізму. У разі жорсткої конкуренції управління попитом дозволяє знизити ризик появи незатребуваної ринком продукції і на збільшити обсяги продажу, а й прибуток підприємства. Треба враховувати що зміна цін на товари, що входять до бюджету споживачів, призводить до зміни їхнього реального доходу до постійного номінального. Причому для споживачів з різним рівнем грошових доходів та саме зміна цін товарів викликає неоднакові зрушення реальних доходів. Такий зсув реального доходу споживачів з грошовим доходом s за зміни цін товарів може бути виміряний індексом реального доходу $I(p,s)$ (індекс Ласпейреса) [11].

Нехай $y_j(s)$ – витрата споживачів із доходом s на товари групи j , а $\sum y_j(s) = s$

Якщо позначити через p вектор цін на товари, за яких визначено криві Енгеля $y_j(s)$ та перейти до часток витрати $\varphi_j(s) = \frac{y_j(s)}{s}$, то $I(p,s)$ визначається співвідношенням

$$I(p, s) = \frac{1}{\sum \frac{p_j \varphi_j(s)}{p_j^0}}$$

тобто відносними змінами цін (індексами). Тому надалі зручніше користуватися індексами цін, зберігаючи колишні позначення та вважаючи $p^0 = 1$. Під терміном «ціна» будемо розуміти індекс цін.

Припустимо, що споживачі із середньодушовим грошовим доходом s при векторі цін p , тобто при зрушенні реального доходу в $I(p, s)$ разів, матимуть таку ж структуру витрат, яку мали споживачі з середньодушовим грошовим доходом $sI(p, s)$ до зміни цін, тобто при векторі цін p^0 . Це припущення називається гіпотезою стійкості структури витрат. Тоді структуру попиту споживачів із доходом s за цінами p можна описати співвідношенням:

$$\varphi_j(s, p) = \varphi_j(sI(p, s)) \quad (1)$$

Гіпотеза, що визначається (1), діє за умов мінімальних змін цін та великої агрегації статей витрати споживачів.

$$y_j(s) > 0$$

$$\varphi_j(p, s) < \varphi_j(s) .$$

Враховуючи, що ціни на товари групи j збільшилися суттєво, натуральне споживання товарів цієї групи має значно зменшитись. Однак можна назвати низку товарних груп, натуральне споживання яких за істотного зростання цін може знизитися не дуже значно. Наприклад, якщо ціни на хліб зросли в кілька разів, то малоймовірно, що купівля хліба також зменшиться, принаймні, у стільки ж разів,

Крім того, сумнів виникає у зв'язку з існуванням парадоксу Гіффена та Веблена [12] для деяких товарів. Сумніви у достовірності гіпотези стійкості структури витрат підкріплюються залежністю поведінки споживачів від загального рівня цін, а не від їх структури.

Виходячи зі сказаного, можна висунути альтернативну гіпотезу поведінки споживачів. Припустимо, що споживачі з доходом s за зміни цін p матимуть структуру натурального споживання таку саму, яку мали споживачі з доходом

$sI(p, s)$ до зміни цін. Це припущення називають гіпотезою стійкості структури споживання. Тоді, з урахуванням нормування, структура попиту споживачів із доходом s за цінами p може бути визначена як

$$\varphi_j(p, s) = \frac{p_j \cdot \varphi_j}{\sum_{k=1}^n p_k \cdot \varphi_k(p, s)} \quad 1.2$$

Нормувальний множник $\frac{1}{\sum_{k=1}^n p_k \cdot \varphi_k(p, s)}$ є індексом Пааше, оскільки зважування цін здійснюється за новою структурою витрат $\varphi_j(p, s)$

Індекс Пааше $I(p, s)$, можна переписати

$$\varphi_j(p, s) = p_j \cdot \varphi_j(p, s) I(p, s) \quad 1.3$$

З приводу гіпотези стійкості структури споживання у свою чергу виникають сумніви щодо її достовірності. Один із аргументів може бути сформульований за допомогою поняття необхідності товарів (товарних груп).

Нехай для деяких j і k $\varphi_d_j(s) > \varphi_d_k(s)$. Тоді говоритимемо, що товар j більш необхідний, ніж k . Якщо $\varphi_d_j(s) = \varphi_d_k(s)$ то потреба цих товарів однакова. Виходячи з цього, всі товари (товарні групи) можна частково впорядкувати за ступенем потреби, визначення якої, природно, ілюструється за допомогою статистики бюджетів населення. Наприклад, зі зростанням доходу частка витрат на хліб та хлібобулочні вироби, цукор, картопля падає, а на фрукти, культурно-побутові товари, транспортні засоби – зростає. Вочевидь, що з зниженні (зростанні) ціни товари першої групи їх натуральне споживання збільшиться (зменшиться) несуттєво. Іншими словами, попит на ці товари мало залежить від рівня цін на них, а на другий групи є еластичною за ціною. Звідси випливає припущення про те, що структура витрат споживачів при зміні цін товарів залежить від упорядкування товарів за ступенем їхньої необхідності: чим потрібніший товар, тим стабільніше натуральне споживання при зміні рівня та структури цін. Це припущення будемо використовувати як третю гіпотезу визначення структури витрат споживачів при зрушенні цін.

Часто її представляють змішаною, оскільки вона може бути як лінійна опукла комбінація двох попередніх.

Структура попиту споживачів з доходом s за цінами p з використанням цієї гіпотези визначається як

$$\varphi_j(p, s) = L(p, s) * [a_j(p, s) * \varphi_j(p, s) + \beta_j(p, s) * \varphi_j(p, s)] \quad 1.4$$

де коефіцієнти $L(p, s) = [\sum a_k(p, s) * \varphi_k(p, s) + \beta_k(p, s) * \varphi_k(p, s)]^{-1}$ – нормувальний

$a_j(p, s)$ – стійкості структури витрат

$\beta_j(p, s)$ – стійкості структури споживання

$$a_j(p, s) \geq 0, \beta_j(p, s) \geq 0, a_j(p, s) + \beta_j(p, s) = 1$$

З формального запису змішаної гіпотези поведінки споживачів можна побачити, що проблема вибору виду та обчислення параметрів функції корисності у нашому у разі трансформується на проблему визначення значень $a_j(p, s)$. При специфікації коефіцієнтів стійкості структури витрат виникають складні питання. Наприклад, значення $a_j(p, s)$ істотно залежать від умов перемикання попиту на інші товари, що входять до складу групи. Припущення, що ці можливості тим ширше, чим більше різноманітність товарів групи часто не працює. Питання про оцінку різноманітності товарів у групах та визначення залежності коефіцієнта стійкості структури витрат від показника різноманітності товарів може бути лише поставлено. Однак є корисним проведення імітаційних експериментів з різними способами формування коефіцієнтів $a_j(p, s)$

Виходячи з розглянутих гіпотез поведінки споживачів, можна поставити завдання відшукування таких цін, за яких прогнозований попит на товари співпадатиме з пропозицією цих товарів. Не можна стверджувати, що встановлення подібних цін призвело б до точного балансу між попитом та пропозицією. Але розраховані ціни рівноваги можуть виявитися корисним «початковим наближенням». Крім того, характер залежності цих цін від параметрів прогнозованого попиту дасть можливість отримати важливі якісні висновки про самі гіпотези поведінки споживачів. В рамках подібної постановки питання можна розглядати різні варіанти спільної зміни цін та пропозиції товарів.

З формальною погляду ціни рівноваги – такі, у яких функції надлишкового попиту перетворюються на нуль. В наших умовах ці функції мають вигляд.

$$F_j(p) = \frac{(\int_0^\infty \varphi_j(p, s) s \mu(s) ds)}{p_j} - V_j$$

де V_j – пропозиція товарів групи j ; $\mu(s)$ – функція щільності розподілу населення за рівнем доходів. Тоді p^* – ціни рівноваги – є рішенням системи

$$p_j V_j = \int_0^\infty \varphi_j(p, s) s \mu(s) ds \quad 1.5$$

або нерухомою точкою відображення

$$R(p) = \{R_1(p), \dots, R_n(p)\}$$

де

$$R_j(p) = \frac{\int_0^\infty \varphi_j(p, s) s \mu(s) ds}{V_j}$$

Оскільки для другої та змішаної гіпотез функції $\varphi_j(p, s)$, не мають умови безперервності на нулі, то не можна стверджувати існування рішення системи (1.5). Те, що вимога безперервності $\varphi_j(p, s)$, – суттєва, можна проілюструвати наступним простим прикладом. Нехай $\varphi_j(s) = \lambda_j (\lambda_j > 0, \sum \lambda_j = 1)$

Тоді, відповідно до другої гіпотези

$$\varphi_j(p, s) = \lambda_j p_j / \sum \lambda_k p_k, \quad F_j(p) = \lambda_j M / \sum \lambda_k p_k - V_j$$

де, $M = \int s \mu(s) ds$ – сумарний прибуток. З цього можна зробити висновок, що в даних умовах система (1.5) має рішення тільки в тому випадку, якщо λ_j пропорціонально V_j .

$$p^{n+1} = \gamma_n p^n + (1 - \gamma_n) R(p^n)$$

У той же час в рамках класичної теорії зазвичай розглядається питання про існування таких цін (рівноважних), за яких усі функції надлишкового попиту непозитивні. При цьому необхідно виконання умов типу замінності, чого у нашому випадку нема.

В практичних розрахунках для знаходження наближеного рішення системи (1.5) можна використовувати стандартний ітеративний процес, який може бути записаний у вигляді

де p^n – вектор цін, одержаний на ітерації n , а γ_n – послідовність параметрів, що демпфують.

В практичних розрахунках часто виявляється корисним спростити функції надлишкового попиту за рахунок заміни функції на її середнє значення M , функції $\varphi_j(p, s)$ – на $\overline{\varphi_j}(p, \bar{s})$ де $\bar{s}$ середньодушовий дохід. У спрощеному вигляді система (1.5) набуде вигляду

$$p_j V_j = \overline{\varphi_j}(p, \bar{s}) M \quad 1.5^*$$

Під час проведення розрахунків спрощена система (1.5*) виявляється корисною. Для аналізу соціальних функцій цін необхідно використовувати більш точну систему (1.5), оскільки в ній враховується соціально-економічна диференціація споживачів.

Якщо вважати зазначені гіпотези справедливими при незначних змінах цін, то можна побудувати граничні функції попиту, що укладаються в рамки класичної теорії. Таким чином, деякі постулати раціональності поведінки споживачів (збереження стійкості структури витрат або натурального попиту) можна пов'язати з його описом, заснованим на припущенні про наявності системи споживчих переваг щодо товарів. Крім того, при такому підході є можливість дати інтерпретацію реакції споживачів на зміну цін та доходів.

Нехай $q_j(p, s)$ система функцій натурального попиту, що задовольняє бюджетного обмеження, тобто $\sum p_j q_j(p, s) = s$ і $\varphi_j(p, s)$ відповідні частки видатків. Будемо говорити, що ця система задовольняє першу гіпотезу при малих змінах цін, якщо

$$\varphi_j(p + \Delta p, s) = \varphi_j(p, s) + o(\Delta p)$$

Де $\Delta p = \{\Delta p_1, \dots, \Delta p_n\}$ – мала зміна цін

$$I = \left(\frac{p_k + \Delta p_k}{p_k} \varphi_k(p, s) \right)^{-1}, \quad \Delta p = \sum |\Delta p_k| \text{ і } \frac{o(\Delta p)}{\Delta p} \text{ прагне до нуля при } \Delta p, \text{ що}$$

прагне до нуля. Покажемо, що за цих умов виконано співвідношення:

$$\frac{d\varphi_j}{dp_k} = -s \frac{\varphi_k}{p_k} \frac{d\varphi_j}{ds} = -q_k \frac{d\varphi_j}{ds} \quad 1.6$$

Дійсно

$$\frac{d\varphi_j}{dp_k} = \lim_{\Delta p_k \rightarrow 0} \frac{\varphi_j(p_j, \dots, p_k + \Delta p_k, \dots, p_n, s) - \varphi_j(p_j, \dots, p_n, s)}{\Delta p_k}$$

Тут I представлено так:

$$\varphi_{j(p_1, \dots, p_k + \Delta p_k, \dots, p_n, s)} = \varphi_j \left(p, s \left[1 + \frac{\Delta p_k}{p_k} \varphi_k(p, s) \right]^{-1} \right) + o(\Delta p_k)$$

$$\begin{aligned} I &= \left[\frac{p_1}{p_1} \varphi_1 + \frac{p_2}{p_2} \varphi_2 + \dots + \frac{p_k + \Delta p_k}{p_k} \varphi_k + \dots + \frac{p_n}{p_n} \varphi_n \right]^{-1} \\ &= \left[\varphi_1 + \varphi_2 + \dots + \varphi_k + \dots + \varphi_n + \frac{\Delta p_k}{p_k} \varphi_k \right]^{-1} = \left[1 + \frac{\Delta p_k}{p_k} \varphi_k \right]^{-1} \end{aligned}$$

В свою чергу

$$\left[1 + \frac{\Delta p_k}{p_k} \varphi_k \right]^{-1} = \frac{p_k + \Delta p_k \varphi_k - \Delta p_k \varphi_k}{p_k + \Delta p_k \varphi_k} = 1 - \frac{\Delta p_k \varphi_k}{p_k + \Delta p_k \varphi_k} = 1 - \frac{\Delta p_k}{p_k} \varphi_k + o(\Delta p_k)$$

тоді

$$\varphi_j \left(p, s - s \frac{\Delta p_k}{p_k} \varphi_k(p, s) + so(\Delta p_k) \right) = \varphi_j(p, s) - \frac{s \Delta p_k}{p_k} \frac{d\varphi_j}{ds} \varphi_k(p, s) + o(\Delta p_k)$$

Підставляючи остаточно вираз $\varphi_j(p_j, \dots, p_k + \Delta p_k, \dots, p_n, s)$ у граничне співвідношення для часткової похідної одержуємо (1.6).

