

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 004.942:681.3

До захисту допущено
В. о. завідувача кафедри ММСА

О.Л.Тимошук

«___» _____ 2018 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз
на тему: «Система підтримки прийняття рішень для оптимізації прибутку
підприємств роздрібної торгівлі»

Виконала:

студентка ІІ курсу, групи КА-71 мп
Акінфієва Єлизавета Олегівна _____

Керівник: професор кафедри ММСА,
д.т.н., проф. Данилов В.Я _____

Рецензент: професор кафедри ІБ,
д.т.н., проф. Качинський А.Б. _____

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів
без відповідних посилань
Студентка _____

Київ
2018

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)
Спеціальність (спеціалізація) — 124 «Системний аналіз» («Системний аналіз і управління»)

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри ММСА

О. Л. Тимощук

« ___ » _____ 2018 р.

ЗАВДАННЯ

на магістерську дисертацію студенту

Акінфієвій Єлизаветі Олегівні

1. Тема дисертації: « Система підтримки прийняття рішень для оптимізації прибутку підприємств роздрібної торгівлі », науковий керівник дисертації Данилов Валерій Якович, професор, д.т.н., затверджені наказом по університету від «07» листопада 2018 р. № 4121-с

2. Термін подання студентом дисертації:

3. Об'єкт дослідження: прибуток підприємства роздрібної торгівлі як велика система

4. Предмет дослідження: оптимізація прибутку підприємства роздрібної торгівлі, складові прибутку

5. Перелік завдань, які потрібно розробити:

- 1) побудувати структурний аналіз підприємства роздрібної торгівлі;
- 2) аналіз споживчого кошику;
- 3) прогнозування прибутків та продажів, побудова плану розвитку.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

- 1) постановка завдання дослідження

- 2) підготовлені дані
- 3) побудова моделей
- 4) порівняння результатів

7. Орієнтовний перелік публікацій:

- 1) XII Міжнародна науково-технічна конференція «Проблеми інформатизації», Київ, Державний університет телекомунікацій, грудень, 2018р.

8. Дата видачі завдання:

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
		07.09.2018—	
		—26.11.2018	

Студент

Є.О. Акінфієва

Науковий керівник дисертації

В.Я. Данилов

РЕФЕРАТ

Магістерська дисертація: 97 с., 31 рис., 31 табл., 1 додаток і 13 джерел.

Актуальність теми: у зв'язку зі зниженням питомої вартості зберігання даних, зростає обсяг інформації, що накопичують підприємства роздрібної торгівлі, а це спричиняє виникнення завдань, пов'язаних з необхідністю обробки великих масивів даних з метою оптимізації прибутку та зменшення витрат підприємством.

Мета дослідження – розробка структури СППР та її реалізація у вигляді програмного продукту для оптимізації прибутку підприємства, побудови асоціацій, моделювання та прогнозування бізнес-процесів у роздрібній торгівлі.

Об'єкт дослідження – прибуток підприємства роздрібної торгівлі.

Предмет дослідження – оптимізація прибутку, моделі та методи інтелектуального аналізу даних, наприклад: алгоритм Apriori для генерації асоціативних правил, моделі регресійного аналізу для прогнозування часових рядів (AR, ARMA, ARIMA, трендові моделі і т.п.), а також сучасні методи розпізнавання та кластеризації.

Методи дослідження – методи асоціативного аналізу, регресійного аналізу, структурного та кластерного аналізу.

Наукова новизна полягає в розробці системного підходу до розв'язання задачі оптимізації прибутку підприємств роздрібної торгівлі, застосовуванні сучасних ідеологій моделювання та прогнозування, та побудові планів розвитку підприємств в короткій та довгостроковій перспективі.

РОЗДРІБНА ТОРГІВЛЯ, ПРИБУТОК ПІДПРИЄМСТВА,
АСОЦІАТИВНИЙ АНАЛІЗ, РЕГРЕСІЙНИЙ АНАЛІЗ, SWOT АНАЛІЗ,
СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

ABSTRACT

Master's thesis: 97 p., 31 f., 31 tables, 1 appendix, 13 sources.

Subject relevance - in connection with the decrease in the specific value of data storage, the amount of information accumulated by retailers is increasing, this entails the emergence of tasks related to the need to process large amounts of data in order to find new patterns, the establishment and discovery of new knowledge. Therefore, in the master's thesis, the design and implementation of decision-support systems are provided, which ensures the search for associations, modeling and forecasting of key financial indicators of the retail enterprise and promotes the establishment of business strategies.

Purpose of research is to develop a decision support system structure and implement it as a new software product for building associations, modeling and forecasting business processes in retail.

Object of research – profit of the retail enterprise.

Subject of research - models and methods of data mining: Apriori algorithm for generating associative rules, regression analysis model for forecasting of time series: AR, ARMA, ARIMA and trend models.

Research methods are based the theory of associative analysis and time series theory.

The scientific novelty consists in developing a systematic approach to solving the problem of optimizing the profit of retail enterprises, applying modern ideologies of modeling and forecasting, and building plans for enterprise development in the short and long term.

RETAIL, ENTERPRISE PROFIT, ASSOCIATIVE ANALYSIS, REGRESSION ANALYSIS, SWOT ANALYSIS, DECISION SUPPORT SYSTEM

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ	8
ВСТУП	9
РОЗДІЛ 1 СТРУКТУРНИЙ АНАЛІЗ РОЗДРІБНОЇ ТОРГІВЛІ.....	10
1.1 Актуальність задачі	10
1.2 Система підтримки прийняття рішень	11
1.3 Структура підприємств роздрібноЇ торгівлі.....	15
Висновки до розділу.....	19
РОЗДІЛ 2 АНАЛІЗ СПОЖИВЧОГО КОШИКУ	20
2.1 Побудова асоціативних правил.....	20
2.2 Побудова дерев рішень	25
2.3 Кластерний аналіз.....	30
Висновки до розділу.....	35
РОЗДІЛ 3 АНАЛІЗ ТА ПРОГНОЗУВАННЯ РОЗВИТКУ	36
3.1 Аналіз часових рядів	36
3.2 Критерії адекватності регресійної моделі	44
3.3 SWOT – аналіз	48
Висновки до розділу.....	55
РОЗДІЛ 4 ПОБУДОВА СППР.....	56
4.1 Опис СППР	58
4.2 Звантаження вхідних даних	61
4.3 Аналіз прибутку.....	63
Висновки до розділу.....	72
РОЗДІЛ 5 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ	73
5.1 Ідея стартап - проекту.....	73

5.2 Технологічний аудит ідеї проекту	74
5.3 Аналіз ринкових можливостей запуску стартап-проекту	75
5.4 Розробка ринкової стратегії проекту	81
5.5 Розроблення маркетингової програми стартап-проекту	85
Висновки до розділу	89
ВИСНОВКИ	90
ПЕРЕЛІК ПОСИЛАНЬ	91
ДОДАТОК А ЛІСТИНГ КОДУ	93

ПЕРЕЛІК СКОРОЧЕНЬ

СППР – підтримка прийняття рішень;

ОПР – об'єкт, що приймає рішення;

АР – авторегресія;

АРКС – авторегресія ковзного;

АРІКС – авторегресія інтегрованого ковзного середнього.

ВСТУП

Підприємства роздрібної торгівлі розвиваються з кожним роком. Цьому сприяють багато причин, зокрема, збільшення асортименту, завдяки інтернет-магазинам відсутня необхідність витратити час на походи в торгові центри, отримання необхідної інформації про товар можливе миттєво. Щоб залишатися сучасними та збільшувати статки, багато ритейлерів відходять від старих засобів аналізу даних та шукають нові, які б дозволили збільшити прибуток. Отже, виникає питання використання модернізованих методів та підходів аналізу даних продажів, тобто створення суцільних сучасних систем оптимізації прибутку підприємств.

Система підтримки прийняття рішень підприємств роздрібної торгівлі має вирішувати задачі структурного аналізу, для цілісного розуміння орієнтованість ритейлера та кінцеве бачення великої системи, задачу формування товарів на складах, закупки їх необхідної кількості, задачу розміщення товарів, створення акційних пропозицій, а також задачу побудови плану розвитку.

Базовими для такої системи є структурний аналіз, асоціативний аналіз та дерева рішень, кластеризація, аналіз часових рядів та SWOT аналіз. Ці інструменти включають в себе методи класифікації, кластеризації, регресії та асоціативний аналіз, а також аналіз і прогнозування часових рядів. Ці інструменти дозволяють обробляти інформацію для покращення показників діяльності підприємств.

На основі отриманих результатів застосованих моделей, важливо систематизувати вихідні дані для прийняття рішень щодо підвищення прибутку та побудови плану розвитку ритейлеру.

РОЗДІЛ 1 СТРУКТУРНИЙ АНАЛІЗ РОЗДРІБНОЇ ТОРГІВЛІ

1.1 Актуальність задачі

Зі зміною ринкової економіки чи законодавства, змінюються також і умови функціонування підприємств роздрібною торгівлі. Основними причинами занепаду багатьох з них є пізня адаптація, нерозуміння факторів впливу на підвищення або пониження прибутку, використання застарілих методів управління підприємством. Саме тому набувають популярності методи оптимізації прибутку, які поза часом дають можливість враховувати будь-які зміни.

Прибуток є найважливішим показником діяльності підприємства, так як він залежить від усіх складових підприємства: виробництво, якість, реалізація, асортимент, доцільність використання тих чи інших ресурсів, собівартість товарів. [1]

Очевидним варіантом збільшення грошового обігу є підвищення цін на продукцію. Проте, існують певні ризики використання такої стратегії. Перш за все, ризик втрати клієнтів. По-друге, незадоволення дистриб'юторів через нечесну оплату купівлі оптом. Підвищення цін має бути доцільним та обгрунтованим. Іншим варіантом оптимізації прибутку є скорочення витрат, що також необхідно робити з врахуванням всіх плюсів і мінусів, так як наступна стратегія полягає в тому, щоб збільшити якість продукції. Як відомо, за більшу вартість часто продається якісніший товар. Зниження ціни призведе до зростання кількості клієнтів, проте можна втратити частину ринку через ту саму низьку ціну. Створення стратегій та плану розвитку вимагає детального аналізу даних та виявлення прихованих закономірностей. Підприємства збирають і накопичують точні дані про кожну окрему покупку, що призводить до потреби обробляти великі обсяги даних.

Стандартні рішення використовують екранну таблицю як форму подання інформації, що для двомірних таблиць цілком природно, а для багатовимірних - явно недостатньо. Справа в тому, що осередок таблиці тут - це не просто поле з даними, а перетин кількох вимірів, агрегат, під яким ховаються деталізована інформація та алгоритми. У спеціалізованій системі повинен бути реалізований інтерфейс, що дозволяє в залежності від заданого режиму роботи враховувати активність користувача, динамічно визначати статуси вимірювань для багатовимірних змінних, що відображаються в таблиці. Отримана інформація повинна використовуватися в пов'язаних модулях або графічних об'єктах. Таким чином можна організувати міжмодульні переходи, багатовимірне обертання, графічне відображення показників, горизонтальний і структурний аналіз. Керуючою панеллю стає весь екран, що при великій функціональній насиченості вельми корисно. Можливості багатовимірного аналізу при цьому розширюються, але настройка такої функціональності вимагає спеціальної технології для предметної адаптації системи.

1.2 Система підтримки прийняття рішень

Для сучасного рішення збільшення прибутку, необхідно системно підійти до аналізу роботи підприємства, тож створення системи підтримки прийняття рішень для ритейлера вирішить питання прогнозування доходу, визначення обсягу продажів, допоможе визначити необхідні для закупівлі товари та надасть необхідну інформацію для подальшого розвитку, так як СППР це автоматизована система, що допомагає систематизувати та структурувати проблему в складних умовах. [1]

На виході такої системи є набір альтернативних варіантів дій щодо розвитку підприємства з боку особи, що приймає рішення, що дасть змогу збільшити прибуток, покращити якість обслуговування клієнтів тощо. Тут прийняття рішень – це процес пошуку найкращого рішення з множини допустимих рішень. Підтримка прийняття рішень відбувається у вигляді надання інформації ОПР, що складається з виявлення пріоритетів для ОПР, генерацію можливих рішень та альтернатив, допомогу в оцінці проблеми, систематизацію обмежень внутрішнього та зовнішнього середовищ, надання можливих наслідків прийнятих рішень та надання кращого варіанту.

При впровадженні СППР проектування являє собою двонаправлений ітераційний процес. Найбільш продуктивний нам може бути інша послідовність проектування аналітичної системи. Спочатку описується попередня архітектура багатовимірного сховища. Опис ґрунтується на системному аналізі складу і структури інформаційних потоків. У базу даних завантажується близько 10% вихідної інформації і частина історичних даних. На цьому етапі не залучаються предметні аналітики, в постановці завдання беруть участь ІТ-фахівці компанії, для роботи з інформацією вони використовують базову функціональність системи і штатні алгоритми агрегації. Дана фаза проекту дозволяє грубо оцінити розміри майбутнього сховища, складність успадкування історичних даних, час регламентної завантаження. Результатом етапу може бути розгорнутий звіт про ступінь узгодженості інформаційних потоків; можливо, будуть вироблені також технічні пропозиції по розробці алгоритмів очищення даних. Наступна фаза пов'язана з реалізацією алгоритмів очищення і узгодження, завантаженням всього обсягу історичних і оперативних даних. Тут здійснюється аналіз адаптивності користувальницького інтерфейсу до реальних обсягів даних, перевіряються можливості засобів супроводу і розширення предметної області. Коли макети робочих місць аналітика і менеджера вже побудовані, в

проектну групу потрібно активно залучати предметних фахівців. Тільки коли аналітик усвідомлює практичну застосовність прототипу системи для вирішення реальних оперативних завдань, можна сподіватися на його конструктивну участь в проекті. Потім настає фаза розробки предметного технічного завдання на реалізацію в системі бізнес-процесів аналізу та підтримки прийняття управлінських рішень. Тільки тут комплексна проектна група остаточно фіксує реквізитний складу об'єктів, логічні зв'язки, алгоритми і методики, що дозволяють обслуговувати бізнес-процеси управління компанією. З цього моменту починається класичне проектування системи зверху вниз. Але і тут дається взнаки специфіка СППР - перш за все вона проявляється в масштабності і логічної складності бізнес-об'єктів управління. Реквізитний складу бізнес-об'єктів може налічувати тисячі і десятки тисяч найменувань. З урахуванням системних сутностей на кожен предметний реквізит в середньому припадає близько десяти об'єктів. Ручна технологія проектування сховища вкрай неефективна, навіть при наявності сучасних графічних засобів. Штатний інструментарій повинен бути дооснащена спеціалізованим автоматичним генератором бази даних. Як джерело генерації тут можна використовувати файл Excel, що описує реквізитний склад і логічні зв'язки бізнес-об'єктів. На базі цього файлу генератор в автоматичному режимі створює або модифікує інформаційне сховище. Така технологія дозволяє распараллелить опис метаданих - цей етап часто стає вузьким місцем корпоративного проекту. Крім того, в якості "сухого залишку" від роботи архітекторів сховища залишається ще і зрозумілий документ Excel, який дозволить службам супроводу розширювати і модифікувати предметну область, не опускаючись до рівня фізичних сутностей. Певну технологічну складність представляє також створення і супровід багатовимірних агрегатів, методик їх розрахунку і моделювання. Кількість показників у великих корпоративних проектах при цьому теж дуже

велике, а зміна моделей досить динамічно, тому стандартний підхід до моделювання виявляється непридатним. І тут можливий варіант використання формалізованих файлів Excel. Ці файли обслуговуються в розрахованому на багато користувачів режимі прикладними адміністраторами і згодом "накочуються" на метадані системи. Зрозумілі для служб експлуатації документи будуть виступати і як опис предметного вирішення, і як джерела для модифікації метаданих.[2]

На етапі постановки задачі, відбувається аналіз поставлених проблем та встановлення кінцевої мети та цілей рішення. Необхідно визначити напрям рішення і відкинути ті підходи, що не відповідають поставленій меті. Етап формування рішень відбувається на основі критеріїв прийняття рішень. Також він складається з пошуку та створення множини допустимих альтернатив рішення. На етапі вибору рішення проводиться оцінка отриманих альтернатив та вибір найкращої альтернативи. [1]

СППР володіє наступними основними характеристиками:

- використовує дані і моделі;
- допомагає в прийнятті рішень для неструктурованих і слабо структурованих завдань;
- допомагає прийняти, але не замінює способу генерації рішень менеджерами;
- збільшує ефективність рішень.
- СППР, що володіє ідеальними характеристиками:
 - працює зі слабо структурованими рішеннями;
 - призначена для ОПР різного рівня підготовки;
 - може використовуватись для групового та особистого використання;
 - підтримує наступні фази процесу прийняття рішення: інтелектуальну, проектування і вибір;

- підтримує різні способи та методи вирішення, що допомагає при вирішенні задач групою ОПР;
- може адаптується до різних змін підприємства та його оточення;
- проста у використанні і модифікації;
- підтримує як взаємозалежні, паралельні, так і послідовні рішення;
- підтримує еволюційне використання і легко адаптується до мінливих вимог;
- використовує накопичені знання.

1.3 Структура підприємств роздрібної торгівлі

Відповідно до стандартного розуміння роботи ритейлера, він складається з багатьох компонент, які впливають на життєдіяльність магазину. Перш за все, виділимо основні складові: постачальники, працівники, товари, розміщення торгівельних точок, оренда приміщень, обладнання, якість продукції, законодавство та власне виробництво, що зображено на рис. 1.1.

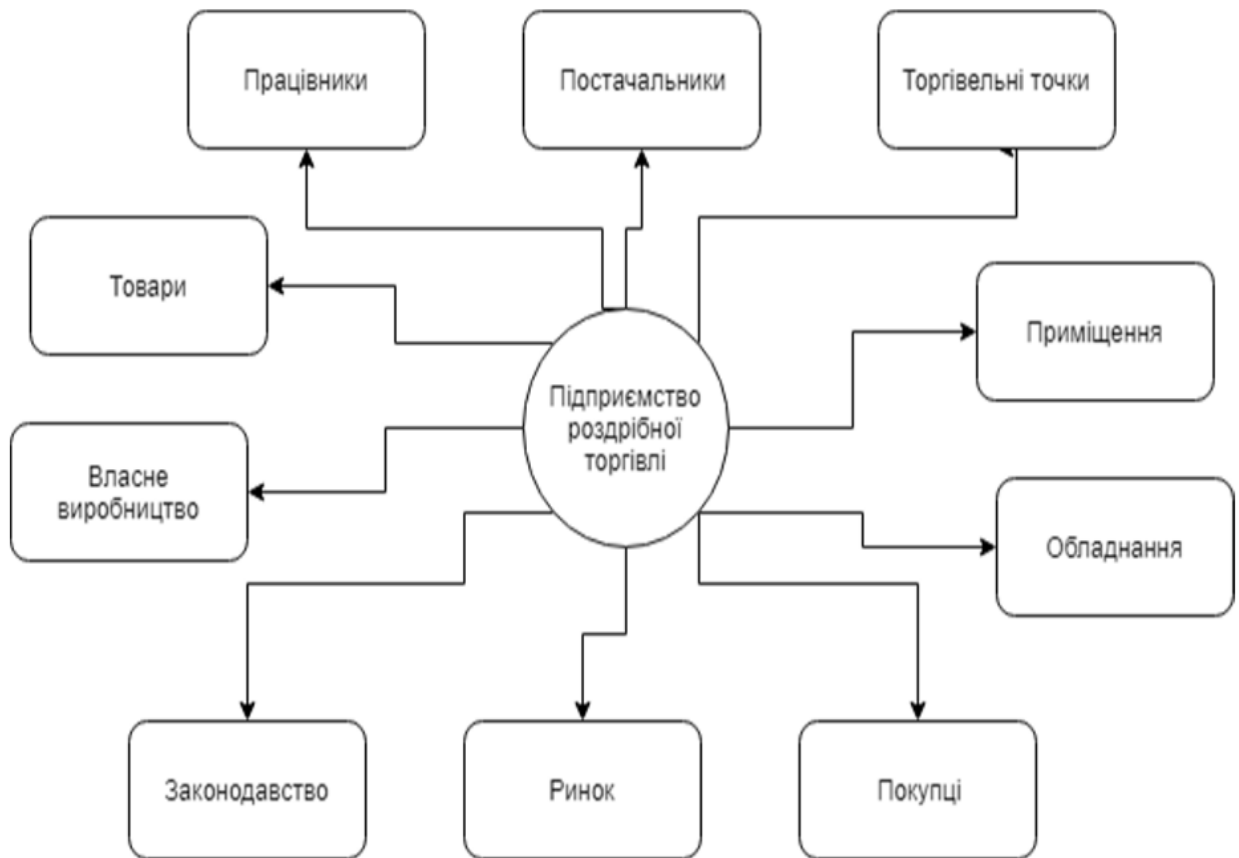


Рисунок 1.1 – Структура підприємств роздрібно́ї торгівлі

Кожен складовий елемент, в свою чергу, складається з більш дрібних компонент.

Розглянемо детальніше задачу закупівлі та продажу товару. Споживчі товари можна поділити на продовольчі, непроовольчі та медичні, в залежності від специфіки роздріного магазину. Поділ товарів на групи та компоненти зображено на рис. 1.2, рис. 1.3, рис. 1.5.



Рисунок 1.2 – Поділ товарів на групи



Рисунок 1.3 – Поділ продовольчих товарів на компоненти

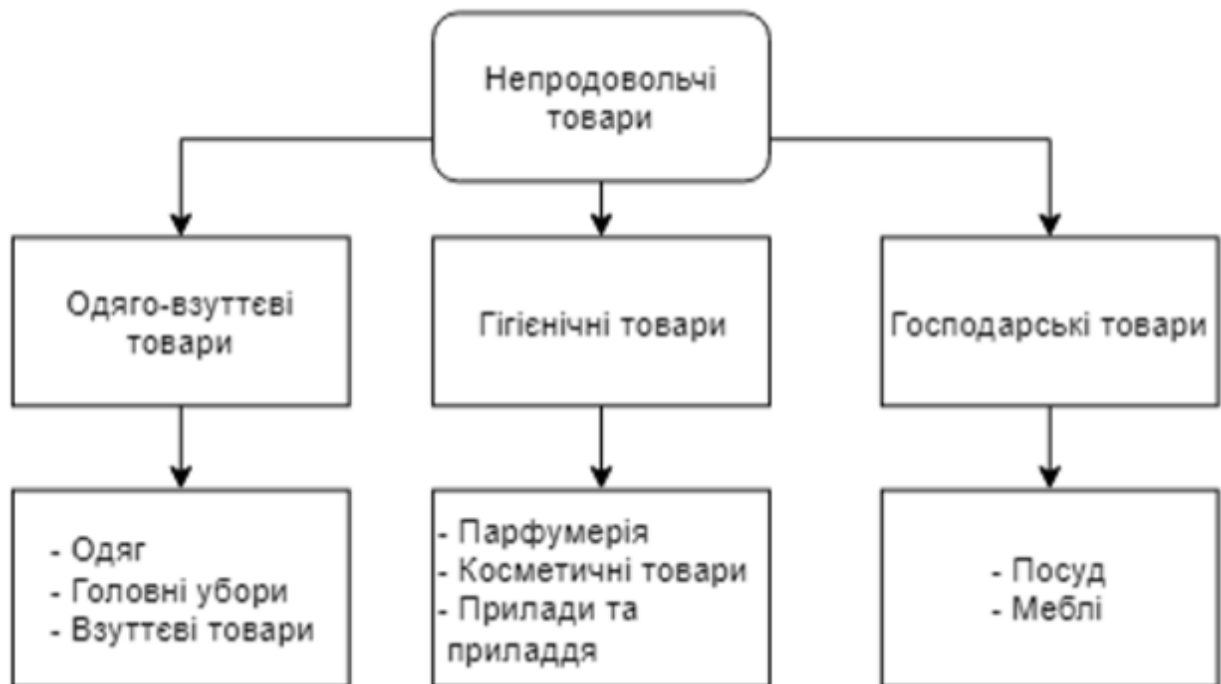


Рисунок 1.4 – Поділ непродовольчих товарів на компоненти

В свою чергу, підприємства також збирають дані щодо кожної транзакції в магазині, що складається з номеру транзакції, проданого товару, категорії товару, групи, ціни продажу, ціни покупки, кількість товару, дату та ім'я або номер клієнта, що зображено на рис. 1.5. Саме ці дані далі аналізуються за допомогою різних методів та підходів. Дану задачу можна розділити на декілька, адже необхідно обробити велику кількість даних, застосувати доцільні алгоритми обробки, після чого відобразити дані для наглядного результату.

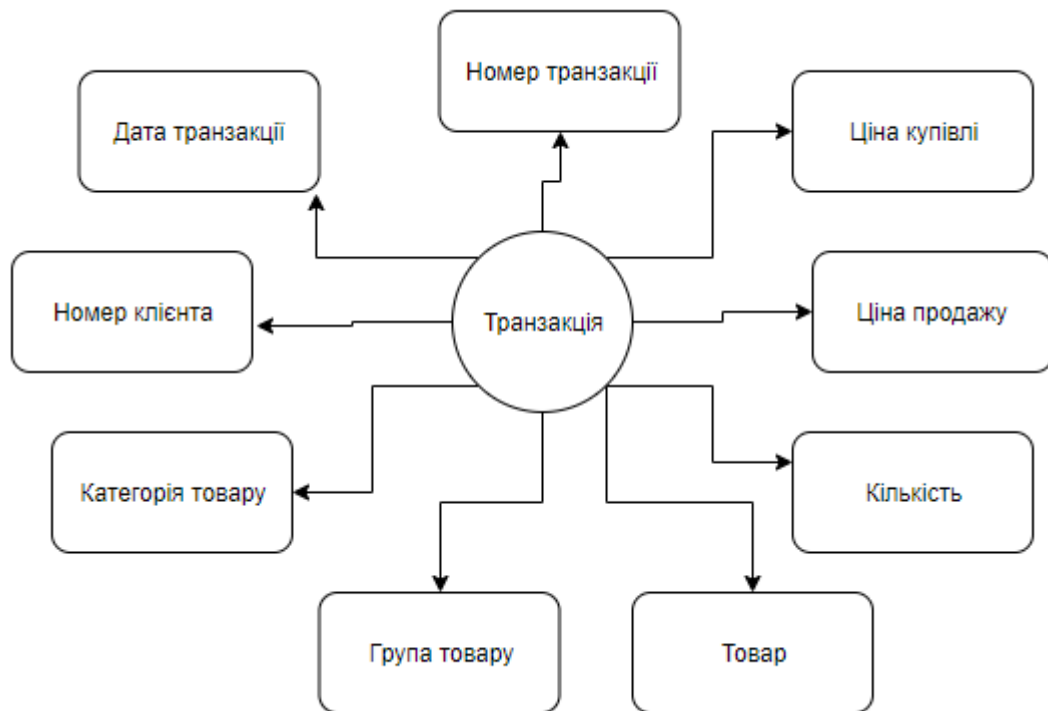


Рисунок 1.5 – Структура транзакцій

Висновки до розділу

Провівши структурний аналіз підприємства роздрібної торгівлі, можна зробити висновки щодо впливу різних компонент на прибутковість підприємства, наприклад, аналіз покупок клієнтів та товари, та запланувати подальші кроки для оптимізації роботи організації.

РОЗДІЛ 2 АНАЛІЗ СПОЖИВЧОГО КОШИКУ

2.1 Побудова асоціативних правил

Для Асоціативні правила являють собою механізм знаходження логічних закономірностей між пов'язаними елементами (подіями або об'єктами). [3]

Деякі області, де асоціативні правила допомагають розв'язувати задачі:

- Аналіз ринкових кошиків - це найбільш характерний приклад видобутку асоціації. Дані збираються за допомогою сканерів штрих-кодів у більшості супермаркетів. Ця база даних, відома як база даних "база ринку", складається з великої кількості записів про минулі транзакції. Єдиний рекорд показує всі товари, придбані клієнтом в одному продажі. Знаючи, які групи нахилиються до того, який набір предметів дає цим магазинам свободу налаштувати розташування магазину та каталог магазинів, вони оптимально ставляться один до одного.
- Медичний діагноз - правила асоціації в медичній діагностиці можуть бути корисними для надання допомоги лікарям у лікуванні хворих. Діагноз не є простим процесом і містить ряд помилок, які можуть призвести до ненадійних кінцевих результатів. Використовуючи вирішальне правило реляційної асоціації, ми можемо визначити ймовірність виникнення хвороби щодо різних чинників та симптомів. Далі, використовуючи методи навчання, цей інтерфейс може бути розширений шляхом додавання нових симптомів та визначення зв'язків між новими ознаками та відповідними захворюваннями.

- Дані перепису - кожному уряду є тони даних перепису. Ці дані можуть бути використані для планування ефективних державних послуг (освіти, охорони здоров'я, транспорту), а також допомагати державним підприємствам (для створення нових заводів, торгових центрів та навіть окремих продуктів для маркетингу). Така практика використання гіпермаркетів та дослідження даних має величезний потенціал для підтримки здорової державної політики та ефективного функціонування демократичного суспільства.
- Білкова послідовність - послідовності, що складаються з двадцяти типів амінокислот. Кожен білок несе унікальну 3D структуру, яка залежить від послідовності цих амінокислот. Невелика зміна послідовності може призвести до зміни структури, яка може змінити функціонування білка. Ця залежність функціонування білка на його амінокислотну послідовність була предметом великих досліджень. Раніше вважалося, що ці послідовності є випадковими, але тепер вважається, що їх немає. Нітін Гупта, Нітин Мангал, Камал Тіварі та Пабітра Мітра розшифрували характер асоціацій між різними амінокислотами, присутніми в протейні. Знання та розуміння цих правил асоціації стануть надзвичайно корисними при синтезі штучних білків.[9]

Наша задача відноситься до аналізу ринкових кошиків. Нехай $A = \{a_1, a_2, a_3, \dots, a_n\}$ – кінцевий список безліч унікальних елементів (list of items). З цих компонентів може бути складено безліч наборів T (sets of items), тобто $T \subseteq A$.

Асоціативні правила $A \rightarrow T$ мають такий вигляд: якщо $\langle \text{умова} \rangle$ то $\langle \text{результат} \rangle$, де на відміну від дерев класифікації, $\langle \text{умова} \rangle$ - НЕ логічне вираз, а набір об'єктів з безлічі A , з якими пов'язані (асоційовані) об'єкти того ж безлічі, включені в $\langle \text{результат} \rangle$ даного правила. Наприклад, асоціативне

правило якщо (смородина, тля) то (мурахи) означає, що якщо на куці смородини зустрілася тля, то шукай поблизу і мурах.[4]

Поняття "вид елемента ак" легко може бути узагальнене на ту чи іншу його категорію або дійсне значення, тобто концепція асоціативного аналізу може бути застосована для комбінацій будь-яких змінних. Наприклад, при прогнозуванні погоди одне з асоціативних правил може виглядати так: 'напрямок вітру' = NNW -> 'завтра буде дощ' = TRUE.

Виділяють три види правил:

- корисні правила, що містять дійсну інформацію, яка раніше була невідома, але має логічне пояснення;
- тривіальні правила, що містять дійсну та легко зрозумілу інформацію, яка відобразить відомі закони в досліджуваній області, і тому не приносять жодної користі;
- незрозумілі правила, що містять інформацію, яка не може бути пояснена (такі правила або отримують на основі аномальних вихідних даних, або вони містять глибоко приховані закономірності, і тому для інтерпретації незрозумілих правил потрібен додатковий аналіз).

Пошук асоціативних правил зазвичай виконують в два етапи:

- в пулі наявних ознак А знаходять найбільш часто зустрічаються комбінації елементів Т;
- з цих знайдених найбільш часто зустрічаються наборів формують асоціативні правила.

Для оцінки корисності і продуктивності перебираються правил використовуються різні частотні критерії, що аналізують зустрічальність кандидата в масиві експериментальних даних. Найважливішими з них є підтримка (support) і достовірність (confidence). Правило $A \rightarrow T$ має підтримку s , якщо воно справедливо для $s\%$ взятих в аналіз випадків:

$$\text{support}(A \rightarrow T) = P(A \cup T) \quad (2.1)$$

Достовірність правила показує, наскільки ймовірним є те, що з наявності в даному випадку умовної частини правила слід наявність заключній його частині (тобто з A слід T):

$$\text{confidence}(A \rightarrow T) = P(A \cup T) / P(A) = \text{support}(A \rightarrow T) / \text{support}(A). \quad (2.2)$$

Алгоритми пошуку асоціативних правил відбирають тих кандидатів, у яких підтримка і достовірність вище деяких наперед заданих порогів: minsupport і minconfidence . Якщо підтримка має велике значення, то алгоритми знаходять правила, добре відомі аналітику або настільки очевидні, що немає ніякого сенсу проводити такий аналіз. Більшість цікавих правил знаходять саме при низькому значенні порогу підтримки. З іншого боку, низьке значення minsupport веде до генерації величезної кількості варіантів, що вимагає істотних обчислювальних ресурсів або веде до генерації статистично необґрунтованих правил.

У пакеті *arules* для R використовуються і інші показники - підйомна сила, або ліфт (*lift*), яка показує, наскільки підвищується ймовірність знаходження T в уже згадуваному випадку, якщо в ньому вже є A :

$$\text{lift}(A \rightarrow T) = \text{confidence}(A \rightarrow T) / \text{support}(T) \quad (2.3)$$

і посилення (*leverage*), яке відображає, наскільки цікавою може бути більш висока частота A і T в поєднанні з більш низьким підйомом:

$$\text{leverage}(A \rightarrow T) = \text{support}(A \rightarrow T) - \text{support}(A) \times \text{support}(T) \quad (2.4)$$

Перший алгоритм пошуку асоціативних правил був розроблений в 1993 р співробітниками дослідницького центру ІВМ, що відразу порушило інтерес до цього напрямку. Щороку з'являлося кілька нових алгоритмів (DHP, Partition, DIC та ін.), з яких найбільш відомим залишився алгоритм "Apriori" (Agrawal, Srikant, 1994).

Пакет *arules* дозволяє знаходити часто зустрічаються поєднання елементів в даних (*frequent itemsets*) і відбирати асоціативні правила, забезпечуючи інтерфейс до модулів на мові С, які реалізують алгоритми "Apriori" і "Eclat". Так як зазвичай обробляються великі безлічі наборів і правил, то для зменшення обсягів необхідної пам'яті пакет містить розвинений інструментарій перетворення розріджених вхідних матриць в компактні набори транзакцій [5]

Для реалізації роботи з алгоритмами виділення асоціацій в *arules* реалізовані спеціальні типи даних, які стосуються об'єктах трьох класів: вхідний масив транзакцій (*transactions*) і на виході - часто зустрічаються фрагменти даних (*itemsets*) і правила (*rules*).

Об'єкти класу *transactions* представляють собою спеціально організовані бінарні матриці з рядками-наборами і стовпцями-ознаками, що містять значення елемента 1, якщо відповідний ознака є в транзакції, і 0, якщо він відсутній. Залежно від типу даних і способу їх завантаження, ці об'єкти можуть мати різні способи організації та склад додаткових слотів. Зокрема, підклас *itemMatrix* є одночасно засобом представлення розріджених матриць з використанням функціоналу пакету *Matrix*. Іншим способом формування примірників класу *transactions* є завантаження даних з файлу функцією `read.transactions ()`.

2.2 Побудова дерев рішень

Алгоритми навчання з учителем, засновані на використанні дерев рішень (англ. "Decision trees": "дерева прийняття рішень", "дерева класифікації і регресії" (від "regression and classification trees, CART"), "вирішальні дерева" та ін.) надзвичайно популярні. Ця популярність обумовлена декількома причинами:

- Дерева рішень дозволяють отримувати дуже легко інтерпретуються моделі, що представляють собою набір правил виду "якщо ..., то ...". Інтерпретація полегшується в тому числі за рахунок можливості подати ці правила у вигляді наочної деревовидної структури.
- В силу свого пристрою дерева рішень дозволяють працювати зі змінними будь-якого типу без необхідності будь-якої попередньої підготовки цих змінних для введення в модель (наприклад, логарифмування, перетворення категоріальних змінних в індикаторні, і т.п.).
- Досліднику немає необхідності в явному вигляді задавати форму взаємозв'язку між відгуком і предикторами, як це, наприклад, відбувається у випадку зі звичайними регресійній моделі. Це виявляється особливо корисним при роботі з великими обсягами даних, про властивості яких мало що відомо.
- Дерева рішень, по суті, автоматично виконують відбір інформативних предикторів і враховують можливі взаємодії між ними. Це, зокрема, робить дерева рішень корисним інструментом розвідувального аналізу даних.
- Дерева рішень можна ефективно застосовувати до даних з пропущеними значеннями, що дуже корисно при вирішенні

практичних завдань, де наявність пропущених значень - це, скоріше, правило, ніж виняток.

- Дерева рішень однаково добре застосовні як до кількісних, так і до якісних залежним змінним.

Звичайно, як і будь-який інший інструмент, дерева рішень не позбавлені недоліків. Найбільш важливими з них є наступні:

- Нестабільність (дисперсія) моделі. Мова тут йде про те, що навіть незначні зміни в навчальних даних можуть суттєво змінювати структуру одержуваної моделі. Це, в свою чергу, може призводити до зовсім іншої інтерпретації моделі.
- Відносно невисока точність прогнозів. Це обумовлено тим, що дерева рішень розбивають простір предикторів на прямокутні області. Якщо взаємозв'язок між відгуком і предикторами неможливо адекватно представити за допомогою таких областей, то відповідна модель буде допускати значних помилок при прогнозі відгуку для нових спостережень.

Дерева прийняття рішень будуються на основі розбиття незалежних змінних і мають три ключові параметри: `minbucket`, `minsplit` і `cp`. [6]

`minbucket` - мінімальна кількість спостережень, яке повинно міститися в кожному аркуші дерева. При великому значенні `minbucket` модель буде занадто загальною, що негативно позначиться на точності. Якщо `minbucket` буде занадто низьким виникає ризик перенавчання, модель буде показувати відмінні результати при навчанні але виявиться непридатною для роботи з реальними даними.

`minsplit` - мінімальна кількість спостережень, які повинні міститися в вершині, щоб спробувати розбити її на кілька вершин.

`cp` (`complexity parameter`) - це метрика, яка зупиняє розгалуження дерева, коли воно вже не покращує показники моделі.