Тепер від частки витрати перейдемо до натурального попиту. Враховуючи що $\varphi_j = \frac{q_j p_j}{s}$, отримуємо

$$\frac{d\varphi_j}{dp_k} = \frac{1}{s} \left(\frac{dq_j}{dp_k} p_j + q \right) = \begin{cases} \frac{dq_j}{dp_k} \frac{p_j}{s} \frac{q_j}{s} & \text{якщо } k = 1 \\ \frac{dq_j}{dp_k} \frac{p_j}{s} & \text{якщо } k \neq 1 \end{cases} \quad 1.7$$

$$\frac{d\varphi_j}{ds} = \frac{p_j}{s} \frac{dq_j}{ds} + \frac{p_j q_j}{s^2} \quad 1.8$$

Використовуючи отримані співвідношення маємо

$$\frac{dq_j}{dp_k} = \begin{cases} \frac{q_j q_k}{s} - q_k \frac{dq_j}{ds} & \text{якщо } k \neq 1 \\ \frac{d^2_j q_j}{s p_j} - q_k \frac{dq_j}{ds} & \text{якщо } k = 1 \end{cases} \quad 1.9$$

Як видно з 1.9

$$\frac{dq_j}{dp_k} + q_k \frac{dq_j}{ds} = \frac{dq_k}{dp_j} + q_j \frac{dp_k}{ds}$$

Проаналізуємо другу гіпотезу. Вважатимемо, що друга гіпотеза виконується при малих змінах цін:

$$\varphi_j^* = (p + \Delta p, s) = p_j \varphi_j(p + \Delta p, s) I$$

Де $\Delta p = \{\Delta p_1, \dots, \Delta p_n\}$ – мала зміна цін

$$\Delta p = \sum \Delta p_k$$

$$I = \left(\sum \frac{p_j + \Delta p_j}{p_j} \varphi_j \right)^{-1}$$

φ_j^* – нова частка j-тої групи товарів у загальному обсязі споживання.

$$\frac{d\varphi_j}{dp_k} = \lim_{\Delta p_k \rightarrow 0} \frac{p_j \varphi_j(p + \Delta p, s) I_n - p_j \varphi_j(p, s) I}{\Delta p_k}$$

Розглянемо два випадки:

$$k \neq j$$

$$\begin{aligned}\frac{d\varphi_j}{dp_k} &= p_j \lim_{\Delta p_k \rightarrow 0} \left(\left(\varphi_j I_H - s \frac{\Delta p_k}{p_k} \frac{d\varphi_j}{ds} \varphi_k(p, s) I_H - \varphi_j I \right) / \Delta p_k \right) \\ &= p_j \lim_{\Delta p_k \rightarrow 0} \frac{\varphi_j (I_H - I)}{\Delta p_k} - p_j \lim_{\Delta p_k \rightarrow 0} \left(\left(\frac{\Delta p_k}{p_k} \frac{d\varphi_j}{ds} \varphi_k I \right) / \Delta p_k \right)\end{aligned}$$

$$\begin{aligned}I_H &= I \frac{1}{\sum p_j \varphi_j + \Delta p_k \varphi_k} - \frac{1}{\sum p_j \varphi_j} = \frac{\sum p_j \varphi_j - \sum p_j \varphi_j - \Delta p_k \varphi_k}{(\sum p_j \varphi_j + \Delta p_k \varphi_k) \sum p_j \varphi_j} \\ &= \frac{\Delta p_k \varphi_k}{(\sum p_j \varphi_j + \Delta p_k \varphi_k) \sum p_j \varphi_j}\end{aligned}$$

Враховуючи вираз для $I - I_H$, – отримаємо

$$\frac{d\varphi_j}{dp_k} = p_j \varphi_k \varphi_j I^2 - p_j \frac{s}{p_k} \frac{d\varphi_j}{ds} \varphi_k I = \frac{\varphi_j \varphi_k}{p_j} - q_k \frac{d\varphi_j}{ds}$$

Таким чином, ми вивели наступну рівність:

$$\frac{dq_j}{dp_k} = -q_k \frac{dq_j}{ds} \quad 1.10$$

$$2. \quad k = j$$

В результаті аналогічних міркувань отримуємо:

$$\frac{dq_j}{dp_k} = -q_k \frac{dq_j}{ds} - \frac{dq_j}{dp_k} = \frac{q_j}{p_j} \quad 1.11$$

Формули 1.9, 1.10, 1.11, що описують граничні зміни функції попиту при варіації цін дозволяють уточнити змістовну інтерпретацію гіпотез.

Таким чином, якщо споживачі схильні до підтримки структури витрат, то попит більш чутливий (проти прагненням до підтримки структури натурального споживання) до зміни ціни цей товар, ніж інші, тобто проявляється цінова еластичність.

1.5. Співвідношення сумарних витрат та доходу, як фактор попиту

Закон попиту звучить так: при незмінності інших факторів попит на товари та послуги в кількісному вираженні змінюється у зворотній залежності від ціни (назад пропорційно ціні). Отже, закон відбиває залежність попиту ціни.

Зміна обсягу попиту при зміні ціни представлено у графічному співвідношенні попиту та пропозиції на рис. 1.1 [10]:

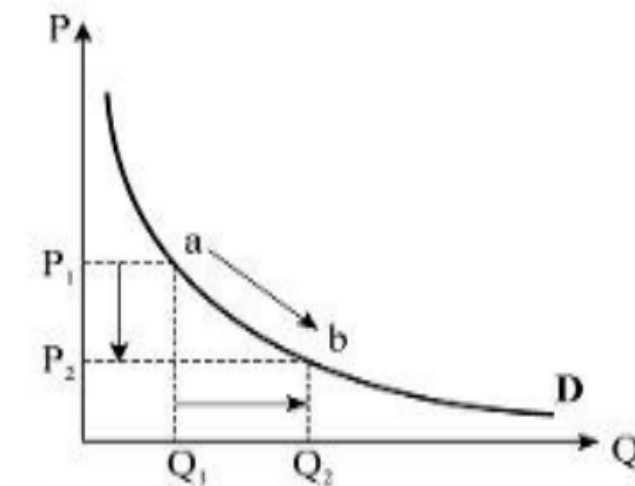


Рисунок 1.1 Зміна обсягу попиту при зміні ціни

І ситуація, у зміні в попиті, внаслідок дії нецінових чинників представлено на рис. 1.2. графічно відповідає зміщенню всієї кривої попиту: праворуч-вгору, якщо попит зростає, і ліворуч-вниз, якщо попит зменшується

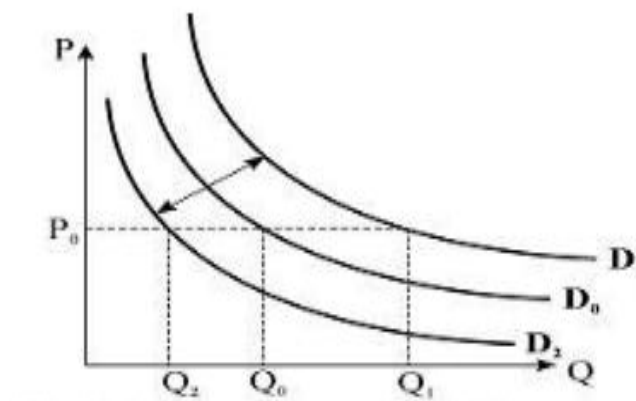


Рисунок 1.2 Зміни у попиті внаслідок дії нецінових чинників

Закон порушується лише у трьох випадках:

1. При очікуваному зростанні цін або в кризових ситуаціях (приклади з ажіотажним попитом на гречку та туалетний папір в часи пандемії Covid-19).
2. При зміщенні попиту в більш дорогі і якісні товари зі зростанням доходів населення.
3. На ринках преміальних та люксових товарів, які часто розглядаються як інвестиція, а також купуються для задоволення вторинних вигод (так званий ефект сноба – підвищення статусу в очах інших, приналежність до «еліти», тощо).

Обсяг пропозиції залежить від наступних факторів:

Виробничі здібності продавця. Наприклад, виробник не може завантажити виробничі потужності більш ніж на 100%. Щоб збільшити обсяги виробництва, йому доведеться або ввести в експлуатацію ще одну виробничу лінію або скористатися аутсорсингом. Цей же вибір постане перед компанією після ухвалення рішення про розширення асортименту або диверсифікації бізнесу. Аналогічним чином салон краси не може надати послуг більше, ніж протягом робочого дня, кожен майстер може обслужити клієнтів, а магазин мати наявності в торговому залі більше SKU, ніж це фізично можливо на наявному устаткуванні.

Попит. Підприємство може виробляти товари та послуги або мати можливості для їх виробництва, але якщо вони нецікаві цільовій аудиторії або мають занадто високі ціни, сенсу в їх виробництві немає. Компанії доведеться працювати на склад і вирішувати проблему з його розпродажем. З іншого боку, в моменти пікового попиту, наприклад, у сезон, саме склади з виробленої продукції можуть задовольнити зростаючі запити. Так наперед готуються до Нового року виробники сувенірної продукції, подарункових наборів та ін. У сфері послуг наростити пропозицію буває простіше. Часто для цього потрібне лише збільшення штату клієнтського персоналу. Таким чином працюють служби таксі, які залучають водіїв з автомобілями, кур'єрські служби та компанії, що надають онлайн-послуги, наприклад, курси іноземних мов. У випадку з кав'ярнями,

салонами краси, магазинами та іншими видами послуг, прив'язаними до площ та залежними від їх розміру, може знадобитися відкриття нових точок та перехід у мережевий формат.

Кількість гравців на ринку. Чим більше гравців на ринку, тим сильніша між ними конкуренція. Відповідно, тим менше їх збутові можливості (пропозиція). З цієї причини знайти підприємства із завантаженням потужностей майже 100% у будь-яких галузях вкрай складно. Середнє завантаження становить близько 50-60%. Завантаження підприємств різних галузей залежить від конкретного регіону, асортименту, цінової політики, маркетингової та збутової активностей та інших факторів.

Етап життєвого циклу ринку. На ринку, що зароджується, попит і пропозиція мінімальні – перші гравці формують ринок. У міру прийняття товарів та послуг ринком зростає інтерес у них і, відповідно, зростає пропозиція. Якщо пропозиція не встигає за попитом, на ринку з'являється дефіцит. Така ситуація може спостерігатися на етапі швидкого зростання ринку. Якщо пропозиція перевищує попит, це надлишок. Він часто характерний для другої половини етапу зрілості та етапу спаду.

Політична та економічна ситуація в країні та у світі. Наприклад, запровадження санкцій призвело до того, що багато ринків зіткнулися з дефіцитом. Розвиток технологій та інновації. Нові технології та впровадження інновацій можуть істотно збільшити продуктивність підприємства, а також зробити його товари та послуги привабливішими для споживачів.

Функція пропозиції від ціни та інших факторів, що на нього впливають, виглядає наступним чином:

$$Q_s = Q_s (P, Y_1, Y_2 \dots Y_N)$$

Q_s – розмір попиту конкретного товару (s – supply, пропозиція).

P – ціна на даний товар.

$Y_1, Y_2 \dots Y_N$ – інші фактори, що впливають на попит, перераховані вище.

Закон пропозиції: при незмінності інших факторів пропозиція товарів та послуг у кількісному вираженні змінюється у прямій залежності від ціни (прямо пропорційно до ціни). Закон визначає залежність пропозиції від ціни.