На рис. 2.1 зображений приклад поділу категорій товарів:

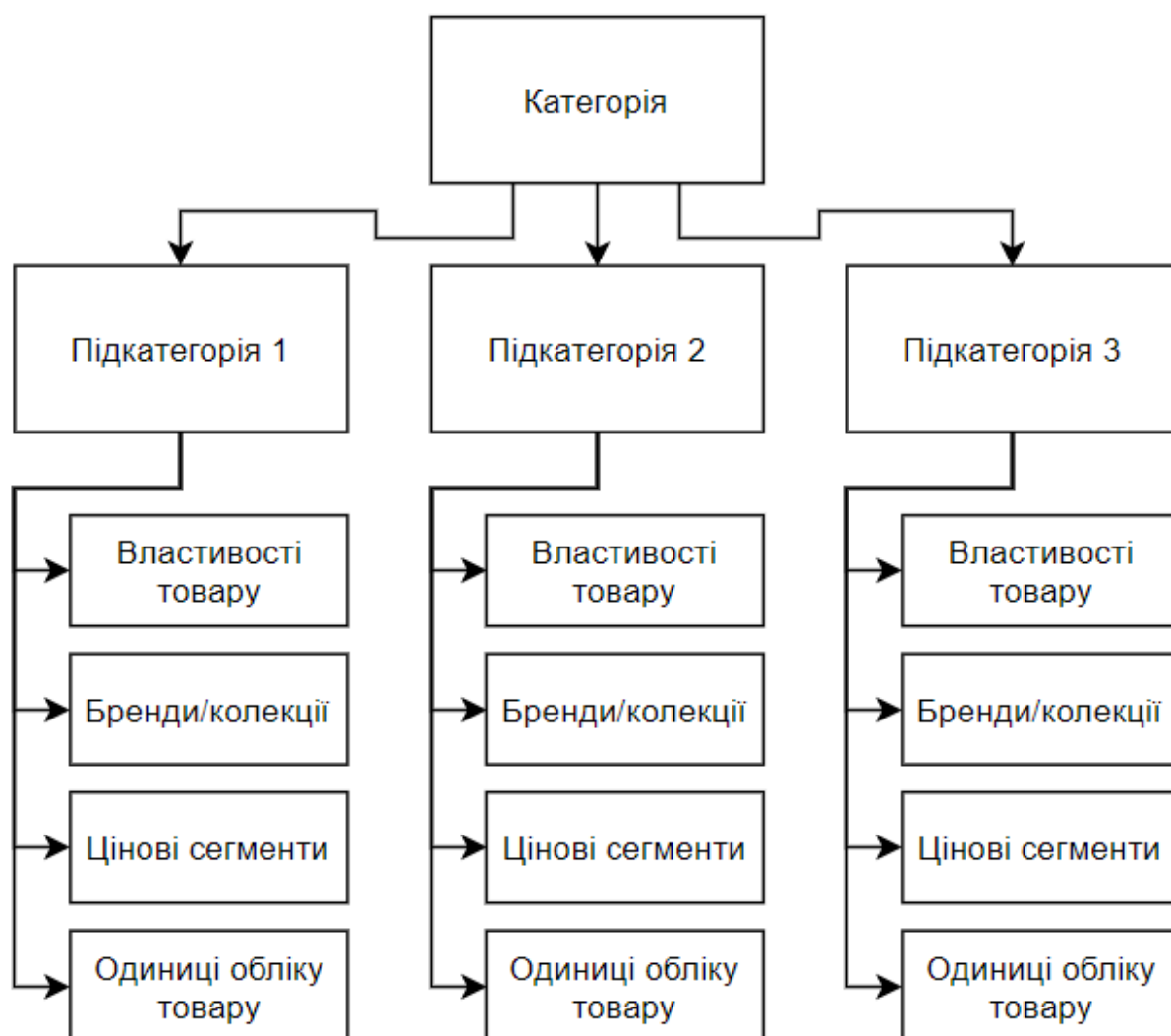


Рисунок 2.1 – поділ товарів по категоріям

Властивості товару виділяються, засновуючих на характеристиках товару:

- покупця або кінцевого користувача (корм для собак або кішки; дитяче, жіноче, чоловіче; джинси для чоловіків, жінок або унісекс; косметика для молодої, дорослої шкіри);
- форма і фасон (роликові, тверді дезодоранти або спреї; брюки стрейч, класика, клеш; туалетна вода, парфумерна вода, парфуми);

- колір (телевізор білий, чорний, срібристий; рубашка в полоску);
- склад (кава з кофеїном або без кофеїну, сметана 15, 20, 30% -ой жирності);
- розмір та вік (костюм S, M, L, XL, або 44, 46, 48, 50; комбінезон на дитину 18 місяців, 24 місяця, 2-3 роки, 3-4 роки, керамічна плитка 15, 20 см, 30 см);
- фактура та матеріал (шпалери вологостійкі, миючі, звичайні; крісла з тканини, шкіри; ванни сталеві або акрилові; кофти - замша, трикотаж, фліс; джинси денієм або стрейч);
- запахи та смаки (парфуми квіткові, мускусні; вина білі, червоні, рожеві);
- можливі інші властивості, наявні у товарі.

Бренди та колекції:

- по торговим маркам (порошки; косметика), по виробнику (телевізори Sony, Panasonic);
- по країні виробництва (плитка керамічна з Іспанії, Італії, Білорусії; вина Грузії, Франції, Чилі);
- по колекціям и серіям.

Цінові сегменти або стиль:

- рівень цін (одяг: вечірній, повсяденний, класика; парфуми: ексклюзив, преміум);
- стилі (в одязі - "діловий", "гламур", "класика", "молодіжний", "щоденний"; меблі в стилі ретро, модерн, високотехнологічний).

Одиниці обліку:

- розмір і форма упаковки (тканина в рулонах або відрізах, ущільнювач в бухтах або в блістерній упаковці);
- міри об'єму, довжини, вага та т. п. - кілограми, метри, літри (пиво 0,3, 0,5, 1,5 л, торти 0,5 кг або 1 кг, фарба в банках по 10, 25, 50 кг).

Представлення покупця є головним фактором, що дозволяє організувати категорію в просторі магазину так, щоб полегшити вибір і прийняття рішення про покупку.

Усі рівні категорії, які розглядалися вище, можуть вибирати покупця в різній послідовності.

В одному випадку це буде: товар, торгова марка, виробник або країна виробництва, властивості та характеристики товару, вага, об'єм або вид упаковки.[7]

В іншому магазині ця категорія буде покупцем вибрати за ознакою: властивості, торгова марка, упаковка, ціна.

Категорія не зміниться, але різною буде послідовність факторів вибору. Наприклад, для покупців з низьким рівнем доходу під час вибору першим буде питання про ціну, а потім - про торговельну марку. А покупцем із середнім рівнем доходу спочатку зацікавиться торговою маркою, а вже потім - ціною.[9]

Чим нижче дохід покупця, тим більше вірогідно, що ціна буде вагомішою при виборі товару, а інші властивості (колір, марка, матеріал, смак і т. п.) виявляться на другому плані.

Чим вище дохід покупця, тим складніше вибір - його визначає не ціна, а інші компоненти (колір, смак, матеріал). Але і ціна в даному випадку може також бути критерієм первинного вибору - за принципом "чим дорожче, тим краще" (наприклад, при виборі коньяку, ікри, парфумерії, хутрових виробів, дорогоцінних каменів - предметів, що визначають статус власника).

2.3 Кластерний аналіз

Кластеризація (або кластерний аналіз) - це задача розбиття множини об'єктів на групи, які називаються кластерами. Середині кожної групи повинні виявитися «схожі» об'єкти, а об'єкти різних груп повинні бути якомога більш відмінні. Головна відмінність кластеризації від класифікації полягає в тому, що перелік груп чітко не заданий і визначається в процесі роботи алгоритму.

Застосування кластерного аналізу в загальному вигляді зводиться до наступних етапів:

- Вибір вибірки об'єктів для кластеризації.
- Визначення безлічі змінних, за якими будуть оцінюватися об'єкти у вибірці. При необхідності - нормалізація значень змінних.
- Обчислення значень міри схожості між об'єктами.
- Застосування методу кластерного аналізу для створення груп схожих об'єктів (кластерів).
- Представлення результатів аналізу.
- Після отримання та аналізу результатів можливе корегування обраної метрики і методу кластеризації до отримання оптимального результату.

Отже, як же визначати «схожість» об'єктів? Для початку потрібно скласти вектор характеристик для кожного об'єкта - як правило, це набір числових значень, наприклад, зростання-вага людини. Однак існують також алгоритми, що працюють з якісними (т.зв. категорійними) характеристиками.

Після того, як ми визначили вектор характеристик, можна провести нормалізацію, щоб всі компоненти давали однаковий внесок при розрахунку «відстані». У процесі нормалізації всі значення приводяться до деякого діапазону, наприклад, $[-1, -1]$ або $[0, 1]$.

Об'єднання або метод деревовидної кластеризації використовується при формуванні кластерів несходства або відстані між об'єктами. Ці відстані можуть визначатися в одновимірному або багатовимірному просторі. Наприклад, якщо ви повинні кластеризувати типи їжі в кафе, то можете взяти до уваги кількість містяться в ній калорій, ціну, суб'єктивну оцінку смаку і т.д. Найбільш прямий шлях обчислення відстаней між об'єктами в багатовимірному просторі полягає в обчисленні евклідових відстаней. Якщо ви маєте дво- або тривимірне простір, то цей захід є реальним геометричним відстанню між об'єктами в просторі (як ніби відстані між об'єктами виміряні рулеткою). Однак алгоритм об'єднання не «дбає» про те, чи є "надані" для цієї відстані справжніми або деякими іншими похідними заходами відстані, що більш значуще для дослідника; і завданням дослідників є підібрати правильний метод для специфічних застосувань.[10]

Евклідова відстань - це, мабуть, найбільш загальний тип відстані. Воно просто є геометричним відстанню в багатовимірному просторі і обчислюється таким чином:

$$r(x, y) = \left\{ \sum_i (x_i - y_i)^2 \right\}^{\frac{1}{2}} \quad (2.5)$$

Евклідова відстань (і її квадрат) обчислюється по вихідним, а не за стандартизованими даними. Це звичайний спосіб його обчислення, який має певні переваги (наприклад, відстань між двома об'єктами не змінюється при введенні в аналіз нового об'єкта, який може виявитися викидом). Проте, на відстані можуть сильно впливати відмінності між осями, за координатами яких обчислюються ці відстані. Наприклад, якщо одна з осей виміряна в сантиметрах, а ви потім перекладете її в міліметри (множачи значення на 10), то остаточне евклідова відстань (або квадрат евклідова відстані), що

обчислюється за координатами, сильно зміниться, і, як наслідок, результати кластерного аналізу можуть сильно відрізнятись від попередніх.

Квадрат евклидової відстані - іноді може виникнути бажання звести в квадрат стандартне евклідова відстань, щоб надати великі ваги більш віддаленим один від одного об'єктів. Це відстань обчислюється таким чином (див. Також зауваження в попередньому пункті):

$$r(x, y) = \sum_i (x_i - y_i)^2 \quad (2.6)$$

Відстань міських кварталів (Манхеттенський відстань) - це відстань є просто середнім різниць по координатах. У більшості випадків ця міра відстані приводить до таких же результатів, як і для звичайного відстані Евкліда. Однак зазначимо, що для цього заходу вплив окремих великих різниць (викидів) зменшується (так як вони не зводяться в квадрат). Манхеттенський відстань обчислюється за формулою:

$$r(x, y) = \sum_i |x_i - y_i| \quad (2.7)$$

Відстань Чебишева - це відстань може виявитися корисним, коли бажать визначити два об'єкти як "різні", якщо вони розрізняються за якоюсь однією координаті (будь-яким одним виміром). Відстань Чебишева обчислюється за формулою:

$$r(x, y) = \max |x_i - y_i| \quad (2.8)$$

Ступінь відстань - іноді бажать прогресивно збільшити або зменшити вагу, що відноситься до розмірності, для якої відповідні об'єкти сильно

відрізняються. Це може бути досягнуто з використанням статичного відстані. Статичне відстань обчислюється за формулою:

$$r(x, y) = \{\sum_i (x_i - y_i)^p\}^{\frac{1}{r}} \quad (2.9)$$

де r і p - параметри, що визначаються користувачем. Кілька прикладів обчислень можуть показати, як "працює" ця міра. Параметр p відповідальний за поступове зважування різниць за окремими координатами, параметр r відповідальний за прогресивне зважування великих відстаней між об'єктами. Якщо обидва параметри - r і p , рівні двом, то це відстань збігається з відстанню Евкліда.

Відсоток незгоди - цей захід використовується в тих випадках, коли дані є категоріальним. Це відстань обчислюється за формулою:

$$r(x, y) = \text{count}(x_i \neq y_i) / i, \quad (2.10)$$

По правилам об'єднання, на першому кроці, коли кожен об'єкт являє собою окремий кластер, відстані між цими об'єктами визначаються обраною мірою. Однак коли зв'язуються разом декілька об'єктів, виникає питання, як слід визначити відстані між кластерами. Іншими словами, необхідно правило об'єднання або зв'язку для двох кластерів. Тут є різні можливості: наприклад, ви можете пов'язати два кластери разом, коли будь-які два об'єкти в двох кластерах ближче один до одного, ніж відповідне відстань зв'язку. Іншими словами, ви використовуєте "правило найближчого сусіда" для визначення відстані між кластерами; цей метод називається методом одиночній зв'язку. Це правило буде "волокнисті" кластери, тобто кластери, "зчеплені разом" тільки окремими елементами, випадково опинилися ближче інших один до одного. Як альтернативу ви можете використовувати сусідів в кластерах, які

знаходяться далі за всіх інших пар об'єктів один від одного. Цей метод називається метод повного зв'язку. Існує також безліч інших методів об'єднання кластерів, подібних до тих, що були розглянуті.

Одиночний зв'язок (метод найближчого сусіда) - як було описано вище, в цьому методі відстань між двома кластерами визначається відстанню між двома найбільш близькими об'єктами (найближчими сусідами) в різних кластерах. Це правило повинне, в даному разі, нанизувати об'єкти разом для формування кластерів, і результуючі кластери мають тенденцію бути представленими довгими "ланцюжками".

Повний зв'язок (метод найбільш віддалених сусідів) - у цьому методі відстані між кластерами визначаються найбільшою відстанню між будь-якими двома об'єктами в різних кластерах (тобто "найбільш віддаленими сусідами"). Цей метод зазвичай працює дуже добре, коли об'єкти відбуваються насправді з реально різних "гаїв". Якщо ж кластери мають в деякому роді подовжену форму, то цей метод непридатний.

Незважене попарне середнє - у цьому методі відстань між двома різними кластерами обчислюється як середня відстань між усіма парами об'єктів в них. Метод ефективний, коли об'єкти в дійсності формують різні "гаї", однак він працює однаково добре і в випадках протяжних ("ланцюжка" типу) кластерів. [11]

Виважена попарне середнє - метод ідентичний методу невиваженого попарного середнього, за винятком того, що при обчисленнях розмір відповідних кластерів (тобто число об'єктів, що містяться в них) використовується в якості вагового коефіцієнта. Тому пропонований метод повинен бути використаний (скоріше навіть, ніж попередній), коли передбачаються нерівні розміри кластерів.

Незважений центроїдний метод - у цьому методі відстань між двома кластерами визначається як відстань між їх центрами тяжкості

Зважений центроїдний метод (медіана) - цей метод ідентичний попередньому, за винятком того, що при обчисленнях використовуються ваги для обліку різниці між розмірами кластерів (тобто числами об'єктів в них). Тому, якщо є (або підозрюються) значні відмінності в розмірах кластерів, цей метод виявляється переважно попереднього. [12]

Метод Варда - цей метод відрізняється від всіх інших методів, оскільки він використовує методи дисперсійного аналізу для оцінки відстаней між кластерами. Метод мінімізує суму квадратів (SS) для будь-яких двох (гіпотетичних) кластерів, які можуть бути сформовані на кожному кроці. [12]

Висновки до розділу

Провівши аналіз споживчого кошику за допомогою побудови асоціативних правил, дерев рішень та кластерного аналізу, підприємство різницевої торгівлі матиме змогу оптимізувати розміщення товарів в лавках магазину, створювати акційні пропозиції та створювати рекламу товарів таким чином, щоб клієнти купували більше та швидше знаходили необхідні товари.

РОЗДІЛ 3 АНАЛІЗ ТА ПРОГНОЗУВАННЯ РОЗВИТКУ

3.1 Аналіз часових рядів

Як і більшість інших видів аналізу, аналіз часових рядів передбачає, що дані містять систематичну складову (зазвичай включає кілька компонент) і випадковий шум (помилку), який ускладнює виявлення регулярних компонент. Більшість методів дослідження часових рядів включає різні способи фільтрації шуму, що дозволяють побачити регулярну складову більш отчетливо. Більшість регулярних складових часових рядів належить до двох класів: вони є або трендом, або сезонної складової. Тренд є загальну систематичну лінійну або нелінійну компоненту, яка може змінюватися в часі. Сезонна складова - це періодично повторюється компонента. Обидва ці виду регулярних компонент часто присутні в ряді одночасно. Наприклад, продажі компанії можуть зростати з року в рік, але вони також містять сезонну складову (як правило, 25% річних продажів припадає на грудень і тільки 4% на серпень). Не існує "автоматичного" способу виявлення тренда в тимчасовому ряді. Однак якщо тренд є монотонним (стійко зростає або стійко спадає), то аналізувати такий ряд зазвичай неважко. Якщо тимчасові ряди містять значну помилку, то першим кроком виділення тренда є згладжування. Згладжування завжди включає деякий спосіб локального усереднення даних, при якому несистематические компоненти взаємно погашають один одного. Самий загальний метод згладжування - ковзне середнє, в якому кожен член ряду замінюється простим або зваженим середнім n сусідніх членів, де n - ширина "вікна". Замість середнього можна використовувати медіану значень, що потрапили у вікно. Основна перевага медіанного згладжування, в порівнянні зі згладжуванням ковзаючим середнім, полягає в тому, що результати стають більш стійкими до викидів (наявними всередині вікна).

Таким чином, якщо в даних є викиди (пов'язані, наприклад, з помилками вимірювань), то згладжування медіаною зазвичай призводить до більш гладким або, принаймні, більш "надійним" кривим, в порівнянні зі змінним середнім з тим же самим вікном. Основний недолік медіанного згладжування в тому, що при відсутності явних викидів, він призводить до більш "зубчастим" кривим (чим згладжування ковзаючим середнім) і не дозволяє використовувати ваги.[15]

Щодо рідше, коли помилка вимірювання дуже велика, використовується метод згладжування методом найменших квадратів, зважених щодо відстані або метод негативного експоненціально зваженого згладжування. Всі ці методи фільтрують шум і перетворюють дані в щодо гладку криву. Ряди з відносно невеликою кількістю спостережень і систематичним розташуванням точок можуть бути згладжені за допомогою Бікубічеський сплайнів. Багато монотонні тимчасові ряди можна добре наблизити лінійною функцією. Якщо ж є явна монотонна нелінійна компонента, то дані спочатку слід перетворити, щоб усунути нелінійність. Зазвичай для цього використовують логарифмічний, експоненціальне або (менш часто) поліноміальний перетворення даних. Періодична і сезонна залежність (сезонність) являє собою інший загальний тип компонент часового ряду. Це поняття було проілюстровано раніше на прикладі авіаперевезень пасажирів. Можна легко бачити, що кожне спостереження дуже схоже на сусіднє; додатково, є повторювана сезонна складова, це означає, що кожне спостереження також схоже на спостереження, що були в тому ж самому місяці минулого року. Загалом, періодична залежність може бути формально визначена як кореляційний залежність порядку k між кожним i -м елементом ряду i ($i-k$) -м елементом. Її можна виміряти за допомогою автокореляції (тобто кореляції між самими членами ряду). k зазвичай називають лагом (іноді використовують еквівалентні терміни: зрушення, запізнювання). Якщо

помилка вимірювання не дуже велика, то сезонність можна визначити візуально, розглядаючи поведінку членів ряду через кожні k тимчасових одиниць.

Аналіз тренду. Не існує перевірених "автоматичних" методів для визначення компонентів тренду в даних часових рядів; однак до тих пір, поки тенденція є одноманітною (послідовно зростаючим чи зменшеним), що частина аналізу даних зазвичай не дуже складна. Якщо дані часової серії містять значну помилку, то першим кроком у процесі ідентифікації трендів є згладжування.

Згладжування завжди передбачає певну форму локального усереднення даних, такими, що несистематичні компоненти окремих спостережень скасовують один одного. Найпоширенішою технікою є рухоме середнє згладжування, яке замінює кожен елемент серії або простою, або середньою з середніх значень n навколишніх елементів, де n - це ширина згладжування "вікна". Медіани можуть бути використані замість засобів. Основна перевага медіани в порівнянні з рухомою середньою згладжуванням полягає в тому, що її результати менш зміщуються від викидів (у межах вікна згладжування). Таким чином, якщо в даних (наприклад, через помилки вимірювання) є дані про викиди, середнє згладжування, як правило, забезпечує більш гладкі або, принаймні, більш "надійні" криві, ніж ковзний середній на основі тієї ж ширини вікна. Основний недолік середнього згладжування полягає в тому, що за відсутності чітких перевищень він може спричинити більші "нерівні" криві, ніж середня середня, і це не дозволяє робити зважування.\

У відносно менш поширених випадках (у даних часової серії), коли похибка вимірювання дуже велика, можна використовувати згладжування, що зважає на найменші квадрати, або негативні експоненціально зважені методи згладжування. Всі ці методи відфільтровують шум і перетворюють дані на плавну криву, яка є відносно неупередженою від викидів. Серії з відносно

нечисленними систематично розподіленими точками можна згладжувати двокубічними сплайнами.

Встановлення функції. Багато одноманітних даних часової серії можна адекватно апроксимувати лінійною функцією; якщо є чіткий монотонний нелінійний компонент, спочатку дані потрібно перетворити, щоб видалити нелінійність. Зазвичай може бути використана логарифмічна, експоненційна або (рідше) полиномиальна функція.

Сезонна залежність (сезонність) - це ще один загальний компонент схеми часових рядів. Концепцію було проілюстровано на прикладі даних пасажирів авіаліній, наведених вище. Формально вона визначається як кореляційна залежність порядку k між кожним i -м елементом ряду та $(i-k)$ -го елемента i вимірюється автокореляцією (тобто кореляцією між двома термами). k зазвичай називають відставанням. Якщо помилка вимірювання не надто велика, то серійність може бути візуально ідентифікована у серії як шаблон, який повторює кожен k -елемент.

Коррелограм автокореляції. Сезонні візерунки часових рядів можна розглянути через корелограми. Корелограма (автокоррелограма) відображає графічно та чисельно функцію автокореляції, тобто коефіцієнти послідовного кореляції (i їх стандартних помилок) для послідовних лагів у заданому діапазоні лагів (наприклад, від 1 до 30). Діапазони двох стандартних помилок для кожного відставання зазвичай позначаються в корелограмах, але, як правило, розмір автокореляції викликає більший інтерес, ніж його надійність, оскільки ми зазвичай зацікавлені лише в дуже сильних (i , таким чином, дуже істотних) автокореляціях.

Часткові автокореляції. Іншим корисним методом вивчення послідовних залежностей є вивчення функції часткової автокореляції - розширення автокореляції, де видаляється залежність від проміжних елементів (тих, що знаходяться у відстані). Іншими словами, часткова

авторелігація схожа на автокореляцію, за винятком того, що при її підрахунку (автоматичні) кореляції з усіма елементами в період відставання виключаються. Якщо вказано відставання 1 (тобто, у проміжку відсутні проміжні елементи), то часткова автокореляція еквівалентна автокореляції. У певному сенсі часткова автокорреляція забезпечує "чистішу" картину серійних залежностей для окремих лагів (не затьмарені іншими серійними залежностями).

Видалення послідовної залежності. Послідовна залежність певного відставання k може бути знята шляхом розмежування ряду, тобто перетворення кожного i -го елемента ряду в його різницю з $(i-k)$ з i th елементом. Є дві основні причини таких перетворень.

По-перше, ми можемо визначити прихований характер сезонних залежностей у серії. Автокореляції для послідовних лагів є взаємозалежними. Тому видалення деяких автокореляцій змінить інші автоматичні кореляції, тобто вони можуть усунути їх, або це може зробити деякі інші сезонності більш очевидними.

Іншою причиною для видалення сезонних залежностей є встановлення стаціонарних рядів, необхідних для ARIMA та інших методів. В реальних дослідженнях та практиці моделі даних незрозумілі, індивідуальні спостереження припускають значну помилку, і нам все ще потрібно не тільки розкривати приховані моделі даних, але і генерувати прогнози. Методологія ARIMA, дозволяє це робити; вона отримала величезну популярність у багатьох сферах, і наукова практика підтверджує свою силу та гнучкість. Однак, завдяки своїй силі та гнучкості ARIMA - це складна техніка, її нелегко використовувати, вона вимагає великого досвіду, і хоча вона часто виробляє задовільні результати, ці результати залежать від рівня знань дослідника.

Авторегресійна модель. Більшість часових серій складаються з елементів, які серійно залежать в тому сенсі, що можна оцінити коефіцієнт

або набір коефіцієнтів, які описують послідовні елементи серії з конкретних відсталих (попередніх) елементів. Це можна звести до рівняння:

$$x_t = \xi + \phi_1 * x_{(t-1)} + \phi_2 * x_{(t-2)} + \dots + \varepsilon, \quad (3.1)$$

де ξ – константа;

ϕ_1, ϕ_2, ϕ_3 – параметри авторегресійної моделі.

Авторегресійний процес буде стабільним тільки в тому випадку, якщо параметри знаходяться в певному діапазоні, наприклад, якщо є лише один параметр авторегресії, то він повинен знаходитись у межах інтервалу від -1 до 1. В іншому випадку, колишні ефекти будуть накопичуватися, а значення послідовних x буде рухатися до нескінченності, тобто ряди не будуть бути стаціонарним. Якщо є кілька авторегресивних параметрів, можна визначити подібні (загальні) обмеження на значення параметрів.

Незалежно від авторегресивного процесу, кожен елемент у серії також може впливати на минулу помилку, яку неможливо врахувати авторегресивний компонент, рухоме середнє, тобто:

$$x_t = \mu + \varepsilon_t - \theta_1 * \varepsilon_{t-1} - \theta_2 * \varepsilon_{t-2} - \dots, \quad (3.2)$$

де

μ – константа;

θ_1, θ_2 – це середньозважені параметри моделі.

Кожне спостереження складається з випадкової помилкової складової та лінійної комбінації попередніх випадкових ударів.

Потреба зворотності. Не заглиблюючись в надто докладну інформацію, існує «подвійність» між процесом переміщення середнього та авторегресивним процесом, тобто вище середнього рівняння може бути переписана (перевернута) у авторегресійну форму (нескінченного порядку). Однак, аналогічно умовам стаціонарності, описаним вище, це може бути зроблено лише тоді, коли ковзних середніх параметрів дотримуються певні умови, тобто якщо модель є зворотною. Інакше серія не буде стаціонарною.

Модель ARMA - це злиття моделей AR (p) і MA (q). Моделі AR (p) намагаються пояснити імпульс та середні ефекти зворотного зв'язку, які часто спостерігаються на торгових ринках (ефекти учасників ринку). Моделі MA (q) намагаються відобразити ударні ефекти, що спостерігаються в умовах білого шуму. Ці ударні ефекти можна розглядати як несподівані події, що впливають на процес спостереження.

Модель ARMA намагається відобразити обидва ці аспекти при моделюванні фінансових часових рядів. Модель ARMA не враховує кластеризацію волатильності - ключових емпіричних явищ багатьох фінансових часових рядів, які ми обговоримо пізніше.

ARMA (1,1) модель - це:

$$x(t) = a * x(t-1) + b * e(t-1) + e(t), \quad (3.3)$$

де

$e(t)$ – білий шум.

Для моделі ARMA часто потрібно менше параметрів, ніж модель AR (p) або MA (q). Тобто він є надмірним за своїми параметрами.

Модель авторегресії з інтегрованим ковзним середнім, загальна модель включає в себе авторегресивні та рухомі середні параметри і явно включає в

себе різницю в формулюванні моделі. Зокрема, три типи параметрів у моделі – авторегресійні параметри (p), кількість диференційованих пропусків (d) та рухомі середні параметри (q). У позначеннях моделі підсумовуються як ARIMA (p, d, q); так, наприклад, модель, що описується як $(0, 1, 2)$, означає, що вона містить 0 (нуль) авторегресивних (p) параметрів і 2 ковзних середніх (q) параметри, які були обчислені для серії після того, як вона була розрізнена один раз.

Вхідні серії для ARIMA повинні бути стаціонарними, тобто вони повинні мати постійну середню, дисперсію та автокореляцію через час. Тому, як правило, серія в першу чергу повинна бути різницею, поки вона не стаціонарна (це також часто вимагає журналу перетворення даних для стабілізації дисперсії). Кількість разів, коли серія потрібно відрізнути, щоб досягти стаціонарності, відображається в параметрі d . Щоб визначити необхідний рівень диференціації, слід вивчити сюжет даних та автокорелограму. Значні зміни рівня (сильні зміни вгору або вниз) зазвичай вимагають відмінності першого порядку не сезонного ($\text{lag} = 1$), сильні зміни схилу зазвичай вимагають не сезонного диференціювання другого порядку. Сезонні візерунки вимагають відповідного сезонного розмежування. Якщо передбачувані коефіцієнти автокореляції знижуються повільно при більш тривалій затримці, зазвичай потрібно відрізнути перше замовлення. Проте слід мати на увазі, що деякі часові ряди можуть вимагати незначного відмінності або взагалі не відрізнятись, і що по різних серіях виробляються менш стійкі оцінки коефіцієнтів.

На цьому етапі також повинні вирішити, скільки параметрів авторегресивного (p) та ковзного середнього (q) необхідні для отримання ефективної, але все ще парсимонічної моделі процесу (парсимонічні означає, що це має найменші параметри і найбільшу кількість ступенів свободи серед

усіх моделей, які відповідають даним). На практиці кількість параметрів p або q дуже рідко повинна бути більшою за 2.

На наступному етапі (оцінка) параметри оцінюються (використовуючи процедури мінімізації функцій для отримання додаткової інформації про процедури мінімізації), так що сума квадратних залишків мінімізується. Оцінки параметрів використовуються на останньому етапі (прогнозування) для обчислення нових значень серії (за винятком тих, що входять до складу вхідних даних) та довірчих інтервалів для цих прогнозованих значень. Процес оцінки виконується на трансформованих (диференційованих) даних; перед створенням прогнозів серії необхідно інтегрувати (інтеграція - зворотна різниця), так що прогнози виражаються в значеннях, сумісних з вхідними даними. Модель авторегресії з інтегрованим ковзним середнім:

$$\tilde{y}(k) = d^m y(k), \quad (3.4)$$

де

$$dy(k) = y(k) - y(k - 1).$$

$$\begin{aligned} \tilde{y}(k) = & a_0 + a_1 \tilde{y}(k - 1) + a_2 \tilde{y}(k - 2) + \dots + a_p \tilde{y}(k - p) + \varepsilon(k) - \\ & - b_1 \tilde{\varepsilon}(k - 1) - b_2 \tilde{\varepsilon}(k - 2) - \dots - b_q \tilde{\varepsilon}(k - q) \end{aligned} \quad (3.5)$$

3.2 Критерії адекватності регресійної моделі

Регресійна модель, побудована безпосередньо за результатами експерименту, дає можливість розрахувати значення відгуку в абсолютно різних точках області зміни факторів.

Коректна перевірка адекватності регресійної математичної моделі дає можливість відповідати на питання, чи буде щойно побудована модель прогнозувати значення вихідної величини з тією ж самою точністю, що і вихідні результати експерименту.

Коефіцієнт детермінації необхідний для виміру доцільності та коректності опису системи побудованою регресією.

Коефіцієнт детермінації безпосередньо інтегрально описує досить точні властивості рівнянь регресії. Він вказує, яка саме частина всього розрахунку експериментальних значень відносно свого середнього, виникла через регресійну залежність:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} \quad (3.6)$$

де $\text{var}(\hat{y})$ – дисперсія частини ряду головної змінної, що описується отриманим рівнянням;

$\text{var}(y)$ – вибіркова дисперсія основної змінної.

Для адекватної моделі коефіцієнт детермінації має прямувати до одиниці $R^2 \rightarrow 1$ [3].

Розрахунок суми квадратів похибок моделі

$$SSE = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2, \quad (3.7)$$

де $\hat{y}(k) = \hat{a}_0 + \hat{a}_1 \hat{y}(k-1) + \hat{a}_2 \hat{y}(k-2) + \hat{b}_1 \hat{x}(k) + b_2 z(k)$;

$y(k)$ – безпосереднє вимірювання;

N – відома довжина вибірки.

Вказаний критерій ефективно використовувати при виборі оптимальної моделі прогнозування в пакетах прикладних програм з прогнозування. Кожне значення помилки прогнозування зводиться до квадрату, що підкреслює великі помилки прогнозу. Ця властивість має велике значення, оскільки модель прогнозування, що дає середні по величині помилки, в деяких випадках може бути кращою відносно іншої моделі, що має менші помилки, але іноді з недопустимими по абсолютній величині викидами. З можливих варіантів необхідно обирати саме таку допустиму модель, щоб для неї SSE набувало мінімального значення.

Інформаційний критерій Акайке обчислюється таким чином, що враховує суму квадратів похибок, а також відому кількість вимірів N і кількість параметрів моделі, що оцінюються p :

$$AIC = N \ln[\sum_{k=1}^N e^2(k)] + 2p. \quad (3.8)$$

Для найкращої обраної моделі критерій має набувати найменше значення, так як він має пряму залежність від суми квадратів похибок.

Статистика Дарбіна-Уотсона обчислюється за формулою:

$$DW = 2 - 2\rho, \quad (3.9)$$

де ρ – коефіцієнт кореляції між сусідніми значеннями випадкової змінної $\varepsilon(k) \approx e(k)$, тобто $\rho = cov[e(k)] = E[e(k)e(k-1)]$.

Цей параметр дозволяє визначити ступінь корельованості похибок моделі. При повній відсутності кореляції між похибками $DW = 2$, це найбільше прийнятне значення даного параметра [5].

Статистика Фішера F визначає ступінь адекватності побудованої моделі загалом. Для адекватної моделі виконується умова:

$$F > F_{\text{крит}}, \quad (3.10)$$

де $F_{\text{крит}}$ знаходиться по таблиці.

Значення статистики Фішера F пропорційне $R^2/(1 - R^2)$, де R^2 – коефіцієнт детермінації.

Отже, моделі вищого ступеня адекватності відповідає більше значення F .

Коефіцієнт Тейла є суттєво важливим індикатором безпосередньої точності побудованої моделі та її доцільності:

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i)^2 + \frac{1}{N} \sum_{i=1}^N (\hat{y}_i)^2}} \quad (3.11)$$

Величина коефіцієнта Тейла знаходиться між 0 і 1. Якщо $U = 1$, то модель не придатна для використання для прогнозу. Прогнозовані та реальні ряди некорельовані. Але в іншому випадку, якщо $U = 0$, то змодельовані та спрогнозовані ряди співпадають зі справжніми і модель описує вихідні дані абсолютно ідеально [6].

Вказаний коефіцієнт може бути розкладений на суму відношення упередженості U^M , відношення варіацій U^S і відношення коваріацій U^C .

Відношення упередженості

$$U^M = \frac{(y_i - \hat{y}_i)^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.12)$$

Він використовується безпосередньо для пошуку присутніх постійних суттєвих відхилень для середніх реальних, а також для прогнозованих рядів. Іншими словами, дослідження, чи модель завищує прогнозування. Чим менша U^M , тим краща модель. Якщо $U^M = 0$, то модель досить якісна і в прогнозованих значеннях повністю відсутня упередженість.

U^S – відношення варіацій, яке визначається так:

$$U^S = \frac{(\sigma - \hat{\sigma})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.13)$$

Відношення варіацій застосовується для перевірки присутності досить динамічних характеристик для поглинання варіацій реальних рядів. U^C – відношення коваріацій, яке визначається так:

$$U^C = \frac{2(1-\rho)(\sigma - \hat{\sigma})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.14)$$

Відношення коваріацій U^C показує, наскільки корельованими є прогнозовані та реальні ряди. Рівність $U^C = 0$ є свідченням того, що прогнозовані і реальні ряди ідеально корельовано [7].

Три зазначені статистики пов'язані між собою таким чином:

$$U^C + U^M + U^S = 1. \quad (3.15)$$

3.3 SWOT – аналіз

SWOT-аналіз є основою, яка використовується для оцінки конкурентної позиції компанії, визначаючи її сильні сторони, слабкі сторони, можливості та загрози. Зокрема, SWOT-аналіз - це основна модель оцінки, яка вимірює, що може і не може зробити організація, а також її потенційні можливості та загрози. Аналіз SWOT - це базова аналітична база, яка оцінює, що таке суб'єкт - зазвичай бізнес, хоча це може бути місцем , промисловість або продукт - може і не може робити, для факторів, як внутрішніх, так і зовнішніх. Використовуючи дані про навколишнє середовище для оцінки позиції компанії, SWOT-аналіз визначає, що допомагає фірмі в досягненні своїх цілей, і які перешкоди вона повинна подолати або звести до мінімуму, щоб досягти бажаних результатів: де така організація сьогодні і де вона може піти. Аналітики представляють SWOT-аналіз як квадрат з кожною з чотирьох областей, що складають один квадрант. Ця візуальна структура забезпечує швидкий огляд позиції компанії. Хоча всі моменти в рамках певного заголовку можуть не мати однакового значення, існує розуміння того, як кількість можливостей вимірюється на кількість загроз тощо.

Використовуючи SWOT-аналіз, організація повинна бути реалістичною щодо своїх хороших та поганих точок. Організація повинна зберігати аналіз конкретних, уникаючи сірих областей та аналізуючи їх у контексті реального життя. Наприклад, як порівнюються продукти та послуги організації з конкуруючими фірмами? SWOT-аналіз повинен бути коротким і простим, і слід уникати складності та надмірного аналізу, тому що більша частина інформації є суб'єктивною. Таким чином, компанії повинні використовувати його як керівництво, а не рецепт.