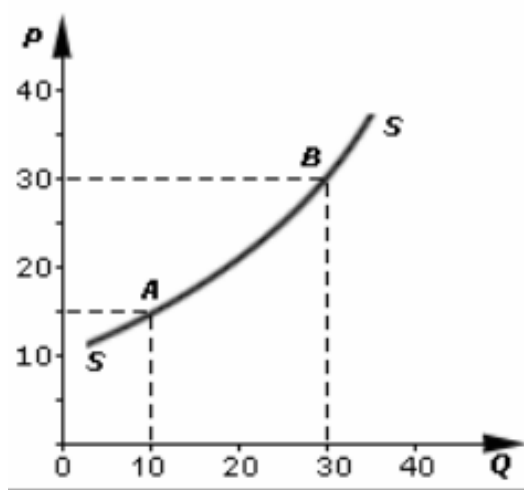


Рисунок 1.3 Крива пропозиції, де P – ціна, Q – кількість пропозиції

Докладніше в випускній кваліфікаційній роботі [10]

ВИСНОВКИ ДО ПЕРШОГО РОЗДІЛУ

Показано один із підходів до кількісної оцінки споживчого попиту. Визначено закономірності попиту. Дано обґрунтування концепції еластичності з використанням запропонованої моделі споживчого попиту.

Моделі є спрощеним способом, які представляють і пояснюють поведінку споживачів на ринку і дуже необхідні на практиці. Деякі моделі роблять це комплексно, інші підкреслюють динаміку рішень споживача та показують певні напрями можливих досліджень поведінки споживачів. Знання щодо поведінки споживача є цінним джерелом інформації про дії споживачів у процесі купівлі товарів. Моделі дозволяють знизити ризик прийняття підприємством невірних рішень і цим уникнути можливих збитків.

РОЗДІЛ ДРУГИЙ

ТЕОРІЯ ТА МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

2.1. Поняття «Data Mining»

Data Mining або у літературі найбільш використовувана аналогія – інтелектуальний аналіз даних, ІАД, спрямована на пошук та розробку методів вилучення з наявних даних знань, що дозволяють приймати на їх основі конкретні, високо обґрунтовані, практично корисні управлінські рішення [11].

Найбільший інтерес до технологій інтелектуальної обробки даних виявляють компанії, що працюють в умовах високої конкуренції та мають чітку групу споживачів (роздрібна торгівля, фінанси, зв'язок, маркетинг). Вони використовують будь-яку можливість підвищення ефективності власного бізнесу через прийняття більш ефективних управлінських рішень. Ці компанії намагаються знайти зв'язок між «внутрішніми» (ціна, затребуваність продукту, компетентність персоналу тощо) та «зовнішніми» (економічні показники, конкуренція, демографія клієнтів тощо) факторами. Це дозволяє їм оцінити (спрогнозувати) рівень продажу та задоволеності клієнтів, розмір доходів, а також сформулювати на основі сукупності всієї наявної інформації практично корисні висновки та рекомендації. Іноді віддача від застосування цих інструментів може становити сотні відсотків за порівняно невисокої вартості впровадження. При цьому результатом обробки даних повинен бути такий інформаційний продукт, який дозволяє прийняти конкретну управлінську дію без надмірного «занурення» особи, яка приймає рішення до деталей базових даних або проміжної аналітики (наприклад, дати рекомендації щодо купівлі/продажу на фінансовому ринку, сформулювати перелік заходів щодо збільшення продуктивності або маркетингу продукту тощо). Причому на практиці можлива ситуація, при якій будь-яке рішення в тій чи іншій частині необхідно приймати обов'язково – питання лише в тому, чи воно приймається на основі об'єктивної інформації чи інтуїтивно.

Вилучення своєчасної і готової безпосередньої інформації з різних джерел для прийняття управлінських рішень передбачає створення деяких інформаційних продуктів. Прикладами таких інформаційних продуктів у бізнесі можуть бути відповіді на такі питання:

- Який із продуктів слід рекламувати більше для збільшення прибутку?
- Як слід удосконалити програму модернізації для зменшення витрат?
- Який процес виробництва змінити, щоб зробити продукт кращим?

Ключ до відповіді на ці питання вимагає глибокого розуміння наявних даних та їх індуктивного аналізу.

Розглянуто деякі приклади застосування методів інтелектуального аналізу даних, що використовуються в бізнес-середовищі та підтверджують на практиці зростаючу актуальність цієї інтелектуальної сфери людської діяльності.

2.2. Застосування інтелектуального аналізу даних

Головна перевага інтелектуального аналізу даних полягає в його здатності визначати зв'язки та закономірності у величезних обсягах даних із різних ресурсів. З доступністю все більшої кількості даних із різноманітних джерел (дистанційні датчики, соціальні мережі, докладні звіти про ринкову активність і переміщення товарів) інтелектуальний аналіз даних пропонує інструменти для використання великих даних і перетворення їх у практичну інформацію.

Компанії, що займаються збиранням даних, допомагають виявляти закономірності та зв'язки в частинах інформації, які, здавалося б, не пов'язані між собою. Оскільки інформація має тенденцію бути розділеною, її, як правило, важко проаналізувати в цілому. Тим не менш, може існувати взаємозв'язок між зовнішніми факторами, такими як економічні фактори чи демографічні показники, і продуктивністю продукції компанії. Незважаючи на те, що керівники здебільшого розглядають показники продажів за лінійкою продуктів, територією, регіоном і каналом збуту, вони зазвичай не мають зовнішнього контексту для інформації. Інтелектуальний аналіз даних допомагає в цьому контексті та в

розумінні «чому» за фактами. Технологія враховує кореляції із зовнішніми факторами, і ці кореляції можуть бути надійними індикаторами для скерування рішень про канал, виробництво та продукт.

Використовуючи методи інтелектуального аналізу даних, пункт роздрібної торгівлі (магазин) може фіксувати інформацію про всі покупки клієнта та розсилати націлені рекламні пропозиції своїм клієнтам на основі історії їх покупок. Аналізуючи демографічну інформацію клієнтів, магазин може пропонувати товари та рекламні пропозиції для конкретного клієнтського сегмента. Всесвітньо відома торговельна мережа США WalMart – «ветеран» інтелектуального аналізу даних, застосованого для модернізації взаємодії з постачальниками. Компанія WalMart проаналізувала транзакції 2900 магазинів з 6 країн, сформувала сховище даних обсягом 7,5 Тбайт. При цьому потрібно було виконати понад 1 млн. складних запитів. Дані використані для визначення патернів покупців при вдосконаленні мерчендайзингових стратегій для 3500 постачальників.

Типовими питаннями за якими йде пошук відповідей під час аналізу даних у роздрібній торгівлі, є:

- Хто покупець?
- Як сегментувати клієнтів?
- На яку цільову аудиторію зробити акцент?
- Які фактори впливають на рішення про покупку?
- Яка значимість кожного з факторів?
- Які товари пропонувати у спільних акціях?
- Які існують залежності у поведінці клієнтів?
- На який обсяг попиту в майбутньому орієнтуватись?

Той самий аналіз також приносить користь іншим частинам бізнесу, таким як дизайн продукту, надання послуг і ефективність роботи.

Інтелектуальний аналіз даних вигідний для більшості підприємств, перш за все, тому, що він може переглядати величезні обсяги даних і виявляти приховані

закономірності, зв'язки та тенденції. Результати корисні для прогнозної аналітики, яка допомагає у стратегічному плануванні, зберігаючи запас поточного бізнес-сценарію. Узагальнено, перевагами інтелектуального аналізу для підприємств є:

1. Цільовий маркетинг і реклама. Інтелектуальний аналіз даних дозволяє командам маркетингу краще розуміти поведінку та переваги клієнтів. Це дозволить їм спрямовувати цільову рекламу на відповідних клієнтів, які демонструють модель поведінки. Крім того, відділ продажів отримує вигоду від інтелектуального аналізу даних, оскільки це допомагає їм орієнтуватися на клієнтів із особливою схильністю до певних продуктів. Крім того, це дозволяє їм продавати більше послуг і продуктів літнім клієнтам.

2. Виявлення проблем із обслуговуванням клієнтів. Інтелектуальний аналіз даних є ефективним інструментом для відстеження проблем із обслуговуванням клієнтів, коли клієнти взаємодіють із агентами контакт-центру за допомогою дзвінків і онлайн-чатів. Це дає їм можливість надавати краще обслуговування клієнтів завдяки поглибленому аналізу, можливого за допомогою аналізу даних.

3. Удосконалене управління ланцюгом поставок (SCM). За допомогою інтелектуального аналізу даних підприємства можуть визначати ринкові тенденції та передбачати майбутню поведінку клієнтів, яка може вплинути на попит на продукт. Це дозволяє підприємствам планувати майбутнє та управляти пропозицією товарів і послуг відповідно до потреб ринку. Крім того, менеджери SCM можуть відповідно планувати свої логістичні операції, оптимізувати розподіл продукції та оптимізувати складські послуги.

4. Підтримка безвідмовної роботи. Збір і видобуток даних із датчиків, пристроїв Інтернету речей, виробничих машин і промислового обладнання допомагає створювати додатки для прогнозованого технічного обслуговування, які визначають потенційні проблеми до того, як фактичний інцидент пошкодить обладнання. Такі завчасні попередження скорочують позаплановий простой машин, тим самим підвищуючи загальну продуктивність.

5. Краще оцінювати ризики. Інтелектуальний аналіз даних дозволяє менеджерам із ризиків і зацікавленому бізнес-персоналу краще оцінювати ризики, пов'язані з фінансами, юридичними питаннями чи факторами кібербезпеки, з якими компанія може зіткнутися в майбутньому. Це дає їм можливість належним чином підготуватися до таких подій і мати план, щоб краще впоратися з такими невдачами.

6. Зниження витрат. Інтелектуальний аналіз даних може легко виявити будь-яку операційну неефективність у типовому бізнес-процесі. Таке раннє виявлення проблеми допомагає оптимізувати корпоративні процеси відповідно до бізнес-цілей компанії, завдяки чому значно заощаджується на корпоративних витратах.

Інтелектуальний аналіз даних відіграє ключову роль у розробці стратегічних планів, які допомагають компаніям отримувати більший бізнес-прибуток і доходи та відрізнити їх від своїх конкурентів.

2.2. Інтелектуальний аналіз даних в бізнесі та торгівлі

Інтелектуальний аналіз даних – це техніка сортування величезних наборів даних. Це допомагає розпізнавати зв'язки та закономірності для вирішення бізнес-проблем. Компанії з аналізу даних використовують інструменти та процеси, щоб допомогти організаціям прогнозувати майбутні тенденції та приймати рішення. Інструменти включають потужні математичні, аналітичні та статистичні можливості, головна мета яких – просіювати величезні обсяги даних для виявлення закономірностей, взаємозв'язків і тенденцій для підтримки обґрунтованого планування та прийняття рішень.

Інтелектуальний аналіз даних, як правило, пов'язаний із запитами відділу маркетингу та розглядається багатьма керівниками як метод, що допомагає їм зрозуміти попит і проаналізувати вплив змін у продуктах, рекламних акціях і ціноутвореннях на продажі. Тим не менш, інтелектуальний аналіз даних також показує численні переваги в інших сферах бізнесу. Дизайнери та інженери

можуть зрозуміти плідність змін продукту та визначити потенційні причини успіху або невдачі продукту, пов'язані з тим, як продукти використовуються. Операції, пов'язані з обслуговуванням і ремонтом, краще планують запаси запчастин і персонал. Організації, які надають професійні послуги, можуть використовувати послуги інтелектуального аналізу даних, щоб розпізнавати нові можливості через демографічні зміни та зміну економічних тенденцій.

Інтелектуальний аналіз даних стає ще кориснішим і ціннішим із розширеним користувацьким досвідом і більшими наборами даних. Вважається, що величезні набори даних володіють більшою інтелектуальністю та розумінням. І коли користувачі більше знайомляться з інструментами інтелектуального аналізу даних і вчать розуміти базу даних, вони стають більш обізнаними та творчими у своєму аналізі та дослідженнях.

2.3. Методи та засоби інтелектуального аналізу даних

Існує велика кількість різних підстав для стратифікації, категоризації, класифікації значної кількості існуючих та новостворених методів Data Mining (рис. 2.1). Наприклад, можна зустріти класифікації за принципом роботи з вихідними навчальними даними (зазначають вони чи ні в результаті обробки змін), за типом одержуваного результату (передбачувальні та описові), за видами застосовуваного математичного апарату (статистичні та кібернетичні) та ін. За типом використовуваного математичного апарату, зазвичай, виділяють такі основні групи методів Data Mining [6]:

1. **Дескриптивний аналіз та опис вихідних даних, попередній аналіз природи статистичних даних** (перевірка гіпотез стаціонарності, нормальності, незалежності, однорідності, оцінка виду функції розподілу, її параметрів тощо).

2. **Багатомірний статистичний аналіз** (лінійний та нелінійний дискримінантний аналіз, кластерний аналіз, компонентний аналіз, факторний аналіз тощо).

3. Пошук зв'язків та закономірностей (лінійний та нелінійний регресійний аналіз, кореляційний аналіз тощо).

4. Аналіз часових рядів (динамічні моделі та прогнозування).