Перші два букви в абревіатурі S (сильні сторони) та W (слабкі сторони) відносяться до внутрішніх факторів, що означає, що ресурси та досвід легко доступні для вас. Приклади таких сфер, як правило, розглядаються:

- Фінансові ресурси (фінансування, джерела доходу, інвестиційні можливості)
- Фізичні ресурси (місце розташування, приміщення, обладнання)
- Людські ресурси (працівники, волонтери, цільові аудиторії)
- Доступ до природних ресурсів, торгових марок, патентів та авторських прав
- Поточні процеси (програми працівників, ієрархії відділів, програмні системи)

Зовнішні сили впливають на кожну компанію, організацію та індивідуальну діяльність. Чи є ці фактори прямо чи опосередковано пов'язані з можливістю або загрозою, важливо взяти до уваги і документувати кожен з них. Зовнішні фактори зазвичай вказують на те, що ви або ваша компанія не контролюєте, наприклад:

- Ринкові тенденції (нові продукти та технології, зміна потреб аудиторії)
- Економічні тенденції (місцеві, національні та міжнародні фінансові тенденції)
- Фінансування (пожертвування, законодавча та інші джерела)
- Демографічні показники
- Відносини з постачальниками та партнерами
- Політичні, екологічні та економічні норми

Сильні сторони (Strengths) описують те, що відрізняє організація і відокремлює її від конкуренції: сильний бренд, лояльна клієнтська база, сильний баланс, унікальні технології тощо. Наприклад, хедж-фонд може створити власну торговельну стратегію, яка повертає результати, що торкаються на ринку. Потім він повинен вирішити, як використовувати ці результати для залучення нових інвесторів.

Слабкі сторони (Weaknesses) припиняють організацію виконувати на своєму оптимальному рівні. Це регіони, де бізнес потребує поліпшення, щоб залишатися конкурентоспроможним: товарообіг вище, ніж середньостатистичний товарообіг, високий рівень заборгованості, неадекватна ланцюжок поставок або брак капіталу.

Можливості (Opportunities) - це сприятливі зовнішні чинники, які організація може використати, щоб надати йому конкурентну перевагу. Наприклад, виробник автомобілів може експортувати свої автомобілі на новий ринок, збільшуючи продажі та частку ринку, якщо країна знижує тарифи.

Загрози (Threats) стосуються факторів, які можуть завдати шкоди організації. Наприклад, посуха є загрозою для компанії з виробництва пшениці, оскільки вона може знищити або зменшити врожайність. Інші загальні загрози включають такі речі, як підвищення витрат на ресурси, збільшення конкуренції, обмежена пропозиція робочої сили тощо.[8]

SWOT-аналіз є чудовим способом ведення бізнес-стратегічних зустрічей. Можливо, що всі в кімнаті обговорюють основні сильні та слабкі сторони компанії, а потім переходять звідти до визначення можливостей та загроз, і, нарешті, ідеї мозкового штурму. Часто аналіз SWOT, який передбачається перед початком сеансу, змінюється, щоб відобразити чинники, про які не було відомо, і ніколи б не захопили, якщо не для введення групи.

Компанія може використовувати SWOT для загальних сесій бізнес-стратегії або для певного сегменту, як-от маркетинг, виробництво чи продаж. Таким чином, можна побачити, як загальна стратегія, розроблена SWOT-аналізом, буде відфільтрована до сегментів нижче, перш ніж взяти на себе зобов'язання. Також можна працювати у зворотному напрямку з SWOT-аналізом, орієнтованим на сегмент, який подає загальний аналіз SWOT.[7]

Складові SWOT аналізу зображені на рис. 3.1:

SWOT аналіз	Сильні сторони - S	Слабкі сторони - W
Можливості - O	SO - Використати можливості та Розширити сильні сторони	WO - Використати можливості та Зменшувати слабкі сторони
Загрози - T	ST - Використати сильні сторони та Зменшити загрози	WT - Зменшувати слабкі сторони та Уникати загрози

Рисунок 3.1 – Компоненти SWOT аналізу

У випадку SWOT аналізу для підприємств різницевої торгівлі, компоненти SWOT мають наступні елементи:

- сильні сторони – існуючий бренд, клієнти та продажі;
- слабкі сторони – сприйняття бренду, посередницьке використання, технології;
- можливості – перехресні продажі, нові ринки, нові сервіси, співпраця;
- загрози – вибір користувача, нові конкуренти, нові продукції, конфлікти.

Компонентами SWOT матриці є значення від 0 до 1, що позначають силу зв'язку вказаних елементів, 0 – немає зв'язку і 1 – пряма залежність. [8] У формулах (3.16) та (3.17) відображається розрахунок коефіцієнтів впливу внутрішніх характеристик на реалізацію загроз та можливостей:

$$D_i = \sum_j K_{S_j T_j} - \sum_k K_{W_k T_i}, \quad (3.16)$$

де

$K_{S_j T_j}$ – сила взаємозв'язку між сильною стороною та загрозою;

$K_{W_k T_i}$ – сила взаємозв'язку між слабкою стороною та загрозою.

$$H_m = \sum_j K_{S_j O_m} - \sum_k K_{W_k O_m}, \quad (3.17)$$

де

$K_{S_j O_m}$ – сила взаємозв'язку між сильною стороною та можливістю;

$K_{W_k O_m}$ – сила взаємозв'язку між слабкою стороною та можливістю.

У формулах (3.18) та (3.19) показаний розрахунок загроз ти можливостей, на які впривають сильні та слабкі характеристики:

$$F_j = \sum_i K_{S_j T_i} + \sum_m K_{S_j O_m}, \quad (3.17)$$

де

$K_{S_j T_i}$ – сила взаємозв'язку між сильною стороною та загрозою;

$K_{S_j O_m}$ – сила взаємозв'язку між сильною стороною та можливістю.

$$G_k = \sum_i K_{W_k T_i} + \sum_m K_{W_k O_m}, \quad (3.17)$$

де

$K_{W_k T_i}$ – сила взаємозв'язку між слабкою стороною та загрозою;

$K_{W_k O_m}$ – сила взаємозв'язку між слабкою стороною та можливістю.

Можливості розраховуються:

$$O = \sum_i H_i \quad (3.18)$$

Загрози розраховуються:

$$T = \sum_i D \quad (3.18)$$

Сильні сторони розраховуються:

$$S = \sum_i F_i \quad (3.18)$$

Слабкі сторони розраховуються:

$$W = \sum_i G_i \quad (3.18)$$

SWOT-аналіз може допомогти процесу прийняття рішень шляхом створення візуального представлення різних факторів, які найбільшою мірою вплинуть на те, чи зможе бізнес, проект, ініціатива або індивідуальний учасник успішно досягти поставленої мети.

Хоча такий знімок важливий для розуміння багатогранної динаміки, яка впливає на успіх, SWOT-аналіз має свої межі. Аналіз може не включати всі релевантні фактори для всіх чотирьох елементів, тим самим даючи зсунуту перспективу. Крім того, тому що він фіксує лише фактори в певний момент часу і не дозволяє тим, як ці фактори можуть змінюватися з часом, запропонована ним інформація може мати обмежений термін достовірності.[7]

Висновки до розділу

Отже, використовуючи аналіз і прогнозування часових рядів, можливо побудувати моделі продажів та спрогнозувати кількість проданих товарів у майбутньому. За допомогою SWOT аналізу розв'язуються задачі врахування впливу внутрішніх і зовнішніх компонент системи рителеру одне на одного задля оптимізації його роботи.

РОЗДІЛ 4 ПОБУДОВА СППР

Для аналізу і виробок пропозицій в СППР використовуються різні методи:

- інформаційний пошук,
- інтелектуальний аналіз даних,
- пошук знань в базах даних,
- міркування на основі прецедентів,
- імітаційне моделювання,
- еволюційні обчислення і генетичні алгоритми,
- нейронні мережі,
- ситуаційний аналіз,
- когнітивне моделювання.

Система підтримки прийняття рішень дозволяє полегшити роботу керівникам підприємств і підвищити її ефективність. Вони значно прискорюють вирішення проблем в бізнесі. Система підтримки прийняття рішень сприяють налагодженню міжособистісного контакту. На їх основі можна проводити навчання і підготовку кадрів. Дані інформаційні системи дозволяють підвищити контроль над діяльністю організації.

Наявність чітко функціонуючої системи підтримки прийняття рішень дає великі переваги в порівнянні з конкуруючими структурами. Завдяки пропозицій, що висуваються СППР, відкриваються нові підходи до вирішення повсякденних і нестандартних завдань.

Якщо не брати до уваги обмеження доступу до даних, аналітик має можливість обробляти і зіставляти гігантські обсяги інформації, використовуючи при цьому всю аналітичну міць системи. Очевидно, що багатовимірні операції пошуку, моделювання і прогнозу можуть створювати

пікові навантаження на апаратні ресурси системи, тому, пропонуючи в якості базової функціональності векторні операції над історичними даними, постачальник рішення зобов'язаний передбачити додаткові вбудовані засоби контролю їх використання. Це можуть бути жорсткі обмеження на обсяг оброблюваної інформації або "м'які" засоби оповіщення про прогнозований час обробки запиту. Для деяких слабо підготовлених користувачів частина найбільш ресурсномісткої функціональності може бути взагалі закрита. У будь-якому випадку повинна бути повністю виключена ситуація, коли аналітик "підвішує" систему на тривалий час. На жаль, саме такі ситуації трапляються при використанні штатного рішення. У базову функціональність просто не закладені необхідні в даному випадку можливості, а її закритість і жорсткість не дозволяють добудовувати відповідні компоненти при впровадженні. Наслідки очевидні - користувачі не контролюють ситуацію і часто завершують роботу нештатним чином. При інтенсивній роботі занадто численних користувачів обчислювальні ресурси системи «входять в клінч». Причому ні організаційними, ні технічними способами вирішувати такі ситуації, як правило, не вдається. Система стає практично непридатною для корпоративного промислового використання.

Система підтримки прийняття рішень має головні складові, це – інтерфейс користувача, база даних та сервіс, що обробляє дані, що зображено на рис. 4.1:

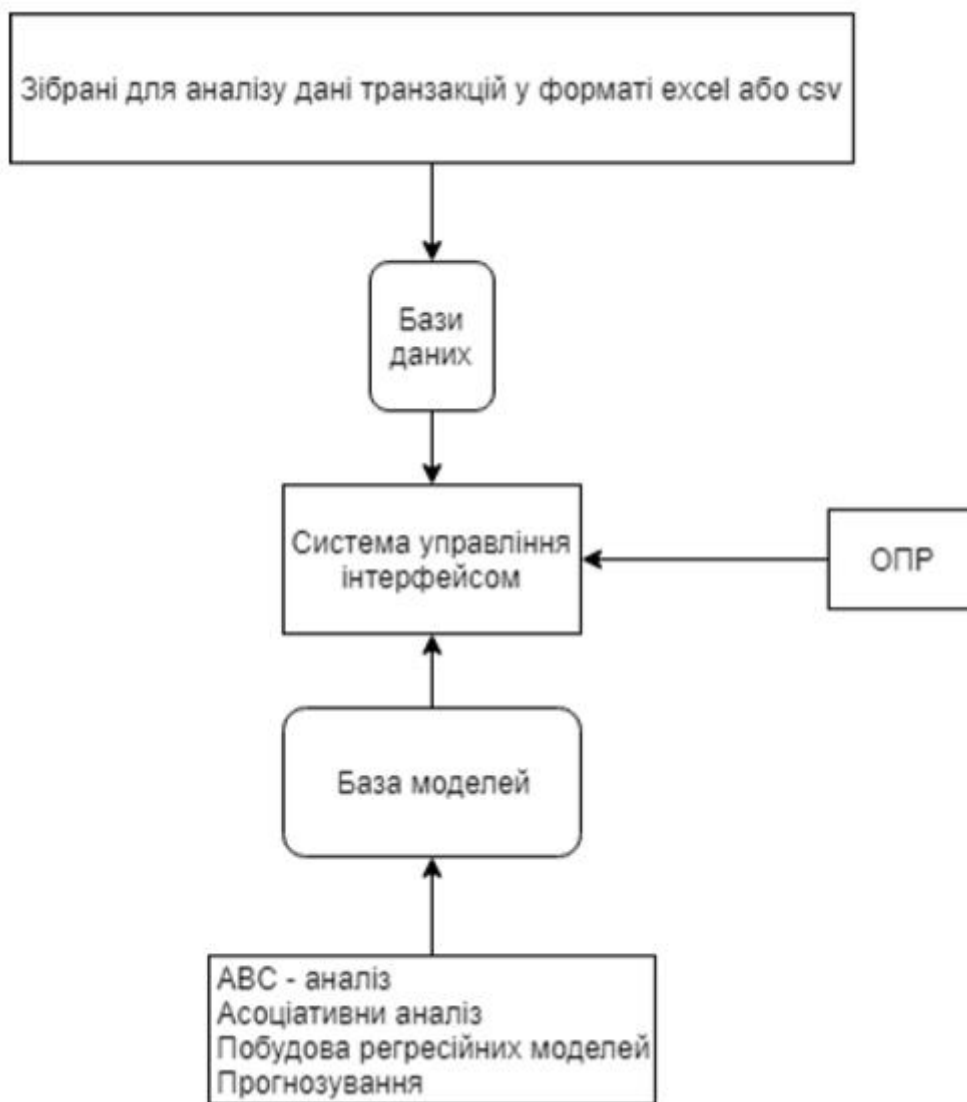


Рисунок 4.1 – структура СППР

4.1 Опиис СППР

СППР розроблена в додатку Power BI з використанням R-скриптів. Для використання системи, необхідно завантажити дані продажів у файл СППР.rbit, який є шаблоном системи. Утворений файл СППР.rбіх відображає розв'язані задачі у вигляді графіків та таблиць.

Дані для аналізу мають бути нормалізованими. Зокрема, на вхід подається чотири файли у форматі CSV або EXCEL. Перший файл містить дані транзакцій, в яких вказується номер транзакції, товар, ціна, номер чеку, кількість і так далі, як зображено на рис. 4.1:

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Transaction No.	Line No.	Receipt No.	Item No.	Sales Staff	Item Category Code	Product Group Code	Price	Net Price	Quantity	VAT	VAT	Date	Customer No.
1010007738	10000	1,01E+11	11536		76	762	2,29	1,89	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	20000	1,01E+11	133038		141	1414	5,99	4,95	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	30000	1,01E+11	141401		94	942	0,99	0,82	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	40000	1,01E+11	149459		142	1421	4,99	4,12	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	50000	1,01E+11	150004		111	1111	4,49	3,71	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	60000	1,01E+11	152159		142	1422	3,99	3,3	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	70000	1,01E+11	152282		82	822	29,99	24,79	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	80000	1,01E+11	165213		93	936	29,99	24,79	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	90000	1,01E+11	177225		76	761	2,29	1,89	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	100000	1,01E+11	201904		76	762	9,99	8,26	-1	CEU	A	6-1-2016	CUSTO2000320
1010007738	110000	1,01E+11	204629		73	731	46	38,02	-1	CEU	A	6-1-2016	CUSTO2000320

Рисунок 4.1 – Дані транзакцій

Наступний файл відображає дані категорій товарів, відповідно код товару та його назва. Приклад зображений на рис.4.2.

A	B
Code	Description
11	Sofas
12	Armchairs, footstool & sofa ta
13	Armchairs
14	Occasional tables range
15	Solitaire sofa-beds, armchairs
16	Seating & Reclining Furniture
21	Multipurpose furniture
22	Store and display furniture
23	Tv furniture & electronics
24	Living-room Storage
25	Living room tables
26	Secondary Storage
31	Workspaces
32	Work storage furniture.
33	Work seating range
34	Flooring
35	Bedroom furniture
36	Mattresses
37	Bedroom storage
38	Storing Toiletries
41	Beds

Рисунок 4.2 - Категорії товарів

Третій файл зберігає дані про групи продуктів для розбиття категорій на групи товарів. На рис. 4.3 зображено код групи товарів, її назва та розподіл по категоріям:

A	B	C
Code	Description	Item Category
111	Sofas	11
112	Leather sofas	11
113	Sofa beds	11
114	Armchairs	11
115	Footstools	11
119	Furniture fabrics	11
121	Sofa tables	12
122	Solitaire sofa beds	12
123	Glass-door cabinets, shelving	12
124	Audio & video furniture - UTG	12
125	Armchairs	12
126	Footstools	12
131	Layer glued armchairs	13
132	Solitaire armchairs	13
133	Natural fibre armchairs	13

Рисунок 4.3 - Приклад таблиці вхідних даних груп продуктів

В останньому файлі зберігається інформація щодо номеру товару, його назви, вартості, доцільність його оборотів, групи та категорії товару, як зображено на рис. 4.4.

A	B	H	I	J	K	X	Y
No.	Description	Unit Price	Profit %	Unit Cost	Vendor No.	Item Cate	Product G
137	MEASURING KITCHEN	0	0	0	zero	973	9731
203	KITCHEN PLUMBING	0	0	0	zero	972	9722
496	BATHROOM PLUMBING	0	0	0	zero	972	9722
528	INST IKEA GAS APP	0	0	0	zero	972	9722
533	inst of wrktp	0	0	0	zero	972	9722
552	CONN UPPEVA TV DEV	0	0	0	zero	972	9722
566	HOME DELIV OUTSD REGION	0	0	0	zero	971	9711
571	LOAD SERV VAT SPEC	0	0	0	zero	971	9713
11013	LACK WLL SHLF 190X26 WHITE4	66,1074	86,9152	8,65	VEN02000006	21	215
11428	GRUNDTAL WLL SHLF 80 STAINLESS STEEL	14,04132	61,7771	5,367	VEN02000006	76	762
11536	GRUNDTAL S-HOOK 11 STAINLESS STEEL 5-P	1,89256	63,84791	0,6842	VEN02000006	76	762
15134	SVALKA WHI WNE GLS 25 CL CLEAR GLASS 6-P	2,8843	27,53874	2,09	VEN02000006	152	1521
17133	VASEN VASE 20 CLEAR GLASS	0,81818	7,07424	0,7603	VEN02000006	163	1633
26707	ENJE RLLER BLIND 120X250 WHITE	16,5207	48,2466	8,55	VEN02000006	121	1213
28508	LEN N PILLOW COT 35X55 WHITE	1,97521	30,55928	1,3716	VEN02000006	91	914
30196	IKEA 365+ SIDE PLATE 18X18 BEIGE	8,2562	100	0,98	VEN02000006	151	1511
30262	FORVAR JAR/LID 1,8 L GLASS/ALUMINIUM-COLOUR	1,2314	-10,77635	1,3641	VEN02000006	143	1432
30993	GRANAT N CUSHION 50X50 WHITE	8,2562	100	1,7899	VEN02000006	123	1231
31841	EDSVIK N KIT MX TP DUAL CHROME-PLATED	41,31405	41,47875	24,1775	VEN02000006	75	751
35127	HASTVEDA ACH RATTAN	47,6281	42,4568	27,4067	VEN02000006	12	125
37980	MELODI PEND LMP	4,12397	32,05576	2,802	VEN02000006	102	1023
39804	BRUNKRISSLA QC/2PWC 200X200/50X60 RED	14,86777	18,52241	12,1139	VEN02000006	111	1112
46786	SKYDD WOOD TREATMENT OIL INDOOR USE 0,5 L	4,12397	46,62425	2,2012	VEN02000006	142	1424
49520	TUPLUR RLLER BLIND 140X195 WHITE	61,9752	100	13,8699	VEN02000006	121	1213
54564	CAPITA N LEG 16 STAINLESS STEEL 4-P	30,5785	100	0	VEN02000006	71	711
55691	LOCK CEIL LMP WHITE	3,29752	29,94432	2,3101	VEN02000006	102	1022
56884	GALANT FRAME 160 SILVER-COLOUR	32,2314	69,08853	9,9632	VEN02000006	31	312
57256	BLANDA BLANK SERV BOWL 36 STAINLESS STEEL	8,2562	35,31649	5,3404	VEN02000006	151	1511
58487	FLYTTA KITCHEN TROLLEY 98X57 STAINLESS STEEL	114,87603	46,81745	61,094	VEN02000006	78	782
58519	GORM SHELV UT 78X55X174 SOFTWOOD	95,0413	100	20,8729	VEN02000006	191	1911
63672	TROFAST FRAME 44X91 PINE	19,8265	13,2724	17,195	VEN02000006	92	921

Рисунок 4.4 - Приклад таблиці вхідних даних товарів

4.2 Звантаження вхідних даних

Перш за все, необхідно встановити програму Power BI Desktop. Далі обрати кнопку «Завантажити дані». Після цього, необхідно вибрати чотири файли, які містять в собі дані для аналізу. На рис. 4.5 зображен шлях до завантаження даних.