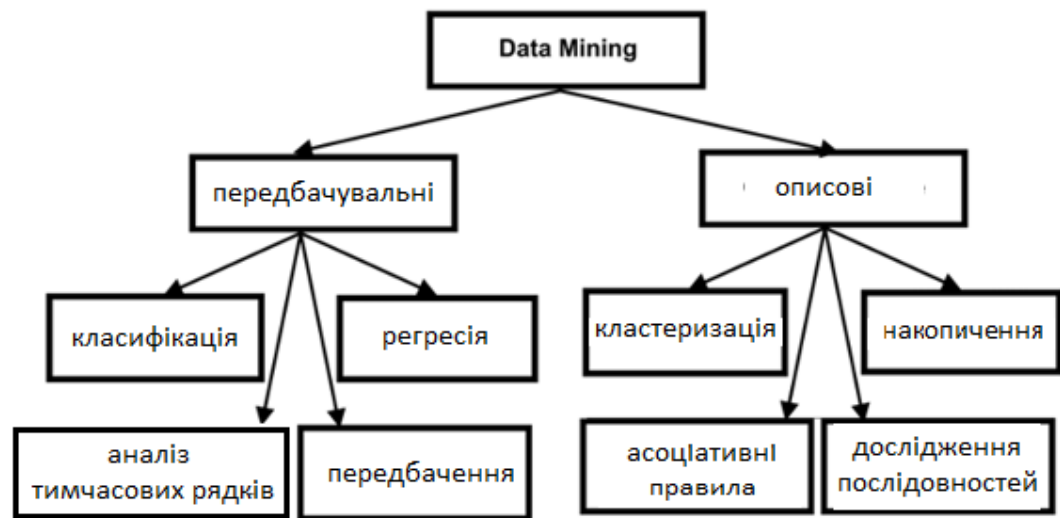


Рис. 2.1 – Ілюстрація прикладу класифікації методів Data Mining

Існують різні класифікації типів закономірностей, що дають змогу відносити використані методи до методів Data Mining але, зазвичай, їх виділяють п'ять [7]:

1. Асоціація.
2. Послідовність.
3. Класи.
4. Кластери.
5. Тимчасові рядки.

Асоціація (англ. Association) має місце у разі, якщо кілька подій пов'язані один з одним. Наприклад, дослідження показує, що 85% покупців, які купували хот-доги, купують і газований напій Coca-cola. Ця асоціація дозволяє запропонувати знижку за такий тип продуктового «комплекту» і, можливо, тим самим збільшити обсяги продажів.

Якщо кілька подій пов'язані друг з одним у часі, має місце тип залежності, названий послідовністю (англ. Sequential Patterns). Наприклад, після покупки

квартири в 75% випадків протягом місяця купується і холодильник, а в межах двох тижнів 80% новоселів купують пральну машину.

Закономірність класів (англ. Classes) у разі, якщо є кілька заздалегідь сформованих класів (груп, типів) об'єктів. Віднесення нового об'єкта до якогось із існуючих класів виконується шляхом класифікації.

Закономірність кластерів (англ. Clusters) відрізняється тим, що класи (групи, типи) заздалегідь не задані, які кількість і склад визначаються автоматично внаслідок процедури кластеризації.

Ретроспективна інформація, що зберігається, дозволяє визначити ще одну закономірність, яка полягає в пошуку існуючих часових рядів (англ. Time Series) і прогнозуванні динаміки значень у них на майбутні періоди часу.

Деталізуючи використовуваний математичний апарат, що є найважливішим компонентом практично будь-яких сучасних методів Data Mining, можна отримати значно глибшу класифікацію існуючих методів [13]:

- Метричні методи класифікації (метод найближчих сусідів та його узагальнення, відбір еталонів та оптимізація метрики).
- Логічні методи класифікації (поняття закономірності та інформативності, вирішальні списки та дерева).
- Лінійні методи класифікації (градієнтні методи, метод опорних векторів)
- Байєсовські методи класифікації (оптимальний байєсовський класифікатор, параметричне та непараметричне оцінювання щільності, поділ суміші розподілів, логістична регресія).
- Методи регресійного аналізу (багатовимірні лінійна регресія, нелінійна параметрична регресія, непараметрична регресія, неквадратичні функції втрат, прогнозування часових рядів).
- Нейромережні методи класифікації та регресії (багатошарові нейронні мережі).

- Композиційні методи класифікації та регресії (лінійні композиції, бустинг, евристичні та стохастичні методи, нелінійні алгоритмічні композиції).
- Критерії вибору моделей та методи відбору ознак (завдання оцінювання та вибору моделей, теорія узагальнюючої здатності, методи відбору ознак).
- Навчання без вчителя (кластеризація, мережа Кохонена, таксономія; пошук асоціативних правил, завдання з частковим навчанням, колаборативна фільтрація, тематичне моделювання, навчання з підкріпленням).

2.3.1. Етапи проведення інтелектуального аналізу даних

Виділяють такі типові етапи, що супроводжують вирішення завдань інтелектуального аналізу даних (рис.2.2) [13]:



Рис. 2.2 – Етапи інтелектуального аналізу даних

1. Аналіз предметної галузі, формулювання цілей та завдань дослідження. На етапі постановки завдання необхідно визначити, що є метою аналізу. Зокрема, потрібно відповісти на низку питань, головне з яких – що саме необхідно визначити в результаті аналізу.

2. Вилучення та збереження даних. Етап підготовки даних включає визначення джерел даних для аналізу, об'єднання даних та їх очищення. Дані, що використовуються, можуть знаходитися в різних базах і на різних серверах. Більш того, якісь дані можуть бути представлені у вигляді текстових файлів, електронних таблиць, перебувати в інших форматах. У процесі об'єднання та перетворення даних часто використовуються можливості служб SQL Server Integration Services. Це дозволяє суттєво автоматизувати процес підготовки. Зібрані таким чином дані, як правило, потребують додаткової обробки даних.

3. Попередня обробка даних [6]:

- очищення (англ. cleaning; виключення протиріч, випадкових викидів та перешкод, перепусток);
- інтеграція (англ. integration; об'єднання даних із кількох можливих джерел в одному сховищі);
- перетворення (англ. transformation; може включати агрегування та стиснення даних, дискретизацію атрибутів та скорочення розмірності тощо).

4. Вивчення даних, яке дозволить зрозуміти, наскільки адекватно підготовлений набір представляє предметну область, що досліджується. Тут може проводитися пошук мінімальних та максимальних значень параметрів, аналіз розподілу значень та інших статистичних характеристик, порівняння отриманих результатів з уявленнями про предметну область.

5. Побудова моделей. Спочатку створюється структура даних, а потім для структури створюється одна або кілька моделей. Модель включає вказівку на алгоритм інтелектуального аналізу даних та його параметри, а також аналізовані дані. При визначенні моделі можна використовувати різні фільтри. Таким чином, не всі дані, що є в описі структури, будуть використовуватися кожною створеною для неї моделлю. Часто для розв'язання задачі створюється кілька моделей, що ґрунтуються на різних алгоритмах, щоб була можливість порівняти результати та вибрати найкращу.

6. Перевірка моделі. Тут метою є оцінка якості роботи створеної моделі перед початком її використання у «виробничому середовищі». Якщо

створювалося кілька моделей, то на цьому етапі робиться вибір на користь тієї, що дасть найкращий результат. При вирішенні передбачуваних завдань інтелектуального аналізу якості прогнозу, що видається моделлю, можна оцінити на перевірочному наборі даних, для якого відомо значення прогнозованого параметра. У MS SQLServer служби Analysis Services надають кошти, що спрощують поділ даних на навчальний та перевірочний набори. Таке секціонування можна виконати автоматично під час побудови інтелектуального аналізу даних. Точність прогнозів, створюваних моделями, можна перевірити за допомогою таких засобів, як діаграма точності прогнозів та матриця класифікації. Інший підхід, званий перехресною перевіркою, полягає в тому, що створюються підмножини даних та порівнюються результати роботи моделі на кожному підмножині. Такий підхід може використовуватися як під час вирішення передбачуваних, і описових завдань. Засоби автоматизації перехресної перевірки доступні при використанні MS SQL Server версії Enterprise або Developer. Найбільш ефективні моделі розгортаються у виробничому середовищі. При цьому, можливі сценарії інтеграції засобів інтелектуального аналізу даних та додатків користувача. І кінцевий користувач у відповідь на сформований запит отримуватиме результати аналізу у вигляді звіту. Під час формування звітів про результати проведеного аналізу можуть використовуватись можливості служби SQLServer Reporting Services. Згодом характеристики предметної області можуть змінюватися, що вимагатиме й зміни шаблонів інтелектуального аналізу даних. Може знадобитися перенавчання існуючих моделей або створення нових. У ряді випадків SQLServer дозволяє автоматизувати процес оновлення моделей за рахунок використання служб Integration Services.

7. Використання нових знань прийняття рішень.

2.3.2. Попередня обробка даних

Практичне застосування методів Data Mining передбачає багатоетапну процедуру. Одним з ключових етапів цієї процедури, що передує, власне,

застосування методів Data Mining є етап попередньої обробки даних, що включає різні типи перетворень.

Одним із ключових перетворень етапу попередньої обробки даних є «очищення» даних (англ. Data Cleaning, Data Cleansing, Data Scrubbing), що передбачає виявлення та коригування/видалення пошкоджених елементів даних.

Дані, які мають такі ушкодження (неточні, неповні, дубльовані, суперечливі, зашумлені) називають «брудними». Джерелами «брудних» даних можуть бути пошкоджені інструменти збору даних, проблеми у введенні вихідних даних, «людський фактор» у разі неавтоматичного варіанту формування даних, проблеми в каналах передачі даних, обмеження технологій передачі даних, використання різних найменувань у межах однієї номенклатури тощо. Особливу актуальність наявності очищення брудних даних підтверджує відомий в інформатиці вираз – «Сміття на вході – сміття на виході» (англ. Garbage In – Garbage Out, GIGO19). Воно означає, що з неправильних вхідних даних буде отримано неправильні результати роботи, в принципі, правильного алгоритму.

Дійсно, практично корисними результати застосування будь-яких методів Data Mining будуть лише у разі використання ними коректних достовірних даних. Враховуючи те, що такі дані можуть бути доставлені з різних джерел і бути досить суттєвими в обсязі, завдання отримання та обробки «чистих» даних може бути дуже непростим. Слід зазначити, що наявність «брудних» даних може бути більш проблематичним, ніж їх відсутність зовсім – вилучення корисних знань з даних може потребувати значного часу, причому безрезультатно. При цьому ще більш проблематичним буде успішне вилучення з таких даних недостовірних знань та подальше їхнє практичне використання з важко передбачуваними наслідками. Саме тому етапу отримання «чистих», готових до аналізу даних надають великого значення, а за витратами часу цей етап може бути одним із найтриваліших [15].

Проблемам отримання «чистих» даних присвячено окремі досить ємні дослідження [15], де обговорюється цілий спектр різних особливостей цієї проблематики, починаючи від концептуальних питань і завершуючи деталями

сучасних технологічних рішень у базах даних і сховищах даних. Можна виділити деякі найважливіші моменти. Усі проблеми очищення даних поділяють на дві групи, викликані інтеграцією різних джерел даних (MultiSource Problems) або обумовлені проблемами єдиного джерела даних (англ. Single-Source Problems). В свою чергу кожна з груп може бути розділена на дві інші групи, що визначаються або недосконалістю схем баз даних, що інтегруються (англ. Schema Level), або недосконалістю на рівні власне елементів даних (англ. Instance Level, записів, об'єктів і т.п.). Далі кожна з гілок одержаного дерева класифікації деталізується конкретним переліком можливих проблем очищення даних (рис. 2.3 проблеми якості даних).

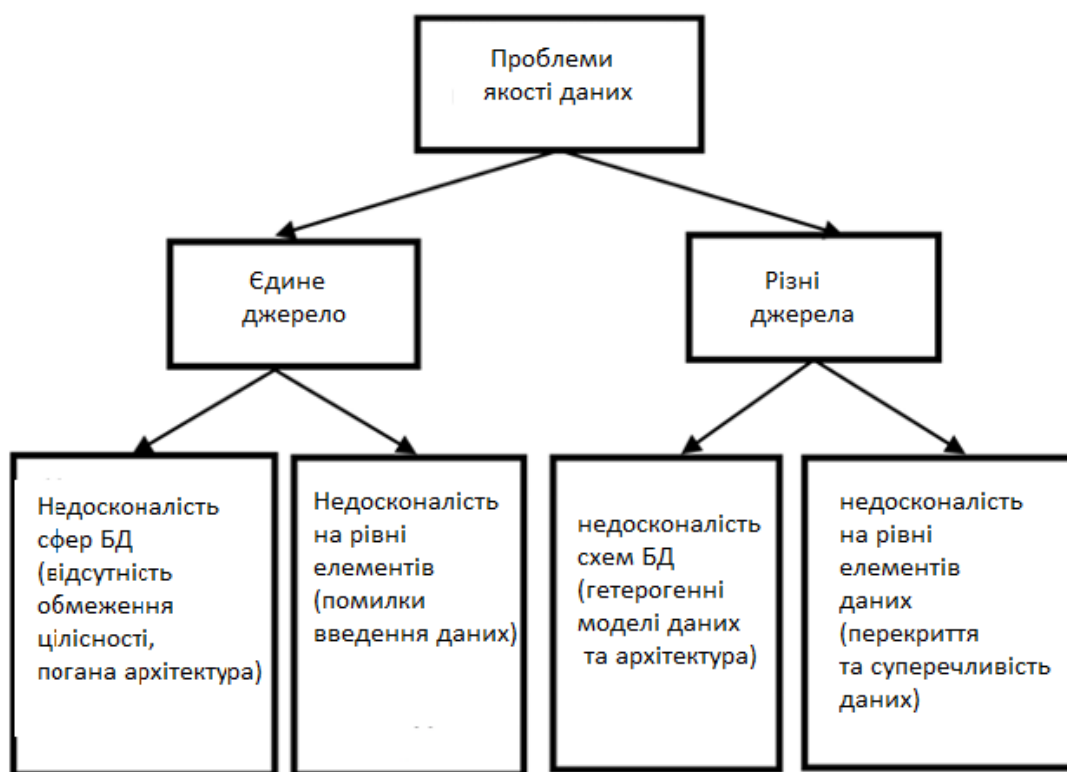


Рис. 2.3 – Проблеми якості даних

Прикладами «брудних» даних єдиного джерела на рівні схеми даних можуть бути:

- співр.1=(ім'я=Михайло, SSN=345) співр.2=(ім'я=Микола, SSN=345)
SSN, означає, що є порушення унікальності

- дата народження 32.12.2002 – Неприпустимі значення, оскільки за межами діапазону
- тощо

Виділяють наступні етапи очищення даних [13]:

1. Аналіз даних (англ. Data analysis). Щоб визначити, які види помилок і невідповідностей мають бути видалені, потрібен детальний аналіз даних. На додаток до інспекції даних або окремих вибірок даних «вручну» слід використовувати і метадані.

2. Визначення способів трансформації потоків даних та правил відображення (англ. Definition of transformation workflow and mapping rules). На даному етапі виконується оцінка кількості джерел даних, ступеня їхньої неоднорідності та «забрудненості». На основі цієї інформації створюються схеми потоків даних, які дозволяють перетворювати безліч джерел даних в один, уникаючи створення помилок Multi-Source злиття (наприклад, поява дублюючих записів).

3. Верифікація (англ. Verification). Оцінка коректності та результативності виконання попереднього етапу (наприклад, на невеликій вибірці даних). При необхідності проводити повернення до етапу 2 для його повторного виконання.

4. Трансформація (англ. Transformation). Завантаження даних до єдиного сховища з використанням правил трансформації, визначених та налагоджених на етапах 2–3. Очищення даних рівня Single-Source.

5. Зворотне завантаження очищених даних (англ. Backflow of cleaned data). Маючи на етапі 4 очищений набір даних в єдиному сховищі, доцільно цими чистими даними замінити аналогічні брудні дані у вихідних джерелах. Це дозволить у майбутньому багато в чому не виконувати повторно всі етапи перетворень щодо очищення даних.

Реалізувати ці етапи можна різними шляхами з використанням існуючих і створених спеціально способів і технологій. Етап аналізу даних передбачає [13] аналіз використання метаданих, яких зазвичай недостатньо для оцінки якості даних з наявних джерел. Тому важливо аналізувати реальні приклади даних,

оцінюючи їх характеристики та сигнатури значень. Це дозволяє знаходити взаємозв'язки між атрибутами у схемах даних різних джерел. Виділяють два підходи розв'язання цього завдання [13]

- профіль даних (англ. data profiling);
- вилучення даних (англ. data mining).

Профільювання даних зорієнтоване на аналіз індивідуальних атрибутів, що характеризуються їх конкретними властивостями: тип даних, довжина, діапазон значень, частота дискретних значень, дисперсія, унікальність, зустрічність «null» значень, типова сигнатура запису (наприклад, у телефонного номера). Саме набір таких властивостей (профіль) дозволяє оцінити різні аспекти якості даних. Вилучення даних передбачає пошук взаємозв'язків між кількома атрибутами досить великого набору даних.

Крім того, для доповнення пропущених значень, коригування неприпустимих значень або ідентифікації дублікатів можуть бути використані обмеження цілісності (англ. integrity constraints), прийняті в реляційних базах даних, накладені додатково на бізнес-зв'язки між атрибутами. Наприклад, відомо, що $Total = Quantity \times Unit_Price$. Усі записи, які не задовольняють цій умові, повинні бути вивчені уважніше, виправлені або виключені з розгляду.

Для вирішення проблем очищення даних в одному джерелі (single-source problems), у тому числі перед його інтеграцією з іншими джерелами даних, реалізують наступні етапи [15]:

- Вилучення значень з атрибутів вільної форми (розбиття атрибутів, англ. Extracting values from free-form attributes (Attribute split)). У цьому випадку може йтися про рядкові значення, що зберігають кілька слів поспіль (наприклад, адресу або повне ім'я людини). У цьому випадку потрібно чітко розуміння того, на якій позиції цього значення знаходиться частина атрибуту, яка цікавить. Можливо, знадобиться навіть сортування складових частин такого атрибуту.

- Валідація та корекція (англ. Validation and correction). Даний етап передбачає пошук помилок введення даних та їх виправлення найбільш автоматичним способом. Наприклад, використовуючи автоматичну перевірку

правопису, щоб уникнути орфографічних помилок і друкарських помилок. Словник географічних назв і поштових кодів також слід використовувати для коригування значень адрес, що вводяться. Залежність атрибутів (дата народження - вік, $Total = Quantity \times Unit_Price$ тощо) також сприяє уникненню багатьох помилок у даних.

- Стандартизація (англ. Standardization). Цей етап передбачає приведення всіх даних до єдиного універсального формату. Прикладами таких форматів є формат написання дати та часу, розмір регістру написання строкових значень. Текстові поля повинні виключати префікси та суфікси, аббревіатури у них мають бути уніфіковані, виключені проблеми з різним кодуванням. Однією з основних проблем, спричинених інтеграцією різних джерел (multi-source problems) даних, є усунення дублювання записів. Цей етап виконується після переважної більшості перетворень та чисток. Він передбачає спочатку ідентифікацію подібних у певному сенсі записів, та був їх злиття з об'єднанням атрибутів. Очевидно, вирішення цього завдання за наявності дублюючих записів первинного ключа досить просто. Якщо такого однозначно ідентифікуючої ознаки немає, завдання усунення дублікатів значно ускладнюється, вимагаючи застосування нечітких (англ. fuzzy) підходів порівняння (близько у певному сенсі) записів між собою.

2.6. Моделі інтелектуального аналізу

Модель інтелектуального аналізу даних створюється шляхом застосування алгоритму даних. Але це більше, ніж алгоритм або контейнер метаданих: це набір даних, статистик та шаблонів, які можна застосовувати до нових даних для формування прогнозів та виведення взаємозв'язків.

У цьому розділі описані моделі інтелектуального аналізу даних та можливі варіанти їх використання: базова архітектура моделей та структур, властивості моделей інтелектуального аналізу даних, способи їх створення та застосування.

2.6.1. Архітектура моделі інтелектуального аналізу даних

Модель інтелектуального аналізу даних отримує дані із структури інтелектуального аналізу даних та аналізує їх, застосовуючи алгоритм інтелектуального аналізу даних. Структура інтелектуального аналізу даних та модель інтелектуального аналізу даних є окремими об'єктами. У структурі інтелектуального аналізу даних зберігаються відомості, що визначають джерело даних.

Модель інтелектуального аналізу даних містить відомості, отримані за підсумками статистичної обробки даних, наприклад, закономірності, виявлені в результаті аналізу.

Модель інтелектуального аналізу даних буде порожня доти, доки не будуть оброблені та проаналізовані дані, передані структурою інтелектуального аналізу даних. Після обробки модель інтелектуального аналізу даних містить метадані, результати та прив'язки до структури інтелектуального аналізу даних.

Метадані визначають ім'я моделі та сервер, де вона зберігається, а також визначення моделі, включаючи список стовпців із структури інтелектуального аналізу даних, які використовувалися для побудови моделі, визначення всіх фільтрів, що застосовувалися при обробці моделі, та алгоритм, який використовувався для аналізу даних. Всі ці варіанти – стовпці даних та їх типи даних, фільтри та алгоритми мають сильний вплив на результати аналізу.

Наприклад, одні й ті самі дані можна використовувати для створення кількох моделей, використовуючи алгоритм кластеризації, алгоритм дерева прийняття рішень і спрощений алгоритм Байеса. У кожному з типів моделей створюються різні набори шаблонів, наборів елементів, правил і формул, які можна застосовувати під час прогнозування. Як правило, кожен з алгоритмів аналізує дані по-своєму, тому вміст моделі також організується в різні структури. В одному з типів моделей дані та шаблони можуть групуватися в кластери; у моделі іншого типу дані можуть бути впорядковані за допомогою дерев, гілок та правил, що розділяють та визначають дані.

Модель також залежить від даних, на яких проводилося її навчання: навіть ті моделі, навчання яких проводились на основі однієї і тієї ж структури інтелектуального аналізу даних, можуть видавати різні результати, якщо під час аналізу фільтрація даних виконувалася по-різному або використовувалися різні початкові значення. Однак фактичні дані не зберігаються у зведеній статистиці, доступній лише моделі, зберігаються з фактичними даними, що у структурі інтелектуального аналізу даних. Якщо під час навчання моделі були створені фільтри даних, визначення фільтрів також зберігаються в об'єкті моделі.

Модель містить набір прив'язок, що вказують на кешовані у структурі інтелектуального аналізу дані. Якщо в процесі обробки структури дані були поміщені в кеш і не були видалені з нього, то ці прив'язки дозволять виконувати деталізацію від результатів до варіантів, що утворює безліч цих результатів. Фактичні дані при цьому зберігаються в кеші структури, а не моделі.

Алгоритми, які надаються в SQL Server інтелектуальному аналізу даних, є найбільш популярними, добре вивченими методами формування шаблонів на основі даних. Щоб взяти один приклад, кластеризація K-середніх є одним із найстаріших алгоритмів кластеризації та широко доступна в багатьох різних засобах та з безліччю різних реалізацій та параметрів. Однак конкретна реалізація кластеризації K-середніх, що використовуються в SQL Server інтелектуальному аналізу даних, була розроблена Корпорацією Майкрософт Research, а потім оптимізована для підвищення продуктивності за допомогою SQL Server Analysis Services.

Всі алгоритми інтелектуального аналізу даних Майкрософт доступні для гнучкого налаштування та програмування з використанням API, що надаються. Є можливість автоматизувати створення, навчання та перенавчання моделей за допомогою компонентів інтелектуального аналізу даних у службах Integration Services.

З іншого боку, підтримується використання сторонніх алгоритмів, відповідних специфікації OLE DB для інтелектуального аналізу даних. Існує також можливість розробляти власні алгоритми, які можна зареєструвати як

служби, а потім використовувати в платформі інтелектуального аналізу даних SQL Server.

Щоб створити модель інтелектуального аналізу даних, необхідно виконати такі дії [15]:

1. Створити базову структуру інтелектуального аналізу даних та увімкнути до неї стовпці даних, які можуть знадобитися.
2. Вибрати алгоритм, який найкраще підходить для аналітичного завдання.
3. Вибрати стовпці зі структури для використання в моделі та вказати, як їх слід використовувати, в якому стовпці міститься результат, який потрібно спрогнозувати, які стовпці призначені лише для введення, і тощо.
4. Задати додаткові параметри для тонкого налаштування обробки, яке проводиться алгоритмом.
5. Заповнити модель даними, виконавши обробку структури та моделі.

2.6.2. Вибір правильного алгоритму

Вибір правильного алгоритму для використання у конкретній аналітичній задачі може бути досить складним. У той час як можна використовувати різні алгоритми для виконання однієї і тієї ж задачі, кожен алгоритм видає різний результат, а деякі алгоритми можуть видавати більше одного типу результатів. Наприклад, алгоритм дерева прийняття рішень (Майкрософт) можна використовувати не тільки для прогнозування, але також як спосіб зменшення кількості стовпців у наборі даних, оскільки дерево прийняття рішень може ідентифікувати стовпці, що не впливають на кінцеву модель інтелектуального аналізу даних.

Вибір алгоритму за типом. SQL Server інтелектуальний аналіз даних включає такі типи алгоритмів [15]:

- Алгоритми класифікації здійснюють прогнозування однієї чи кількох дискретних змінних з урахуванням інших атрибутів у наборі даних.