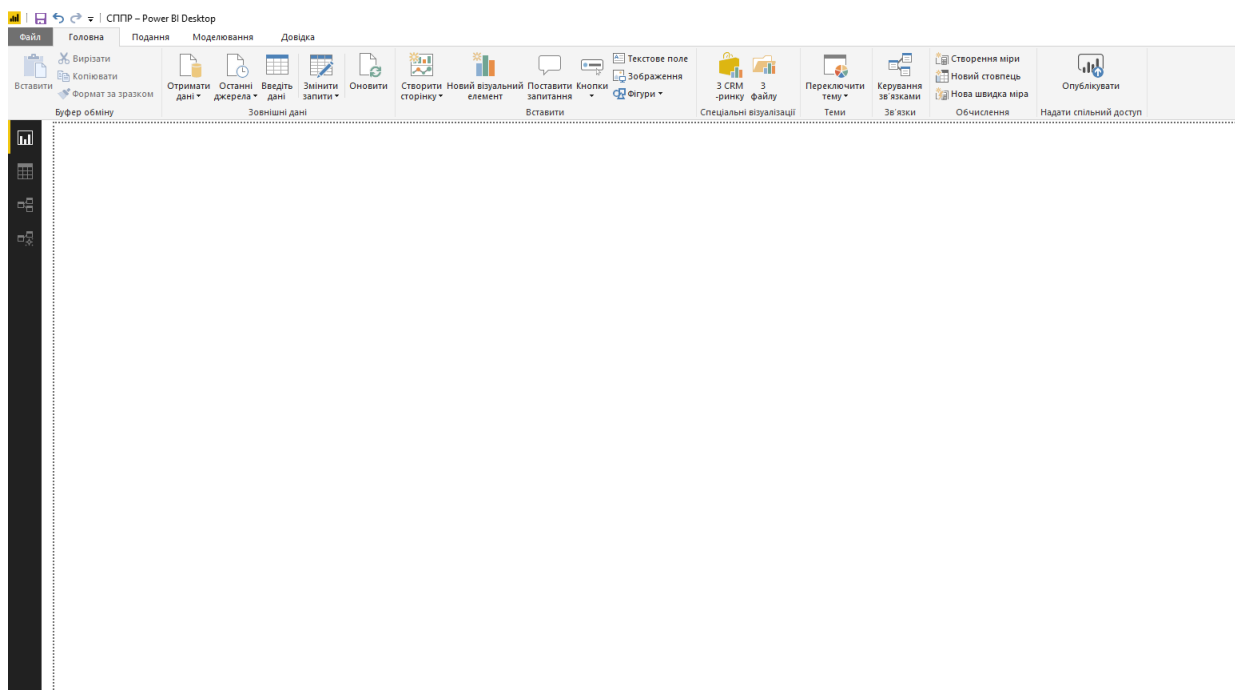


Рисунок 4.5 – Підключення даних для аналізу

На рис. 4.6. зображений загальний інтерфейс програми після завантаження даних.

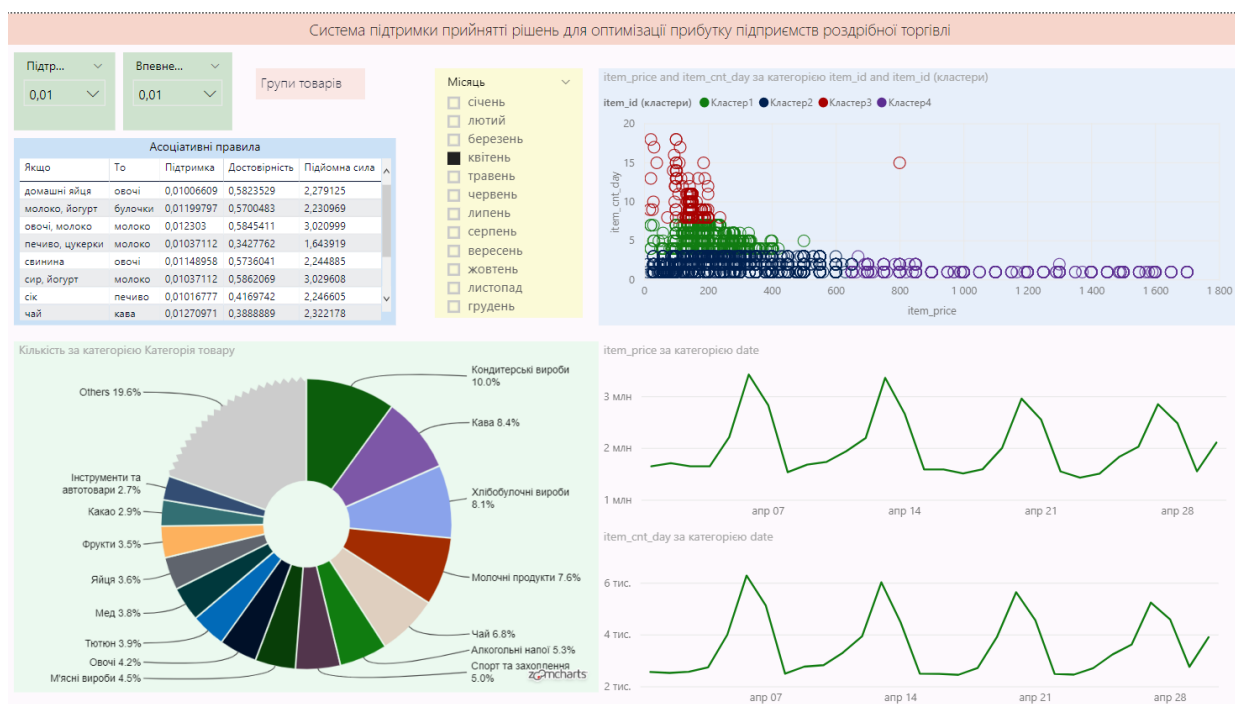


Рисунок 4.6 – Інтерфейс програми

Завантажені файли можна побачити, перейшовши на вкладку «Дані», що зображено на рис. 4.7.

date	date_block_num	shop_id	item_id	item_price	item_cnt_day	item_id (кластер)
30 марта 2013 г.	2	31	8845	149	1	1
30 марта 2013 г.	2	31	8423	149	1	1 Кластер3
30 марта 2013 г.	2	31	8645	149	1	1 Кластер3
30 марта 2013 г.	2	31	8655	149	1	1
30 марта 2013 г.	2	31	8707	149	1	1 Кластер1
30 марта 2013 г.	2	31	10022	149	1	1 Кластер1
30 марта 2013 г.	2	31	10056	149	1	1 Кластер1
30 марта 2013 г.	2	31	10317	149	1	1
30 марта 2013 г.	2	31	10340	149	1	1
30 марта 2013 г.	2	31	9842	149	1	1 Кластер3
30 марта 2013 г.	2	31	7243	149	1	1 Кластер1
30 марта 2013 г.	2	31	7709	149	1	1 Кластер1
30 марта 2013 г.	2	31	7704	149	1	1
30 марта 2013 г.	2	31	7690	149	1	1 Кластер2
30 марта 2013 г.	2	31	12817	149	1	1
30 марта 2013 г.	2	31	12781	149	1	1 Кластер1
30 марта 2013 г.	2	31	12927	149	1	1 Кластер3
30 марта 2013 г.	2	31	12924	149	1	1 Кластер2
30 марта 2013 г.	2	31	12370	149	1	1 Кластер3
30 марта 2013 г.	2	31	12361	149	1	1
30 марта 2013 г.	2	31	12392	149	1	1 Кластер2
30 марта 2013 г.	2	31	12492	149	1	1 Кластер3
30 марта 2013 г.	2	31	14066	149	1	1
30 марта 2013 г.	2	31	14443	149	1	1
30 марта 2013 г.	2	31	14408	149	1	1
30 марта 2013 г.	2	31	13965	149	1	1 Кластер3
30 марта 2013 г.	2	31	11177	149	1	1 Кластер2
30 марта 2013 г.	2	31	11251	149	1	1
30 марта 2013 г.	2	31	11214	149	1	1
30 марта 2013 г.	2	31	11198	149	1	1 Кластер1
30 марта 2013 г.	2	31	10980	149	1	1 Кластер3
30 марта 2013 г.	2	31	11030	149	1	1
30 марта 2013 г.	2	31	12017	149	1	1
30 марта 2013 г.	2	31	12064	149	1	1
30 марта 2013 г.	2	31	12062	149	1	1
30 марта 2013 г.	2	31	12058	149	1	1
30 марта 2013 г.	2	31	11576	149	1	1
30 марта 2013 г.	2	31	97	149	1	1
30 марта 2013 г.	2	31	5416	149	1	1

Рисунок 4.7 – Перегляд вхідних даних

4.3 Аналіз прибутку

Перейшовши на вкладку «Звіт», користувач переглядає проаналізовані дані. В верхньому лівому кутку програми можна детально переглянути асоціативний аналіз. При цьому, можливо задавати параметри, такі як «Підтримка», «Впевненість» та тип, по якому будується асоціативний аналіз, як показано на рис. 4.8.

Підтримка: 0,01

Впевненість: 0,01

Групи товарів

Асоціативні правила

Якщо	То	Підтримка	Достовірність	Підйомна сила
домашні яйця	овочі	0,01006609	0,5823529	2,279125
молоко, йогурт	булочки	0,01199797	0,5700483	2,230969
овочі, молоко	молоко	0,012303	0,5845411	3,020999
печиво, цукерки	молоко	0,01037112	0,3427762	1,643919
свинина	овочі	0,01148958	0,5736041	2,244885
сир, йогурт	молоко	0,01037112	0,5862069	3,029608
сік	печиво	0,01016777	0,4169742	2,246605
чай	кава	0,01270971	0,3888889	2,322178

Рисунок 4.8 - Асоціативний аналіз по групам товарів

На рис. 4.9 зображені фільтри для побудови асоціативного аналізу.

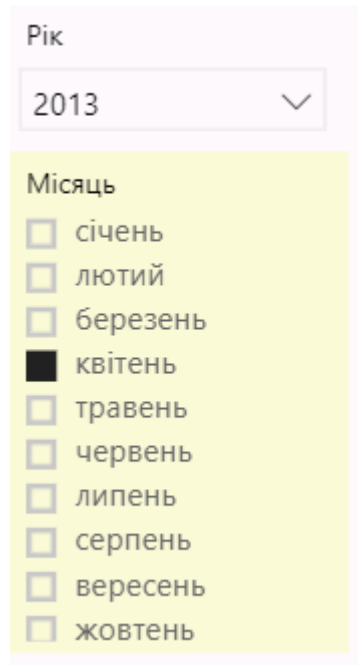
Підтримка: Усі

Впевнен...: Усі

Підтримка	То	Впевнен...
<input type="checkbox"/> 0,01		<input type="checkbox"/> 0,01
<input type="checkbox"/> 0,02		<input type="checkbox"/> 0,02
<input type="checkbox"/> 0,03		<input type="checkbox"/> 0,03
<input type="checkbox"/> 0,04		<input type="checkbox"/> 0,04
<input type="checkbox"/> 0,05	ово	<input type="checkbox"/> 0,05
<input type="checkbox"/> 0,06	бул	<input type="checkbox"/> 0,06
<input type="checkbox"/> 0,07	мол	<input type="checkbox"/> 0,07
<input type="checkbox"/> 0,08	мол	<input type="checkbox"/> 0,08
<input type="checkbox"/> 0,09	ово	<input type="checkbox"/> 0,09
<input type="checkbox"/> 0,10	мол	<input type="checkbox"/> 0,10
чай	печ	0,416
	кава	0,01270971 0,388

Рисунок 4.9 – Фільтри для асоціативного аналізу

Також в програмі можливо задати фільтр по даті, на яку необхідно зробити аналіз, що зображено на рис. 4.10.



Рік

2013

Місяць

- січень
- лютий
- березень
- квітень
- травень
- червень
- липень
- серпень
- вересень
- жовтень

Рисунок 4.10 – Фільтр по даті

Нижче відображається розподіл товарів від найбільш популярних груп, до найменш популярних, за ознакою прибутковості, що зображено на рис. 4.11 – 4.12.

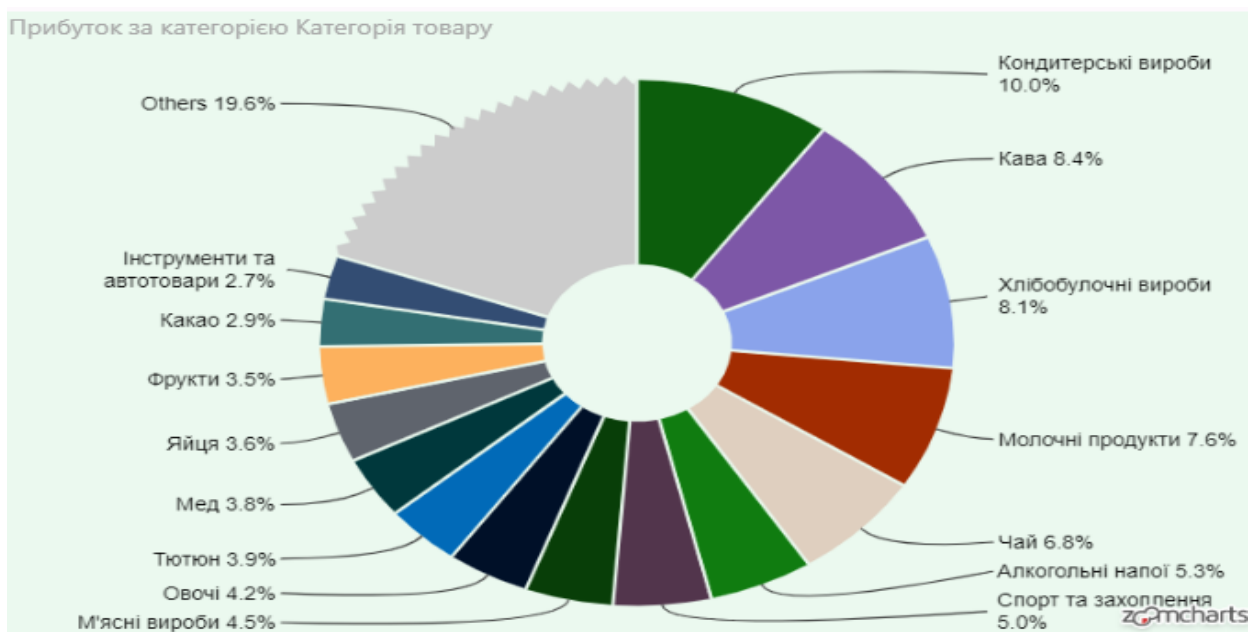


Рисунок 4.11 – Найбільше прибутку від групи товарів

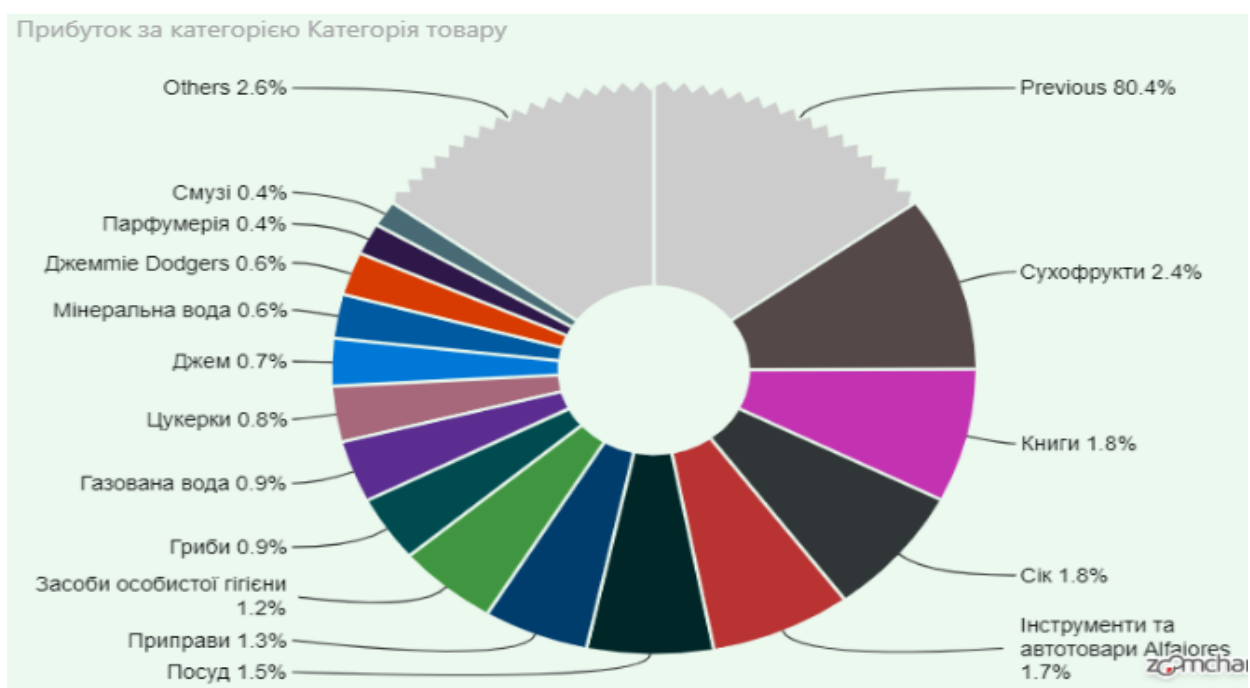


Рисунок 4.12 – Групи товарів за прибутком, що менше 3%

Далі знаходиться графік кластерного аналізу, що відображає поділ товарів на кластери, відповідно до їх ціни та кількості придбаного найменування. На рис.4.13 відображається кластерний аналіз в системі.