- Регресивні алгоритми здійснюють прогнозування однієї чи кількох безперервних числових змінних, наприклад прибутку чи збитків, з урахуванням інших атрибутів у наборі даних.
- Алгоритми сегментації поділяють дані на групи або кластери елементів, що мають схожі властивості.
- Алгоритми взаємозв'язків здійснюють пошук кореляції між різними атрибутами набору даних. Найбільш частим застосуванням цього алгоритму є створення правил взаємозв'язку, які можуть використовуватися для аналізу споживчого кошика.
- Алгоритми аналізу послідовностей узагальнюють послідовності, що часто зустрічаються в даних, такі як серія переходів по веб-сайту або подій, зареєстрованих в журналі перед ремонтом обладнання.

Однак ніщо не змушує користувача обмежуватись одним алгоритмом у своїх рішеннях.

SQL Server інтелектуальний аналіз даних дозволяє створювати кілька моделей в одній структурі інтелектуального аналізу даних, тому в рамках одного рішення інтелектуального аналізу даних можна використовувати алгоритм кластеризації, модель прийняття рішень і модель Naïve Bayes для отримання різних уявлень про дані. В одному рішенні можна використовувати кілька алгоритмів для виконання окремих завдань. Наприклад, за допомогою регресії можна отримувати фінансові прогнози, а за допомогою алгоритму нейронної мережі виконувати аналіз факторів, що впливають на прогнози.

2.6.3. Стовпці моделі інтелектуального аналізу даних

Модель інтелектуального аналізу даних містить стовпці даних, одержувані з певних структур інтелектуального аналізу даних стовпців. Можна вибрати стовпці з структури інтелектуального аналізу даних, які використовуються в моделі, а також створити копії стовпців із структури інтелектуального аналізу даних, перейменувати їх або змінювати спосіб їх використання. У процесі

створення моделі необхідно визначити способи використання стовпців в моделі. Наприклад, стовпець може бути ключем, використовуватися для прогнозу чи взагалі пропускатися алгоритмом. При створенні моделі замість автоматичного додавання всіх доступних стовпців даних рекомендується уважно проаналізувати дані в структурі та включити в модель ті стовпці, аналіз яких має сенс. Наприклад, не слід включати в модель кілька стовпців з ідентичними даними, а також використовувати стовпці, що містять переважно унікальні значення. Якщо стовпець не підходить для використання, його не потрібно видаляти із структури чи моделі інтелектуального аналізу даних. Можна просто встановити для стовпчика прапор, що вказує, що стовпець повинен пропускатися під час побудови моделі. Це означає, що стовпець залишиться у структурі інтелектуального аналізу, але не використовуватиметься в моделі інтелектуального аналізу даних. Якщо деталізацію з моделі до структури інтелектуального аналізу даних включено, то відомості зі стовпця можна буде отримати пізніше.

Залежно від вибраного алгоритму, деякі стовпці в структурі інтелектуального аналізу даних можуть виявитися несумісними з конкретними типами моделей або викликати погіршення якості результатів. Наприклад, якщо дані містять числові дані в безперервному інтервалі (наприклад, у стовпці доходу), а моделі потрібні дискретні значення, то може знадобитися перетворити дані на дискретні значення або виключити їх з моделі. У деяких випадках алгоритм автоматично перетворює дані або розподіляє їх за групами, але результати таких операцій можуть виявитися непередбачуваними або небажаними. Також необхідно розглянути можливість створення додаткових копій стовпця та перевірки застосування різних моделей. Також можна задати для окремих стовпців прапори, що вказують, яка потрібна спеціальна обробка. Наприклад, якщо дані містять порожні значення (NULL), то керувати їх обробкою можна скористатися прапором моделі. Якщо певний стовпець у моделі повинен вважатися регресором, цього можна домогтися за допомогою прапора моделі.

Після створення моделі можна вносити зміни, наприклад, додавати або видаляти стовпці або змінювати ім'я моделі. Однак після будь-якої зміни, навіть якщо воно стосувалося тільки метаданих моделі, необхідно виконати повторну обробку моделі.

2.6.4. Обробка моделей інтелектуального аналізу даних

Модель інтелектуального аналізу даних до обробки є порожнім об'єктом. Під час обробки моделі дані, які були поміщені в кеш структурою, передаються через фільтр, якщо він був визначений моделі, і піддаються аналізу відповідно до заданого алгоритму. Алгоритм обчислює набір зведених статистичних показників, що описують дані, виявляє правила та закономірності даних, а потім на основі правил і закономірностей виробляє заповнення моделі.

Після обробки модель інтелектуального аналізу даних містить цінні відомості про дані та виявлені при аналізі закономірності, включаючи статистичні показники, правила та формули регресії. Переглянути ці відомості можна за допомогою засобів перегляду або створивши запити інтелектуального аналізу даних, які витягуватимуть ці відомості та використовувати їх для аналізу та подання.

2.6.5. Властивості моделі інтелектуального аналізу даних

Кожна модель інтелектуального аналізу даних має властивості, які визначають модель та її метадані. До цих властивостей входять: ім'я, опис, дата останньої обробки моделі, дозволи на модель, а також всі фільтри для даних, які використовувалися для навчання.

Кожна модель інтелектуального аналізу даних також містить властивості, успадковані від структури інтелектуального аналізу даних, які описують стовпці даних, що використовуються в моделі. Якщо будь-який зі стовпців, що

використовуються моделлю, є вкладеною таблицею, то до нього також може застосовуватися окремий фільтр.

Крім того, кожна модель інтелектуального аналізу даних має дві спеціальні властивості: Algorithm та Usage.

Властивість Algorithm визначає алгоритм, який використовується для створення моделі. Набір доступних алгоритмів залежить від постачальника. Можна змінити алгоритм пізніше, але деякі стовпці в моделі інтелектуального аналізу даних можуть стати неприпустимими, якщо вони не підтримуються вибраним алгоритмом.

Після зміни якості моделі завжди необхідно виконувати повторну обробку моделі.

Властивість Usage визначає, які шпальти будуть використані моделлю. Можна визначити стовпець як Вхідні дані, Прогноз, Тільки прогноз, Ключ. Властивість Usage застосовується до окремих стовпців моделі інтелектуального аналізу даних і має задаватися окремо для кожного стовпця, включеного до моделі. Якщо структура містить стовпець, який не використовується в моделі, для використання задається значення Пропустити. Як приклад даних, які можуть включатися до структури інтелектуального аналізу даних, але не використовуватися при аналізі, можна навести імена клієнтів або адреси електронної пошти. У такому разі до них можна буде виконувати запити пізніше, але не включати їх на етапі аналізу.

Значення властивостей інтелектуального аналізу даних можна змінити після створення моделі. Однак після будь-якої зміни, навіть якщо воно стосувалося лише імені моделі інтелектуального аналізу даних, необхідно виконати повторну обробку моделі. Після повторної обробки моделі можна отримати інші результати.

ВИСНОВКИ ДО ДРУГОГО РОЗДІЛУ

Визначено що перевагою інтелектуального аналізу даних є його здатність визначати зв'язки та закономірності у величезних обсягах даних із різних ресурсів. Інтелектуальний аналіз даних пропонує зручні інструменти для використання великих обсягів даних і перетворення їх у практичну інформацію чим вигідний для більшості підприємств.

Проаналізовано моделі інтелектуального аналізу даних та можливі варіанти їх використання. Встановлено що кожна модель інтелектуального аналізу даних має властивості, які визначають модель та її метадані.

ТРЕТІЙ РОЗДІЛ

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ СПОЖИВЧОГО ПОПИТУ НА ПРИКЛАДІ ПРИВАТНОГО ПІДПРИЄМСТВА «ЗДОРОВИЙ ПРОДУКТ»

3.1. Аналіз та сегментування клієнтської бази

Метою в практичній частині випускної кваліфікаційної є розробка методики сегментації клієнтської бази з використанням методів кластерного аналізу даних про продаж. В роботі виконано кластерний аналіз з використанням методів машинного навчання щодо динамічних ознак клієнтів, коли не потрібно вести персоніфікований облік клієнтів.

Для того, щоб зробити найвигідніші пропозиції чи відкоригувати план продажу, тобто, оптимізувати діяльність компанії в зовнішніх умовах, що постійно змінюються необхідно дослідити та побудувати моделі споживчого попиту на різні групи товарів та послуг. Складність вирішення цих завдань обумовлена мінливістю таких важливих зовнішніх факторів, як конкурентне середовище, а також наміри покупців, які важко спрогнозувати, тому що постійно виникає дисбаланс в поінформованості учасників ринку (продавців та покупців) про умови угоди – якість товару, його ціну, характеристики та ін. Таке явище інформаційної асиметрії [5, 6] є нерівномірним розподілом між учасниками ринку інформації про умови здійснення ринкової угоди та наміри один одного. У цьому доцільним є опис «портрета» цільової аудиторії, тобто, ознак клієнтів [11]. Для цього було враховано велику кількість прихованих факторів їхньої поведінки, які не мають яскраво виражених причинно-наслідкових зв'язків. Звичайно, необхідно збирати якнайбільше статичних даних про клієнтів (стаття, вік, географічні дані), але найпростіше вести знеособлений облік клієнтів, наприклад, за виданими на руки дисконтними (бонусними) картками з унікальним ідентифікатором. До цієї карти можна прив'язати індивідуальний кабінет, де за згодою клієнта можна отримати його додаткові статичні ознаки [11].

Ще складніша ситуація з динамічними показниками, які формуються на підставі поведінки окремого клієнта щодо інших клієнтів. В цьому випадку можливо вести не тільки суворий облік покупок окремого клієнта, а й виконувати узагальнений аналіз клієнтської бази за відомими методиками та показниками ефективності: RFM-класифікація, визначення розміру середнього чека, цінності клієнта тощо. [1, 9]. Так чи інакше необхідно зіставити ознаки покупців з ознаками угод та товарів, що вони споживають протягом певного періоду часу. Однак у більшості випадків відомі методики аналізу та сегментування клієнтської бази [8, 9] або орієнтовані отримання статичних ознак клієнтів, що дозволяє отримати збалансований статистично значимий портрет цільової аудиторії, або є класифікаційними, тобто, поділяють клієнтську базу за однією з наперед визначених ознак, наприклад, давність покупки, середній обсяг покупки тощо. [4], при цьому не враховуються комплексні фактори, що узагальнюють та поділяють клієнтів на окремі кластери. Кластерний аналіз виконується для об'єднання клієнтів за окремими ознаками чи групами ознак. Причому ці ознаки можуть ставитися як до статичних персоніфікованих ознак клієнта, і до динамічним ознаками, що характеризує його дії – замовлення, купівлі, тобто, ознаки угод.

3.1. Методологія дослідження

Щоб дослідити наміри покупців, насамперед ідентифікувати кожного покупця, з зіставленням ознак товарів, які він набував протягом певного періоду часу. Однак, при великій кількості клієнтів-покупців, висока трудомісткість побудови моделі поведінки для кожного клієнта окремо, до того ж це буде малокорисним та інформативним по відношенню до моделей поведінки інших клієнтів. В такому випадку, сегментування клієнтської бази і для кожного сегмента побудувати власну модель, яка усереднює моделі переваг всіх клієнтів, що потрапили в даний сегмент за різними характеристиками угод є найкращим розв'язанням [2]. Маючи в своєму розпорядженні дані про портрет цільової

аудиторії, продавці легко зможуть звертатися до різних сегментів клієнтської бази з певними пропозиціями, тим самим збільшуючи лояльність клієнтів [2].

Найбільш популярним класичним алгоритмом кластеризації є метод k -середніх [7]. Реалізація такого алгоритму та його різновидів відносно проста та показана у роботі з використанням інструментарію електронних таблиць [12]. Першим кроком і водночас недоліком даного алгоритму є необхідність завдання кількості кластерів k , на яке ділитиметься простір ознак. Існують окремі методи обчислення оптимальної кількості кластерів. В інших випадках зручніше задати значення k вручну та проаналізувати результат розбиття. Слід зазначити, що розбиття може здійснюватися й у двовимірному просторі ознак, наприклад, ID клієнта/№ угоди, а подробиці угоди зможуть додатково відображати характеристики купленого товару. На наступному кроці k довільних точок призначаються центрами кластерів, і обчислюється відстань між призначеними кластерними центрами та іншими точками всередині простору кластеризації. Приналежність точки до кластера визначається обчисленням найменшої відстані одного з k -центрів. В якості застосовуваних заходів може бути використана евклідова відстань між m об'єктами

$$x_i = (x_{i1}, \dots, x_{im}) \in X_Q, i = \overline{1, Q} \text{ та центрами кластерів}$$

$$d_{2k}(x_i, x_j) = \sqrt{\sum_{t=1}^m (x_{it} - x_{jt})^2} \quad 3.1$$

Алгоритм є ітеративним і на наступному кроці змінюється положення кластерних центрів і знову проводиться розподіл точок по k -кластерам, і операція повторюється до тих пір, поки значення відстаней всіх об'єктів до кожного кластеру не мінімізуються, це означає, що досягнуто оптимального розбиття. Цільовою функцією є мінімізація відстаней до кожного i -го об'єкта всередині j -го кластера:

$$I(k, m) = \sum_{i=1}^m d_{2imin} \quad 3.2$$

Завершальним етапом роботи алгоритму є аналіз та інтерпретація сформованих кластерів з метою визначення узагальнюючих ознак, якими точки в

кластерах є схожими, тобто які з особливостей поведінки користувачів є систематичними.