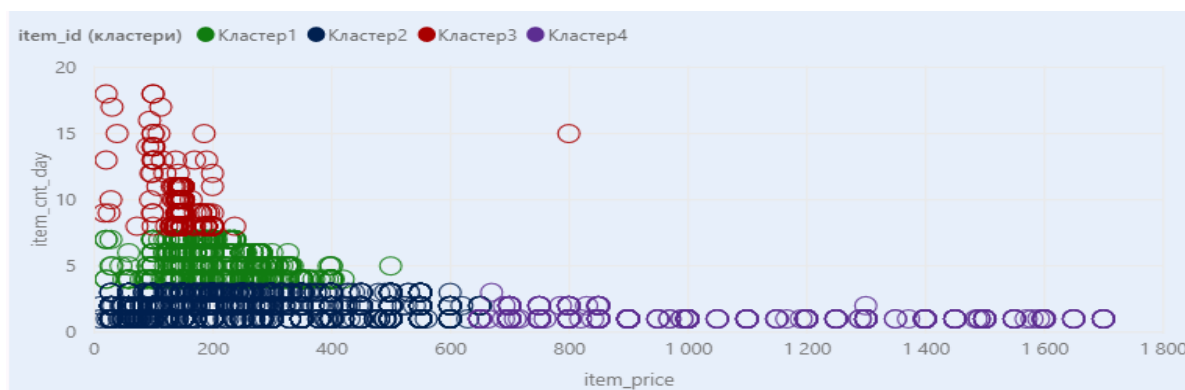


Рисунок 4.13 – Кластерний аналіз

На другій сторінці додатку знаходяться результати SWOT-аналізу, що зображено на рис. 4.14 – 4.19:

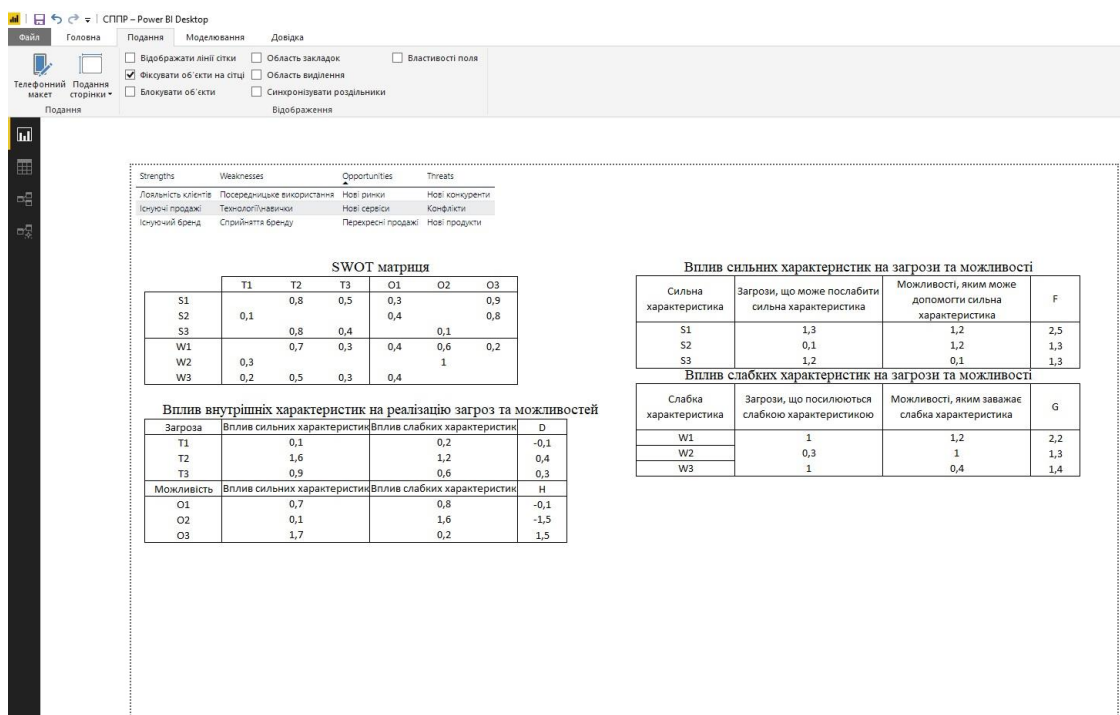


Рисунок 4.14 – Вигляд програми для SWOT аналізу

Strengths	Weaknesses	Opportunities	Threats
Існуючий бренд	Сприйняття бренду	Перехресні продажі	Нові продукти
Існуючі продажі	Технології/навички	Нові сервіси	Конфлікти
Лояльність клієнтів	Посередницьке використання	Нові ринки	Нові конкуренти

Рисунок 4.15 – Елементи SWOT-аналізу

	T1	T2	T3	O1	O2	O3
S1		0,8	0,5	0,3		0,9
S2	0,1			0,4		0,8
S3		0,8	0,4		0,1	
W1		0,7	0,3	0,4	0,6	0,2
W2	0,3				1	
W3	0,2	0,5	0,3	0,4		

Рисунок 4.16 – Заповнена SWOT матриця

Загроза	Вплив сильних характеристик	Вплив слабких характеристик	D
T1	0,1	0,2	-0,1
T2	1,6	1,2	0,4
T3	0,9	0,6	0,3
Можливість	Вплив сильних характеристик	Вплив слабких характеристик	H
O1	0,7	0,8	-0,1
O2	0,1	1,6	-1,5
O3	1,7	0,2	1,5

Рисунок 4.17 – Вплив внутрішніх характеристик на реалізацію загроз та можливостей

Сильна характеристика	Загрози, що може послабити сильна характеристика	Можливості, яким може допомогти сильна характеристика	F
S1	1,3	1,2	2,5
S2	0,1	1,2	1,3
S3	1,2	0,1	1,3

Рисунок 4.18 – Вплив сильних характеристик на загрози та можливості

Слабка характеристика	Загрози, що посилюються слабкою характеристикою	Можливості, яким заважає слабка характеристика	G
W1	1	1,2	2,2
W2	0,3	1	1,3
W3	1	0,4	1,4

Рисунок 4.19 – Вплив слабких характеристик на загрози та можливості

З отриманих результатів SWOT аналізу, можна зробити висновок, що необхідно використовувати такі сильні сторони, як лояльність клієнтів та існуючий бренд для покращення якості обслуговування, що допоможе заходити на нові ринки, продавати нові категорії товару, продавати нові категорії товару. Нові технології також допоможуть знайти нових клієнтів завдяки зручності у користуванні послугами ритейлера. Щодо слабких сторін, то підприємству необхідно уникати конфліктів з клієнтами та надавати ширші можливості покупцям.

На наступній сторінці зображені побудовані моделі регресійного аналізу. Зокрема, модель AP(8):

$$\begin{aligned}
 Y = & 21875,23 + 0,08 * y(k - 1) - 0,64 * y(k - 2) - \\
 & -0,71 * y(k - 3) - 0,32 * y(k - 4) - 0,11 * y(k - 5) + \\
 & +0,04 * y(k - 6) + 0,18 * y(k - 7) + 0,2 * y(k - 8) \quad (4.1)
 \end{aligned}$$

На рис. 4.20 зображена модель AP(8).

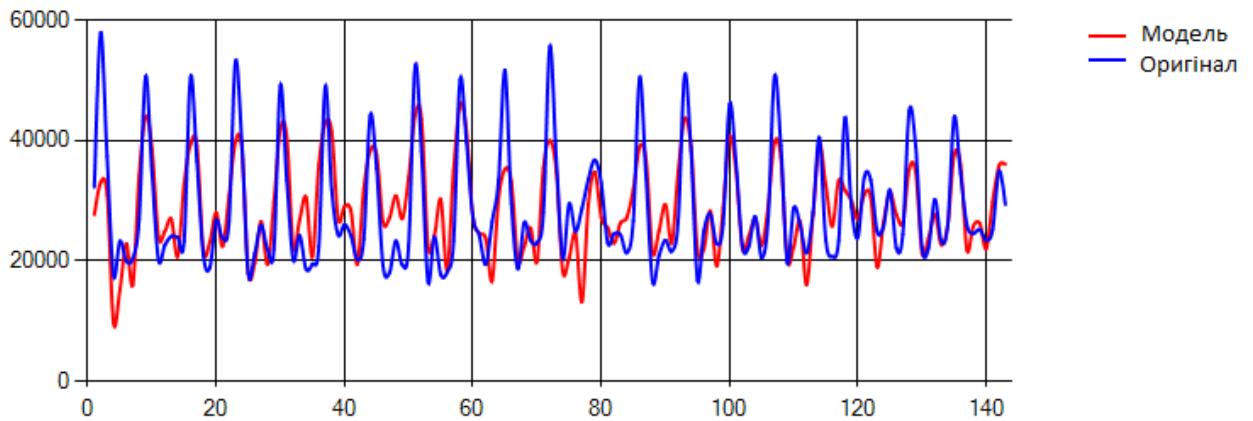


Рисунок 4.20 – АР (8) прибутковості

АРКС(8,3):

$$\begin{aligned}
 Y = & 21875,23 + 0,08 * y(k - 1) - 0,64 * y(k - 2) - \\
 & - 0,71 * y(k - 3) - 0,32 * y(k - 4) - 0,11 * y(k - 5) + \\
 & + 0,04 * y(k - 6) + 0,18 * y(k - 7) + 0,2 * y(k - 8) - 73,69 - \\
 & - 0,08 * ma(k - 1) + 1,17 * ma(k - 2) - 1,22 * ma(k - 3) \quad (4.2)
 \end{aligned}$$

Графік АРКС(8,3) зображено на рис. 4.21.

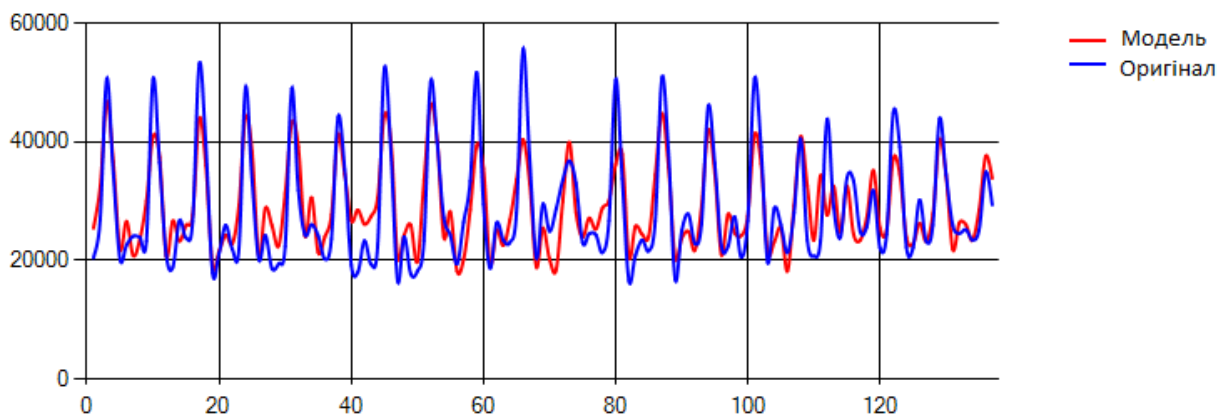


Рисунок 4.21 – АРКС(8,3) прибутковості

АРІКС(8,3,3):

$$\begin{aligned}
Y = & 2676,67 - 0,4 * \tilde{y}(k - 1) - 0,42 * \tilde{y}(k - 2) - \\
& - 0,41 * \tilde{y}(k - 3) - 0,41 * \tilde{y}(k - 4) - 0,42 * \tilde{y}(k - 5) - \\
& - 0,4 * \tilde{y}(k - 6) + 0,6 * \tilde{y}(k - 7) - 122,34 + 0,38 * ma(k - 1) + \\
& + 1,11 * ma(k - 2) - 0,47 * ma(k - 3) + \\
& + y(k) - 2 * y(k - 1) + y(k - 2) + y(k - 3)
\end{aligned} \tag{4.3}$$

Графік побудованої моделі зображено на рис. 4.22.

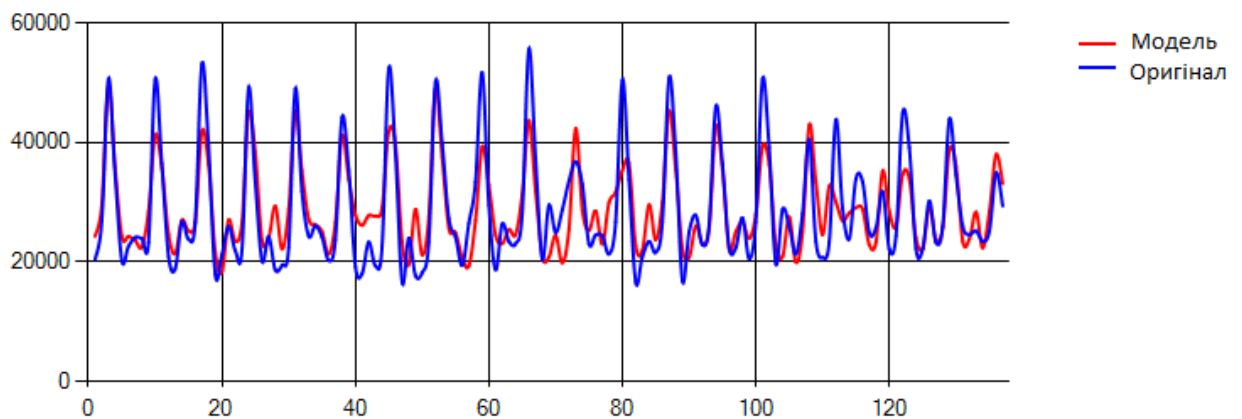


Рисунок 4.22 – АРІКС(8,3,3) прибутковості на навчальній вибірці

Оцінки адекватності для побудованих моделей відображені в таблицях 4.1-4.2.

Таблиця 4.1 – Оцінки моделі прибутковості

	Середньоквадратична похибка	Коефіцієнт детермінації	Акайке	Байєса-Шварца	Фішера	Дарбіна-Уотсона
АР(8)	6231,88	0,59	1259.55	1257.59	1.43	2.04
АРКС(8,3)	5090,52	0,54	1179,5	1177,55	1,152	1,92
АРІКС(8,8,3)	5091	0,54	1179,33	1177,37	1,152	1,92

Таблиця 4.2 – Оцінки моделі прибутковості

	Середньоквадратична похибка	Коеф. детермінації	Акайке	Байєса-Шварца	Фішера	Дарбіна-Уотсона
АР(8)	218.34	0.69	361.27	354.66	2.71	0.758
АРКС(8,3)	1524,76	0,93	358,84	358,97	11,47	1,1
АРІКС(8,3,8)	1530,31	0,85	360,14	359,15	11,47	1,1

Відповідно до оцінок моделей, найкращою є модель АРКС(8,3).

Висновки до розділу

Отже, було побудовано СППР зі зручним інтерфейсом та швидкими підрахунками аналізу прибутковості та споживчого кошику для підприємства роздрібною торгівлі. В систему включені такі елементи, як регресійний аналіз, кластерний аналіз, асоціативний аналіз та SWOT - аналіз.

РОЗДІЛ 5 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ

5.1 Ідея стартап - проекту

Ідею стартап проекту описано в таблиці 5.1.

Таблиця 5.1 - Опис ідеї стартап-проекту

Ідея	Напрямки використання	Вигоди для користувача
Розробка системи підтримки прийняття рішень для оптимізації прибутку підприємств роздрібною торгівлі	Роздрібна торгівля	Збільшення прибутків підприємства
	Розподілення запасів	Своєчасна закупіля актуальних продуктів торгівлі

Сильні, слабкі та нейтральні властивості ідеї проекту відображено в таблиці 5.2.

Таблиця 5.2 - Визначення характеристик ідеї проекту

п/п	Техніко-економічні характеристики ідеї	Потенційні) товари/концепції конкурентів			W (слабка сторона)	N (нейтр. сторона)	S (сильна сторона)
		FS	Sales Loft	Plan. Analyt			
.	Універсальний екстеншн даних	-	+	-	+		

Продовження таблиці 5.2

.	Кластерний аналіз	+	+	-			+
.	Асоц. аналіз	+	-	+			+
.	Моделювання і прогнозування	+	+	+		+	
.	Популярність бренду	+	-	-	+		

5.2 Технологічний аудит ідеї проекту

Технологічний аудит ідеї проекту наведений у таблиці 5.3.

Таблиця 5.3 - Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Дерева рішень	R	Наявна. Потрібно реалізувати алгоритм.	Загальнодоступна
2	Асоціативний аналіз	R	Необхідно реалізувати алгоритм Apriori.	Загальнодоступна

Продовження таблиці 5.3

3	Моделювання та прогнозування	R	Необхідно реалізувати моделі: AR, ARMA, ARIMA	Технологія загальнодоступна
Обрана технологія реалізації ідеї проекту: для реалізації проекту обрана мова програмування R.				

5.3 Аналіз ринкових можливостей запуску стартап-проекту

Властивості потенційного ринку стартап-проекту наведені у таблиці 5.4.

Таблиця 5.4 - Характеристика потенційного ринку стартап-проекту

п/п	Показники стану ринку (найменування)	Характеристика
	К-сть головних гравців, од.	5
	Сумарний обсяг продажів, грн/ум.од.	1000 ум.од
	Динаміка потенційного ринку проекту	Зростає
	Наявність обмежень для входу	Відсутні
	Специфічні вимоги до стандартизації та сертифікації	Відсутні
	Середня норма рентабельності в галузі (або по ринку), %	30%

Опис потенційних користувачів наведений в таблиці 5.5.

Таблиця 5.5 - Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія(цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Прогнозування та передбачення потреб користувачів у роздрібній торгівлі безпосередньо	Середній та великий бізнес, що застосовують системи керування ресурсів підприємства	ERP система підприємства, розміри отриманих даних, технологічні обмеження	Простий інтерфейс Результативність прогнозування Якісний аналіз даних Оптимальний розподіл ресурсів

Потенційні загрози для стартап-проекту наведені у таблиці 5.6.

Таблиця 5.6 - Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Неповнота даних	Клієнтські бази можуть містити невідповідні дані, дані помилкові	Додання модуля попередньої обробки інформації

Продовження таблиці 5.6

2	Дефіцит технічних засобів	Користувачі можуть мати недостатньо потужні технічні засоби для коректної роботи системи	Створення модуля розрахунків на віддалених серверах
---	---------------------------	--	---

Фактори можливостей наведені у таблиці 5.7.

Таблиця 5.7 - Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Хмарні калькуляції	Змога виконання обчислень на віддалених серверах	Пристосування модулів обчислення для роботи на сервері
2	Коригування прогнозу	Можливість виправлення прогнозу в режимі реального часу на основі існуючих даних	Розробка модулів інтеграції уснуючими системами аналітики, оптимізація використання для роботи в онлайн режимі

Проведений ступеневий аналіз конкуренції на ринку зображено у таблиці 5.8.

Таблиця 5.8 - Ступеневий аналіз конкуренції на ринку

Властивості конкурентного середовища	В чому безпосередньо проявляється дана властивість	Дія на підприємство (допустимі вчинки організації, щоб бути конкурентоспроможною)
1. Тип конкуренції - олігополія	Існує досить невелика кількість фірм.	Впровадження технологічних інновацій. Співпраця з
2. За рівнем конкурентної боротьби - глобальний	Продукт абсолютно не залежить від країни чи локалізації користувача	дослідницькими інститутами. Збільшення функціоналу та задоволення потреб користувачів.
3. За галузевою властивістю внутрішньогалузева	Продукт спрямований на роздрібну торгівлю	
5. Конкуренція за типами товарів: - за побажанням	Полягає у випередженні задоволення бажань клієнта	
5.3а За характером конкурентних переваг - нецінова	Переваги передбачають собою ефективність та різноманіття функціоналу	
6. За інтенсивністю - не марочна	Торгова марка майже не впливова	

Проаналізована конкуренція в галузі зображена в таблиці 5.9.