В роботі приведено приклад сегментації клієнтської бази за допомогою електронних таблиць. На початку необхідне вивантаження даних із бази з продажу товарів. На рис. 3.1 показаний приклад структури бази даних з обліку продажу дисконтних карт клієнтів.

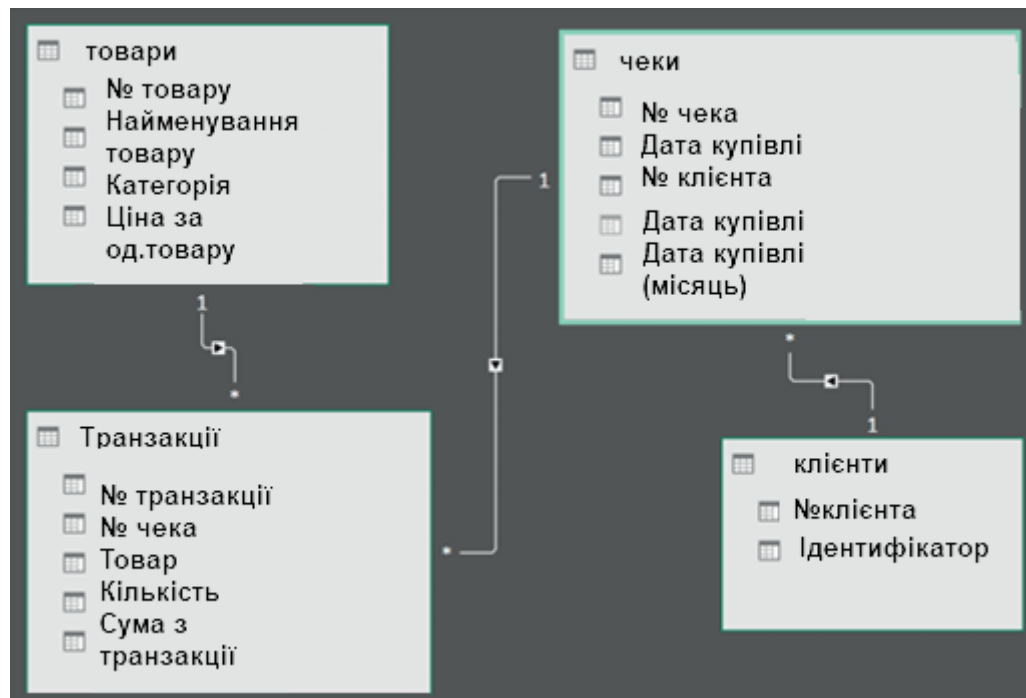


Рис. – Структура бази даних з обліку продажі

З цієї бази було взято інформацію щодо найменування товарів та тимчасовий інтервал, протягом якого ці товари купувалися, тому на основі бази побудовано зведену таблицю за угодами, клієнтами та періодами продажів (рис. 3.2). Тим самим визначається ключові ознаки угод, якими виконуватиметься кластеризація клієнтів.

	A	B	C	D	E	F
2						
3	Ідентифікатор клієнта					
4	Найменування	1003500010001	1003500010002	1003500010004	1003500010006	1003500010007
5	Апельсини	2,56			9,84	
6	Баклажани			5,12		
7	Банани	3,09		8,71		
8	Виноград	6,61				
9	Груша	3,41				
10	Капуста	14,9		6,28		
11	Картопля	11,03			3,15	
12	Огірок	8,55	2,97		5,52	
13	Помідор	9,25	3,67			
14	Яблука			7,33		1,12
15	Дата купівлі (Місяць)					
16	січ лют берез квіт					
17						
18						

Рисунок 3.2 – Зведена таблиця по товарам, клієнтам та періодам продажів

В роботі аналіз проводився з урахуванням даних по продажах за чотири місяці і виявляються особливості поведінки всього масиву покупців-власників дисконтних карток в той період. Слід зазначити, що у цьому випадку використовується обмежений набір товарів із десяти товарних позицій для зручності ілюстрації роботи алгоритму. Як значення елементів даної крос-таблиці використовується кількість одиниць (в даному випадку кг) продукції. Насправді, на даному етапі важливим є сам факт покупки, тому далі перетворено цю таблицю так, щоб на перетині рядків і стовпців були показані одиниці тільки у разі покупки (як показано на рис. 3.3).

	A	B	C	D	E	F	G	H	I	J
1	Місяць	Найменування	Ціна	Категорія	кластер 1	кластер 2	кластер 3	кластер 4	1003500010001	1003500010002
2	січень	апельсини	67	фрукти					1	
3	січень	баклажани	46	овочі					1	
4	січень	банани	25	фрукти					1	
5	січень	виноград	65	овочі					1	
6	січень	груша	50	фрукти					1	
7	січень	капуста	12	овочі					1	
8	січень	картопля	13	овочі					1	
9	січень	огірок	67	овочі					1	1
10	січень	помідор	65	овочі					1	1
11	січень	яблука	15	фрукти						
12	лютий	апельсини	67	фрукти						
13	лютий	баклажани	46	овочі						
14	лютий	банани	25	фрукти						
15	лютий	виноград	65	овочі						
16	лютий	груша	50	фрукти						

Рисунок 3.3 – Підготовка даних під кластеризацію (частина даних прихована)

На рис. 3.3 у стовпцях A-D показані ознаки угод щодо кожної товарної позиції за місяцями. Всього вийшло 40 рядків, які є узагальнюючими ознаками всіх угод за 4 місяці. Стовпці E-H відведені під масиви вагових коефіцієнтів належності кожної групи угод певному кластеру. Далі починаючи зі стовпця I і далі розташовані індикатори вчинення того чи іншого типу угоди кожним із врахованих за дисконтними картками клієнтів. У прикладі використовується вибірка більш ніж 700 клієнтів. На рис. 3.4 показано початковий стан кластеризації, коли всі стандартні об'єкти належать до кластера №1. Для кожного клієнта розраховується відстань до кожного кластера з використанням метрики за формулою 3.1, потім мінімальної відстані визначається належність даного клієнта до певного кластера. Як цільову функцію використовується мінімум суми відстаней кожного клієнта до кластера, а як змінні – вагові коефіцієнти в стовпцях E-H, що лежать в межах від 0 до 1. Ці вагові коефіцієнти показують, наскільки цей тип угоди по всім клієнтам близький до центру кластера. Також після визначення належності, до якого кластеру відносяться всі угоди щодо кожного клієнта, легко вести підрахунок частки клієнтів, що належать до того чи іншого кластера, а також здійснити інтерпретацію результатів (рис. 5).

	A	B	C	D	E	F	G	H	I	J	K
	Місяць	Найменуванн	Ціна	Категор	кластер 1	кластер 2	кластер 3	кластер 4	1003500010001	1003500010002	1003500010002
1	січень	апельсини	67	фрукти					1		
2	січень	баклажани	46	овочі							1
3	січень	банани	25	фрукти					1		1
4	січень	виноград	65	овочі					1		
5	січень	груша	50	фрукти					1		
6	січень	капуста	12	овочі					1		
7	січень	картопля	13	овочі					1		1
8	січень	огірок	67	овочі					1		
9	січень	помідор	65	овочі					1	1	
10	січень	яблука	15	фрукти					1		
11	січень	апельсини	67	фрукти							1
12	лютий	баклажани	46	овочі							1
13	лютий	банани	25	фрукти							
14	лютий	виноград	65	овочі							
15	лютий	груша	50	фрукти							
16	лютий	капуста	12	овочі							
17	лютий	картопля	13	овочі							
18	лютий	огірок	67	овочі							
19	лютий	помідор	65	овочі							
20	лютий	яблука	15	фрукти							
21	лютий										
42					доля клієнтів в кластері	100	відст. до кластеру 1		3,16227766	2	2,645751311
43					доля клієнтів в кластері	0	відст. до кластеру 1		3,16227766	2	
44					доля клієнтів в кластері	0	відст. до кластеру 1		3,16227766	2	
45					доля клієнтів в кластері	0	відст. до кластеру 1		3,16227766	2	
46					цільова функція	1727,039929	мін. відстань		3,16227766	2	
47							приналежність		1	1	

Рисунок 3.4 – Розрахунок мір близькості об'єктів до кластерів

Мінімізація суми відстаней угод кожного клієнта до кластера виконується з допомогою підбору вагових коефіцієнтів в стовпцях E-H за допомогою еволюційного пошуку рішення. Зрозуміло, що точніше збіжність алгоритму

пошуку рішення і що більше час пошуку, то точніше розбиття. На рис. 3.5 показані результати розбиття пар клієнт/угода за кластерами. При цьому частка клієнтів, віднесених до кластера 1, склала 22%, до кластера 2 – 17%, до кластера 3 – 28% та до кластера 4 – 33%. На прикладі зазначених інтерпретовано отримане розбиття кластерами. Так, наприклад, кластер №1, для якого спостерігається піковий ваговий коефіцієнт, що дорівнює 0,98, для угоди, що характеризується купівлею товару «капуста» у квітні місяці, зроблено висновок, що 22% клієнтів, яких можна ідентифікувати, схильні купувати саме цей продукт в квітні. Також у даному кластері можна спостерігати множину дрібних піків серед вагових функцій, які означають не ідеальне визначення меж кластера.

Кластер №2 інтерпретовано як покупців, схильних купувати банани та виноград також у квітні, але ця категорія покупців менш численна і становить лише 17% від загальної кількості клієнтів. Невисокі значення коефіцієнтів для інших типів угод також характеризують, які товари ще воліють клієнти в даному кластері.

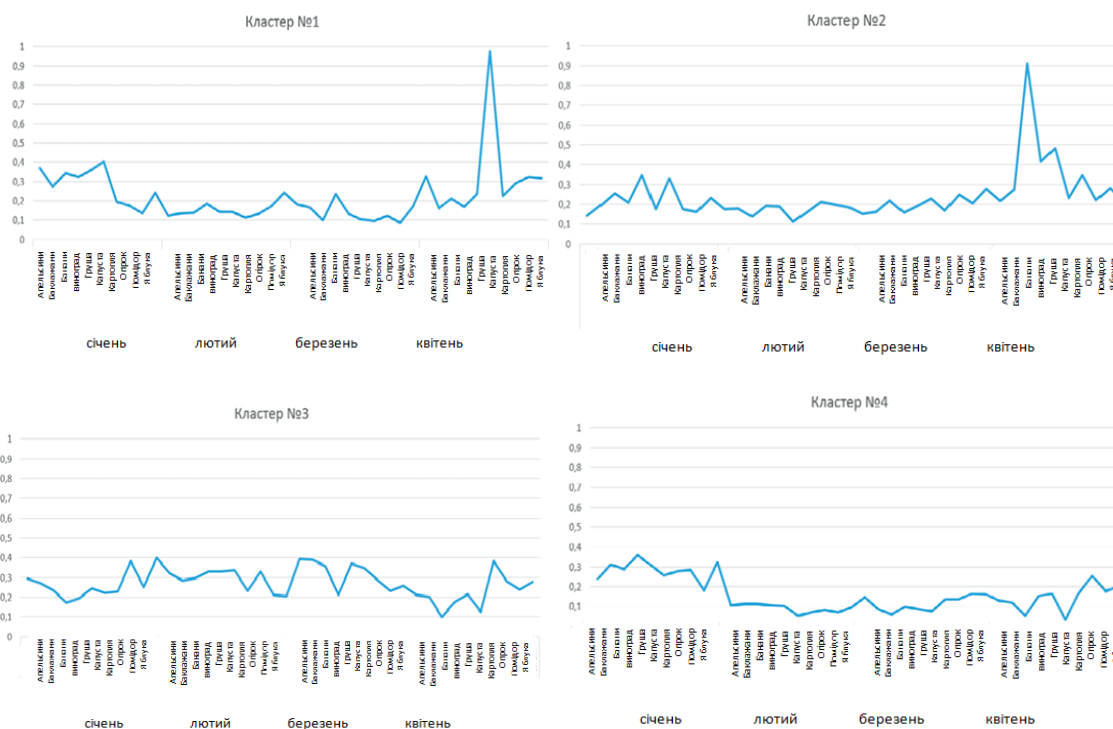


Рис. 3. 5 – Сегментація клієнтської бази за типами угод

Третій кластер має найгірше розбиття і слабо піддається інтерпретації, тому що не має яскраво виражених піків і по ньому складно визначити наміри клієнтів щодо наведених товарів. Можливо, для цього необхідно провести додаткове дослідження або скоротити або збільшити кількість кластерів, а також збільшити кількість ознак.

Останній четвертий кластер також не має піків вагових коефіцієнтів, що перевищують значення 0,5, проте всередині кластера спостерігається збільшення цих коефіцієнтів у січні місяці щодо інших часових інтервалів. Це може означати збільшення купівельної активності 33% клієнтів на новорічні свята, що загалом не суперечить здоровому глузду.

Проведені дослідження показали, що при заданій кількості кластерів результат розбиття з тими ж характерними рисами залишається незмінним, а змінюються лише номери кластерів, що є важливим при інтерпретації результатів.

ВИСНОВКИ ДО ТРЕТЬОГО РОЗДІЛУ

Вказані методи дозволяють ефективно вибудовувати маркетингову стратегію, спираючись на моделі споживчого попиту окремих груп покупців на основі їх намірів щодо вибору товару, що їм підходить. На основі запропонованої методики зручно виконувати кластеризацію, яка не вимагає використання складного інструментарію та програмного забезпечення. Особливо корисним підхід виявиться на дослідження моделей споживчого попиту із використанням даних про стан намірів покупців і з використанням підходів без персоніфікованого обліку клієнтів.

ВИСНОВКИ

Дано обґрунтування концепції еластичності з використанням запропонованої моделі споживчого попиту.

Встановлено що моделі є спрощеним способом, які представляють і пояснюють поведінку споживачів на ринку і дуже необхідні на практиці. Деякі моделі роблять це комплексно, інші підкреслюють динаміку рішень споживача та показують певні напрями можливих досліджень поведінки споживачів. Знання щодо поведінки споживача є цінним джерелом інформації про дії споживачів у процесі купівлі товарів. Моделі дозволяють знизити ризик прийняття підприємством невірних рішень і цим уникнути можливих збитків.

Встановлено що інтелектуальний аналіз даних пропонує зручні інструменти для використання великих обсягів даних і перетворення їх у практичну інформацію чим вигідний для більшості підприємств.

Проаналізовано моделі інтелектуального аналізу даних та можливі варіанти їх використання. Показано сегментацію клієнтської бази, виявлення намірів споживачів та дослідження споживчого попиту на обрані категорії товарів та послуг протягом певного періоду часу з метою оптимізації плану продажу.

Вказані методи дозволяють ефективно вибудовувати маркетингову стратегію, спираючись на моделі споживчого попиту окремих груп покупців на основі їх намірів щодо вибору товару, що їм підходить. На основі запропонованої методики зручно виконувати кластеризацію, яка не вимагає використання складного інструментарію та програмного забезпечення. Особливо корисним підхід виявиться на дослідження моделей споживчого попиту із використанням даних про стан намірів покупців і з використанням підходів без персоніфікованого обліку клієнтів.

Розглянуто питання вибору кількості кластерів, мір близькості об'єктів для їхнього поділу між кластерами, а також питання інтерпретації результатів кластеризації та подальшого аналізу та застосування отриманого розбиття. Реалізована методика дозволяє, маючи в своєму розпорядженні дані про портрет

цільової аудиторії, звертатися до різних сегментів клієнтської бази з певними пропозиціями, тим самим збільшуючи лояльність клієнтів. За допомогою описаних методів і підходів можна ефективно дослідити моделі споживчого попиту на окремі набори товарів залежно від намірів різних груп покупців

ПЕРЕЛІК ПОСИЛАНЬ

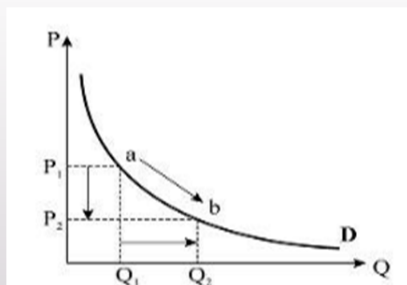
1. Белова Т.Г. МАРКЕТИНГОВІ ДОСЛІДЖЕННЯ КОНСПЕКТ ЛЕКЦІЙ. Електронний ресурс. – Режим доступу. – <https://dspace.nuft.edu.ua/server/api/core/bitstreams/dd90a3a6-1178-40cc-bae9-3e4896466b47/content>
2. Вініченко І. І. Мікроекономіка: Навч. посібник для студ. вузів/І. І. Вініченко, Н. В. Дацій, С. О. Корецька; М-во освіти і науки України, ГУ «ЗІДМУ». – К.: Центр навчальної літератури, 2005. – 265 с.
3. Вікіпедія. Вільна Енциклопедія. Попит. Електронний ресурс. – Режим доступу. – <https://uk.wikipedia.org/wiki/Попит>
4. Гончарова Н.М. Мікроекономіка: Навч. посібник для студ. вищих навч. закл./Міжрегіональна академія управління персоналом. – К.: МАУП, 2005. – 304 с.
5. Горобчук Т. Мікроекономіка: Навчальний посібник/Тетяна Горобчук,; М-во освіти і науки України, Житомирський інженерно-технологічний ун-т. – К.: Центр навчальної літератури, 2004. – 271 с.
6. Інтелектуальний аналіз даних [Електронний ресурс] : навчальний посібник/А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя: ЗНТУ, 2012. – Електронний ресурс. – Режим доступу. – <http://eir.zntu.edu.ua>
7. Калько Д. Р. Методи та задачі інтелектуального аналізу даних. Електронний ресурс. – Режим доступу. – <https://jktod.donnu.edu.ua/article/view/13044>
8. Каменева Н. М., М. В. Косич, О. Ю. Александрова та ін. Мікроекономіка: Навч. посібник – Харків: УкрДУЗТ, 2022. – 305 с
9. КЛАСИЧНІ ТЕХНОЛОГІЇ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ. Електронний ресурс. – Режим доступу. – https://moodle.znu.edu.ua/pluginfile.php/767747/mod_resource/content/1/Тема%207.pdf
10. Кваліфікаційна робота здобувачки вищої освіти Кострубіцька А.В.

11. Колодчак О. М. Інтелектуальний аналіз даних/О. М. Колодчак//Вісник Національного університету «Львівська політехніка». Комп'ютерні системи та мережі. – 2013. – № 773. – С. 49-58. – [Електронний ресурс]. – Режим доступу: http://nbuv.gov.ua/UJRN/VNULPKSM_2013_773_11
12. Макаренко Н.О., Лищенко М.О. Маркетингове ціноутворення. Теоретичні основи: навч. посіб. Буринь: ПП «Буринська районна друкарня», 2020. 129 с
13. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. – Київ. – 2017. – 150 с.
14. Оверченко В. І., Мажак З. М., Софій М. І. Мікроекономіка: Навчальний посібник./за наук. ред. О. Л. Ануфрієвої. – Івано-Франківськ, «Лілея-НВ». – 2015. – 348 с закон попиту. Еластичність
15. Служби Analysis Services – інтелектуальний аналіз даних. Електронний ресурс. – Режим доступу. – <https://learn.microsoft.com>
16. Черняк О.І. Інтелектуальний аналіз даних: підручник/О.І. Черняк, П.В. Захарченко. – К.: Знання, 2014. – 599 с

Інтелектуальний аналіз споживчого попиту для забезпечення максимального прибутку підприємства

Виконала студентка Кострубіцька Анастасія В'ячеславівна

Попит. Закон попиту



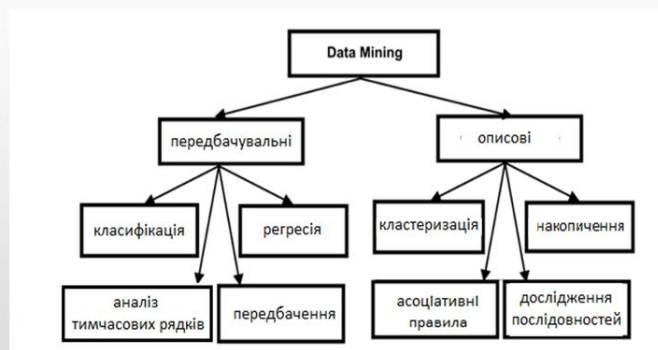
- Попит - це кількість споживачів, які готові придбати товари чи послуги за певною ціною.
- Закон попиту звучить так: при незмінності інших факторів попит на товари та послуги в кількісному вираженні змінюється у зворотній залежності від ціни (назад пропорційний ціні). Отже, закон відбиває залежність попиту ціни.

Попит



- структуру попиту споживачів із доходом s за цінами p можна описати співвідношенням:
- $\varphi_j(s, p) = \varphi_j(sI(p, s))$

Методи Data Mining



Етапи проведення інтелектуального аналізу даних

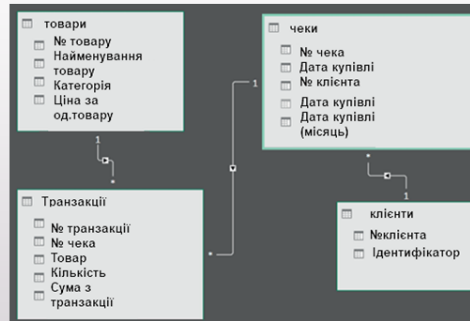


Вибір алгоритму



- SQL Server інтелектуальний аналіз даних дозволяє створювати кілька моделей в одній структурі інтелектуального аналізу даних

Аналіз та сегментування клієнтської бази



- В роботі приведено приклад сегментації клієнтської бази за допомогою електронних таблиць. На початку зроблено вивантаження даних із бази з продажу товарів (з обліку продажу дисконтних карт клієнтів)

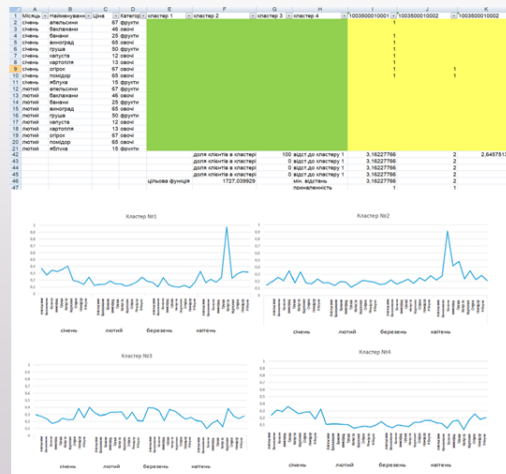
Підготовка даних під кластеризацію

№	А	В	С	Д	Е	Ф
2	Ідентифікатор клієнта					
3	Найменування					
4	1003500010001	1003500010002	1003500010004	1003500010006	1003500010007	
5	Апельсини	2,56			9,84	
6	Бананани			5,12		
7	Банани	3,09		8,71		
8	Виноград	6,61				
9	Груша	3,41				
10	Калюта	14,9		6,28		
11	Картопля	11,03			3,15	
12	Огірок	8,55	2,97		5,52	
13	Помідор	9,25	3,67			
14	Яблука			7,33		1,12
15	Дата купівлі (Місяць)					
16	січ	лют	берез	квіт		
17						
18						

№	А	В	С	Д	Е	Ф	Г	Н	І	К
1	Місяць	Найменування	Ціна	Категорія	кластер 1	кластер 2	кластер 3	кластер 4	1003500010001	1003500010002
2	січень	апельсини	67	фрукти					1	
3	січень	бананани	46	овочі						1
4	січень	банани	25	фрукти					1	
5	січень	виноград	65	овочі					1	
6	січень	груша	50	фрукти					1	
7	січень	калюта	12	овочі					1	
8	січень	картопля	13	овочі					1	
9	січень	огірок	67	овочі					1	
10	січень	помідор	65	овочі					1	
11	січень	яблука	15	фрукти						1
12	лютий	апельсини	67	фрукти						
13	лютий	бананани	46	овочі						
14	лютий	банани	25	фрукти						
15	лютий	виноград	65	овочі						
16	лютий	груша	50	фрукти						

- аналіз проводився з урахуванням даних по продажах за чотири місяці і виявлялися особливості поведінки всього масиву покупців-власників дисконтних карток в той період

Сегментація клієнтської бази



- Показано результати розбиття пар клієнт/угода за кластерами. Частка клієнтів, віднесених до кластера 1, склала 22%, до кластера 2 – 17%, до кластера 3 – 28% та до кластера 4 – 33%

Висновки

- Встановлено що інтелектуальний аналіз даних пропонує зручні інструменти для використання великих обсягів даних і перетворення їх у практичну інформацію чим вигідний для більшості підприємств.
- Проаналізовано моделі інтелектуального аналізу даних та можливі варіанти їх використання.
- Показано сегментацію клієнтської бази, виявлення намірів споживачів та дослідження споживчого попиту на обрані категорії товарів та послуг протягом певного періоду часу з метою оптимізації плану продажу.
- Вказані методи дозволяють ефективно вибудовувати маркетингову стратегію, спираючись на моделі споживчого попиту окремих груп покупців на основі їх намірів щодо вибору товару, що їм підходить.
- На основі запропонованої методики зручно виконувати кластеризацію, яка не вимагає використання складного інструментарію та програмного забезпечення.



Дякую за увагу