Таблиця 5.9 - Аналіз конкуренції в галузі

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти
	Microsoft, IBM, Analytics Inc	SlikPie
Висновки	Володіють великою частиною ринку, мають узагальнені рішення	Спрямовані на малий та середній бізнес, не мають локалізацій для більшості Європейських країн

Значимі фактори конкурентоспроможності та їх безпосереднє обґрунтування зображені в таблиці 5.10.

Таблиця 5.10 - Обґрунтування значимих факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування факторів (наведення впливових чинників, що роблять фактор для порівняння конкурентних систем вагомим)
1	Інновації	Інноваційні системи мають безпосередньо забезпечити вагому перевагу клієнтам над конкурентами
2	Функціонал	Функціонал має покривати розв'язання задач першої необхідності клієнтів
3	Цінова політика	Вартість системи відіграє велику роль при виборі продукту потенційним клієнтом
4	Ресурсоемність	Досить великі затрати технічних ресурсів призводять до необхідності залучення додаткових коштів

Порівняльний аналіз сильних та слабких сторін системи відображено у таблиці 5.11.

Таблиця 5.11 - Порівняльний аналіз сильних та слабких сторін RFS

п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з RFS						
			3	2	1		1	2	3
	Інновації	2							
	Функціонал	8							
	Цінова політика	7							
	Ресурсоемність	0							

SWOT-аналіз проекту наведено в таблиці 5.12.

Таблиця 5.12 - SWOT-аналіз стартап-проекту

Сильні сторони: самодостатня цінова політика, функціонал надає рішення більшості задач клієнта	Слабкі сторони: відсутність безпосередньої співпраці з інноваційними центрами
Можливості: впровадження новітніх розв'язків, повна оптимізація роботи системи	Загрози: недостатньо оптимізована робота системи, узагальнені розв'язки, недоцільні для задач користувача

Альтернативи ринкового заровадження проекту розглянуто в наступній таблиці 5.13.

Таблиця 5.13 - Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	В отримання ресурсів	Строки реалізації
1	Спеціалізовані рішення	Висока	3-4 місяців
2	Хмарний сервіс	Висока	6-9 місяців
3	Узагальнення рішення, вихід на нові сфери ринку	Середня	9-12 місяці

5.4 Розробка ринкової стратегії проекту

Опис та безпосередній вибір цільових груп потенційних клієнтів та користувачів зображено в таблиці 5.14.

Таблиця 5.14 - Вибір цільових груп потенційних користувачів

п/п	Опис цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
	Малий бізнес	Низька, через необхідність безпосереднього вкладання великих коштів.	Низький	Середня	Складний
	Середній бізнес	Середня. В залежності від виду підприємства або бізнесу, готовність відрізняється одна від одної.	Середній	Вище середнього	Достатньо складний
	Великий бізнес	Абсолютна. Допустима більшість організацій та підприємств такого рівня застосовували чи планують застосовувати системи прогнозування, через що готові розглядати альтернативні рішення	Високий	Середня	Складний
Які цільові групи обрано: 2, 3					

В таблиці 5.15 зображено вибір основної стратегії розвитку.

Таблиця 5.15 - Визначення основної стратегії розвитку

п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
	Розробка та створення додаткових функціональних модулів	Таргетні пропозиції бізнесу, проведення презентації функціональних рішень на ярмарках та конференціях	Відсутність безпосередньо аналогічних до новостворених функціональних модулів у конкурентів	Розробка, створення, планування та удосконалення існуючих модулів безпосередньо на основі потреб ринку та інформації від клієнтів

В таблиці 5.16 наведено визначення базової стратегії конкурентної поведінки.

Таблиця 5.16 - Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
1	Ні	Вірогідні обидва варіанти події	Основні функціональні модулі будуть виконувати суттєво схожі функції, що і конкуренти	Розумна цінова політика, функціональні інновації та нові модулі

В таблиці 5.17 наведено визначення основної стратегії позиціонування.

Таблиця 5.17 - Визначення стратегії позиціонування

п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
	Висока якість прогнозування в клієнтській сфері застосування	Розробка та удосконалення існуючих модулів на основі важливих потреб ринку та інформації від клієнтів та користувачів	Спеціалізовані базові рішення, хмарні сервіси, віддалені сервери	Прогнозування, передбачення, аналіз, побудова плану розвитку

5.5 Розроблення маркетингової програми стартап-проекту

В таблиці 5.18 зображені основні переваги потенційного товару.

Таблиця 5.18 - Визначення основних переваг потенційного товару

п/п	№ Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Широкий функціонал	Вирішення задач	Забезпечує вирішення більшої кількості важливих задач бізнесу
2	Спеціалізовані рішення	Вирішення задач	Забезпечує ефективніше вирішення задач у звуженій сфері застосування
3	Технічні ресурси	Хмарні сервіси	Дозволяє користуватись рішенням за рахунок віддалених технічних потужностей

Опис трьох рівнів моделі товару відображено у таблиці 5.19.

Таблиця 5.19 - Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за здумом	Аналіз даних. Прогнозування та передбачення потреб споживача		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	Швидкодія	Нм	Тх/Тл/Е
	Ефективність	Нм	Тх/Тл
	Користувацький інтерфейс	Нм	Е
	Якість: стандарти відповідні до законодавства. Створені функціональні скріпти.		
Пакування: Власний сайт			

Продовження таблиці 5.19

III. Товар із підкріпленням	До продажу: застосунок для інтеграції в існуючі системи керування підприємством для прогнозування та передбачення потреб споживачів на основі великих масивів даних
	Після продажу: Швидкодія, ефективність, легкість у користуванні
Закритий код. Захищений від можливості декомпіляції.	

Визначення меж встановлення ціни показано в таблиці 5.20.

Таблиця 5.20 - Визначення меж встановлення ціни

№ п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	-	200\$/міс	Рівень доходів підприємств надзвичайно високий	150-200\$/міс

Формування системи збуту зображено в таблиці 5.21.

Таблиця 5.21 - Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
	Таргетні пропозиції для компаній	Презентації функціоналу	-	-

Концепція маркетингових комунікацій відображена у таблиці 5.22.

Таблиця 5.22 - Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонува ння	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Середній бізнес – оптимальні рішення за невисоку ціну	Соціальні мережі, внутрішньо ринкова комунікація	Прогнозува ння покупок споживача	Короткий опис переваг продукту, заохочення дізнатись більше	Передбаче ння покупок споживачів

Продовження таблиці 5.23

2	Великий бізнес – повноцінні рішення для покращення продажів	Таргетні дзвінки до клієнтів	Прогнозування покупок споживача	Донести інформацію про оптимальність рішення для бізнесу клієнта	Прогнозування покупок клієнтів
---	---	------------------------------	---------------------------------	--	--------------------------------

Висновки до розділу

Отже, відповідно до отриманих вище результатів, можна робити висновок про існування попиту на запропоновану систему. Варто зауважити, що на ринку присутня сильна конкуренція, але інноваційна складова продукту дозволяє суттєво збільшити конкурентоспроможність проекту. Подальша імплементація проекту доцільна за умови фокусування на середній та великий бізнес, оскільки саме цей ринок динамічно зростає і є достатньо рентабельним.

ВИСНОВКИ

В даній роботі проведено розгляд прибутку підприємства роздрібною торгівлі як великої системи з внутрішніми та зовнішніми параметрами. Для оптимізації прибутку було запропоновано використовувати асоціативний, кластерний аналізи, побудову регресійних моделей та застосування SWOT – аналізу.

Результатом виконання роботи є система підтримки прийняття рішень, яка дозволяє робити швидкий аналіз даних та робити висновки, такі як доцільність покупки товарів на продаж, розміщення товарів в магазинах, проведення акцій та пропозицій, планувати майбутні продажі та враховувати сильні, слабкі сторони підприємства, а також мати змогу уникати загроз та використовувати можливості.

Використані методи запрограмовані на мові R та використані для комплексної візуалізації в сучасному додатку для звітів Power BI.

Система розроблена на прикладі даних підприємства роздрібною торгівлі. По результату роботи програми, було ідентифіковано найбільш прибуткові категорії товарів, проаналізовані групи покупців, спрогнозовано прибуток в короткостроковій перспективі та проаналізовано сильні та слабкі сторони типового підприємства роздрібною торгівлі, а також його загрози та можливості.

Подальший розвиток роботи полягає в додаванні інших моделей аналізу даних, використання ключових показників ефективності та порівняння їх з отриманими результатами за допомогою машинного навчання.

ПЕРЕЛІК ПОСИЛАНЬ

1. Багиев Г.Л. Маркетинг: підручник / Багиев Г.Л., Тарасевич В.М. – СПб. : Питер, 2010. – 576 с.
2. Коротков А. В. Управление маркетингом : навчальний посібник / Коротков А.В., Синяева И.М. – М. : Юнити-Дана, 2005. – 463 с.
3. Наумов В.Н. Маркетинговые решения в розничной торговле: навчальний посібник / Наумов В.Н., Кукура С.П. – СПб. : Политехника-сервис, 2008. – 198 с.
4. The story and concept of development of the payment system in the Eurasian economic union [Електронний ресурс]. – Режим доступу: <https://cyberleninka.ru/article/v/kontsepsiya-razvitiya-platezhnoy-sistemy-v-evraziyskom-ekonomicheskom-soyuze>
5. Дихтль Е. Практический маркетинг: навчальний посібник / Дихтль Е., Хершген Х. – М. : Высш.шк., 2007. – 255 с.
6. Бузукова Е. А. Ассортимент розничного магазина: методы анализа и практические советы / Бузукова Е. А. – СПб. : Питер, 2007. – 176 с.
7. Федько В. П. Основы маркетинга / Федько В. П., Федько Н. Г. – Ростов н/Д : Феникс, 2005. – 497 с.
8. Згуровський М.З. Основи системного аналізу: підручник / Згуровський М.З., Панкратова Н.Д. – К.: Видавнича група ВНУ, 2007. – 546 с.
9. Згуровський М.З. Принципи дослідження міждисциплінарних проблем / Згуровський М.З., Панкратова Н.Д. – К.: Наук. думка, 2005.-743 с.
10. Брагин Л. А. Торговое дело: экономика, маркетинг, организация: підручник / Брагина Л. А., Данько Т. П. – М. : Инфра-М, 2006. – 172 с.
11. Панкратов Ф. Г. Коммерческая деятельность: учебник для вузов/ Панкратов Ф. Г., Серегина Т. К. – М. : Дашков и К°, 2006. – 503 с.

12. Снегирева В. В. Розничный магазин: управление ассортиментом по товарным категориям / Снегирева В. В. – СПб. : Питер, 2005. – 416 с.
13. Гуняков Ю. В. Коммерция. Теория коммерческой деятельности / Гуняков Ю. В. – Красноярск : КГТЭИ, 2008. – 178 с.
14. Modern system of marketing communications and structural correlation of its elements [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/v/instrumenty-i-metody-upravleniya-assortimentom-tovarov-v-riteyle>
15. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/v/primenenie-tehnologiy-intellektualnogo-analiza-dannyh-v-estestvennonauchnyh-tehnicheskikh-i-gumanitarnyh-oblastyah>

ДОДАТОК А ЛІСТИНГ КОДУ

```

libraryRequireInstallWithWorkarounds = function(packageName, ...)
{
  print(paste("***", packageName, "***"))
  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "http://cran.us.r-project.org", ... )
    packOK = require(packageName, character.only = TRUE)
  }
  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "https://cloud.r-project.org", ... )
    packOK = require(packageName, character.only = TRUE)
  }

  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "http://cran.rstudio.com/", ... )
    packOK = require(packageName, character.only = TRUE)
  }
  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "https://cran.r-project.org", ... )
    packOK = require(packageName, character.only = TRUE)
  }

  if(!require(packageName, character.only = TRUE))
  {
    if(!(as.numeric(version$major)>=3 && as.numeric(version$minor)>= 3.0 ))
      setInternet2(use = TRUE)

    install.packages(packageName)
    packOK = require(packageName, character.only = TRUE)
  }
}

allpackages = c("scales", "reshape", "ggplot2", "htmlwidgets", "XML", "plotly")

sapply(allpackages, libraryRequireInstallWithWorkarounds)
module powerbi.extensibility.visual {
  export class Circle {
    private svg: d3.Selection<SVGElement>;
    private container: d3.Selection<SVGElement>;
    private circle: d3.Selection<SVGElement>;
    private textValue: d3.Selection<SVGElement>;
    private textLabel: d3.Selection<SVGElement>;

    constructor(element: HTMLElement) {
      this.svg = d3.select(element)
        .append('svg')
        .classed('circleCard', true);

      this.container = this.svg.append("g")
        .classed('container', true);

      this.circle = this.container.append("circle")
        .classed('circle', true);

      this.textValue = this.container.append("text")
        .classed("textValue", true);

      this.textLabel = this.container.append("text")
        .classed("textLabel", true);
    }

    public update(
      width: number,
      height: number,
      textValue: string,
      textLabel: string,
      color: string
    ) {
      this.svg.attr({
        width: width,

```

```

        height: height
    });

    let radius: number = Math.min(width, height) / 2.2;

    this.circle
        .style("fill", color)
        .style("fill-opacity", 0.5)
        .style("stroke", "black")
        .style("stroke-width", 2)
        .attr({
            r: radius,
            cx: width / 2,
            cy: height / 2
        });

    let fontSizeValue: number = Math.min(width, height) / 5;

    this.textValue
        .text(textValue)
        .attr({
            x: "50%",
            y: "50%",
            dy: "0.35em",
            "text-anchor": "middle"
        }).style("font-size", fontSizeValue + "px");

    let fontSizeLabel: number = fontSizeValue / 4;

    this.textLabel
        .text(textLabel)
        .attr({
            x: "50%",
            y: height / 2,
            dy: fontSizeValue / 1.2,
            "text-anchor": "middle"
        }).style("font-size", fontSizeLabel + "px");
    }
}

constructor(element: HTMLElement) {
    this.svg = d3.select(element)
        .append('svg')
        .classed('circleCard', true);

    this.container = this.svg.append("g")
        .classed('container', true);

    this.circle = this.container.append("circle")
        .classed('circle', true);

    this.textValue = this.container.append("text")
        .classed("textValue", true);

    this.textLabel = this.container.append("text")
        .classed("textLabel", true);
}

public update(
    width: number,
    height: number,
    textValue: string,
    textLabel: string,
    color: string
) {
    this.svg.attr({
        width: width,
        height: height
    });

    let radius: number = Math.min(width, height) / 2.2;

    this.circle
        .style("fill", color)
        .style("fill-opacity", 0.5)
        .style("stroke", "black")
        .style("stroke-width", 2)
        .attr({
            r: radius,
            cx: width / 2,
            cy: height / 2
        });

    let fontSizeValue: number = Math.min(width, height) / 5;

```

```

    this.textValue
      .text(textValue)
      .attr({
        x: "50%",
        y: "50%",
        dy: "0.35em",
        "text-anchor": "middle"
      }).style("font-size", fontSizeValue + "px");

    let fontSizeLabel: number = fontSizeValue / 4;

    this.textLabel
      .text(textLabel)
      .attr({
        x: "50%",
        y: height / 2,
        dy: fontSizeValue / 1.2,
        "text-anchor": "middle"
      }).style("font-size", fontSizeLabel + "px");
  }
}

libraryRequireInstallWithWorkarounds = function(packageName, ...)
{
  print(paste("***", packageName, "***"))
  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "http://cran.us.r-project.org", ... )
    packOK = require(packageName, character.only = TRUE)
  }
  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "https://cloud.r-project.org", ... )
    packOK = require(packageName, character.only = TRUE)
  }

  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "http://cran.rstudio.com/", ... )
    packOK = require(packageName, character.only = TRUE)
  }
  if(!require(packageName, character.only = TRUE))
  {
    install.packages(packageName, repos = "https://cran.r-project.org", ... )
    packOK = require(packageName, character.only = TRUE)
  }

  if(!require(packageName, character.only = TRUE))
  {
    if(!(as.numeric(version$major)>=3 && as.numeric(version$minor)>= 3.0 ))
      setInternet2(use = TRUE)

    install.packages(packageName)
    packOK = require(packageName, character.only = TRUE)
  }
}

allpackages = c("scales", "reshape", "ggplot2", "htmlwidgets", "XML", "plotly")

sapply(allpackages, libraryRequireInstallWithWorkarounds)
module powerbi.extensibility.visual {
  export class Circle {
    private svg: d3.Selection<SVGElement>;
    private container: d3.Selection<SVGElement>;
    private circle: d3.Selection<SVGElement>;
    private textValue: d3.Selection<SVGElement>;
    private textLabel: d3.Selection<SVGElement>;

    constructor(element: HTMLElement) {
      this.svg = d3.select(element)
        .append('svg')
        .classed('circleCard', true);

      this.container = this.svg.append("g")
        .classed('container', true);

      this.circle = this.container.append("circle")
        .classed('circle', true);
    }
  }
}

```

```

    this.textValue = this.container.append("text")
      .classed("textValue", true);

    this.textLabel = this.container.append("text")
      .classed("textLabel", true);
  }

  public update(
    width: number,
    height: number,
    textValue: string,
    textLabel: string,
    color: string
  ) {
    this.svg.attr({
      width: width,
      height: height
    });

    let radius: number = Math.min(width, height) / 2.2;

    this.circle
      .style("fill", color)
      .style("fill-opacity", 0.5)
      .style("stroke", "black")
      .style("stroke-width", 2)
      .attr({
        r: radius,
        cx: width / 2,
        cy: height / 2
      });

    let fontSizeValue: number = Math.min(width, height) / 5;

    this.textValue
      .text(textValue)
      .attr({
        x: "50%",
        y: "50%",
        dy: "0.35em",
        "text-anchor": "middle"
      }).style("font-size", fontSizeValue + "px");

    let fontSizeLabel: number = fontSizeValue / 4;

    this.textLabel
      .text(textLabel)
      .attr({
        x: "50%",
        y: height / 2,
        dy: fontSizeValue / 1.2,
        "text-anchor": "middle"
      }).style("font-size", fontSizeLabel + "px");
  }
}

constructor(element: HTMLElement) {
  this.svg = d3.select(element)
    .append('svg')
    .classed('circleCard', true);

  this.container = this.svg.append("g")
    .classed('container', true);

  this.circle = this.container.append("circle")
    .classed('circle', true);

  this.textValue = this.container.append("text")
    .classed("textValue", true);

  this.textLabel = this.container.append("text")
    .classed("textLabel", true);
}

public update(
  width: number,
  height: number,
  textValue: string,
  textLabel: string,
  color: string
) {
  this.svg.attr({
    width: width,
    height: height
  });
}

```



```
});  
  
let radius: number = Math.min(width, height) / 2.2;  
  
this.circle  
  .style("fill", color)  
  .style("fill-opacity", 0.5)  
  .style("stroke", "black")  
  .style("stroke-width", 2)  
  .attr({  
    r: radius,  
    cx: width / 2,  
    cy: height / 2  
  });  
  
let fontSizeValue: number = Math.min(width, height) / 5;  
  
this.textValue  
  .text(textValue)  
  .attr({  
    x: "50%",  
    y: "50%",  
    dy: "0.35em",  
    "text-anchor": "middle"  
  }).style("font-size", fontSizeValue + "px");  
  
let fontSizeLabel: number = fontSizeValue / 4;  
  
this.textLabel  
  .text(textLabel)  
  .attr({  
    x: "50%",  
    y: height / 2,  
    dy: fontSizeValue / 1.2,  
    "text-anchor": "middle"  
  }).style("font-size", fontSizeLabel + "px");  
}  
}
```