НАЦІОНАЛЬНИЙ УНІВЕРСТИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

УДК 004.9:005.912

«ПОГОДЖЕНО»	**«ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ»**
Декан факультету інформаційних технологій	Завідувач кафедри комп'ютерних наук
Болбот І.М., д.т.н., професор	Голуб Б.Л., к.т.н., доцент
_____	_____
____ _____ **2024 р.**	____ _____ **2024р.**

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему Система підтримки прийняття рішень для керівництва платформою з продажу кави

Спеціальність 122 «Комп'ютерні науки»

_____
(код і назва)

Освітня програма «Інформаційно управляючі системи та технології»

_____
(назва)

Орієнтація освітньої програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Гарант освітньої програми

Кандидат технічних наук, доцент _____ Голуб Белла Львівна
(науковий ступінь та вчене звання) (підпис) (ПІБ)

Керівник магістерської кваліфікаційної роботи

Кандидат технічних наук, доцент _____ Голуб Белла Львівна
(науковий ступінь та вчене звання) (підпис) (ПІБ)

Виконав _____ Мамонтова Діана Віталіївна
(підпис) (ПІБ студента)

КИЇВ-2024

Зміст

СПИСОК УМОВНИХ ПОЗНАЧЕНЬ

СППР – Система підтримки прийняття рішень

СД – сховище даних

UML – Unified Modeling Language

OLAP – On-Line Analytical Processing

MOLAP – Multidimensional On-Line Analytical Processing

ROLAP – Relational On-Line Analytical Processing

HOLAP - Hybrid On-Line Analytical Processing

SSAS – SQL Server Analysis Services

SSIS - SQL Server Integration Services

DSV – Data Source View

ВСТУП

Розвиток культури споживання кави в Україні, з появою нових кав'ярень, спеціалізованих магазинів та обжарювальних, сприяє формуванню стійкого ринку та зростанню попиту на різні види кави. Кава - один з найпопулярніших продуктів у всьому світі. Вона є актуальною для бізнесу в Україні з кількох причин. Вона залишається одним із найпопулярніших напоїв у світі, і Україна не є винятком. Зростаючий попит на якісну каву серед українських споживачів створює значні можливості для бізнесу. Ринок кави пропонує широкий асортимент продуктів, що дозволяє бізнесам задовольняти різні смакові уподобання та потреби клієнтів, збільшуючи таким чином свій потенційний ринок.

Зростання продажів призводить до збільшення кількості підриємств з продажу кави. А це спонукає підприємцям шукати нові можливості щодо збільшення продажу. Прогнозування продажів кави є критично важливим для бізнесу в Україні. Це дозволяє ефективно планувати ресурси, уникати надлишків або дефіциту продукції, зменшувати витрати та підвищувати прибутковість. Визначення найбільш популярних та прибуткових видів кави сприяє оптимізації асортименту, дозволяючи зосередитися на найбільш перспективних продуктах. Прогнозування також допомагає виявляти нові тренди та зміни в споживчих вподобаннях, що дає змогу бізнесу швидко адаптуватися до ринкових умов та випереджати конкурентів.

Створення системи підтримки та прийняття рішень (СППР) [1] для керівництва платформою з продажу кави є важливим кроком. СППР допомагає керівництву приймати більш обґрунтовані рішення на основі аналізу даних, що знижує ризики суб'єктивних помилок та підвищує ефективність управлінських рішень. Використання DSS дозволяє аналізувати ринкові тенденції, виявляти нові можливості для зростання та швидко адаптуватися до змін у споживчих уподобаннях та конкурентному середовищі. Крім того, DSS оптимізує різні операційні процеси, такі як управління

запасами, логістика, маркетинг та продажі, що знижує витрати та підвищує ефективність роботи платформи.

В даній роботі було використано методи Mining Structure [2] для прогнозування продажів кави, які мають кілька важливих переваг. Вони дозволяють аналізувати великі обсяги даних про споживачів, їхні звички, уподобання та поведінку, що допомагає виявити, яка кава користується найбільшою популярністю серед різних сегментів клієнтів. Знання про те, яка кава є прибутковішою, дозволяє оптимізувати асортимент продукції, зосереджуючи зусилля на виробництві та просуванні найбільш популярних і прибуткових видів кави. Використовуючи методи Data Mining, компанії можуть створювати персоналізовані маркетингові кампанії, спрямовані на конкретні групи споживачів, що підвищує ефективність реклами та сприяє збільшенню продажів. Крім того, Data Mining допомагає виявляти нові тренди на ринку кави, що дозволяє компаніям швидко адаптуватися до змін і випереджати конкурентів.

Таким чином, актуальність кави для бізнесу в Україні, важливість прогнозування продажів, створення систем підтримки прийняття рішень і використання методів Data Mining обумовлені необхідністю ефективного управління, оптимізації ресурсів, задоволення споживчого попиту та забезпечення конкурентних переваг.

Об’єктом дослідження є платформа з продажу кави.

Предметом дослідження є система підтримки прийняття рішень (СППР) для керівництва платформою з продажу кави.

Метою дослідження є проведення аналізу недоліків, що існують у платформах з продажу кави та підвищення продажу кавової продукції. Отримані результати дослідження та впровадження системи підтримки прийняття рішень стануть основою для підвищення продуктивності бізнес-процесів та досягнення стратегічних цілей платформи в умовах сучасного ринкового середовища.

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Опис предметної області

Кава є одним із найпопулярніших товарів на світовому ринку, і конкуренція у цій галузі є надзвичайно високою. Для ефективного управління продажами кави керівництво стикається з низкою викликів, таких як коливання попиту, зміни ринкових трендів, сезонність, вподобання клієнтів, а також потреба швидко реагувати на зміни умов постачання та цін.

Процеси постачання і продажу кави в Україні є багатоступеневими та охоплюють кілька ключових етапів, від імпорту до реалізації кінцевому споживачеві. Кава імпортується з традиційних країн-виробників, таких як Бразилія, Колумбія, Ефіопія, Гондурас, Гватемала, Ямайка, Танзанія, Сальвадорк та інші. Найпоширеніші сорти включають арабіку та робусту.

Зелена кава або вже обсмажені зерна транспортуються морськими шляхами, що є найбільш економічним способом доставки. Порти в Європі часто використовуються як проміжні точки, звідки кава направляється в українські порти, наприклад, в Одесу чи Миколаїв.

Імпортовані зелені зерна можуть бути обсмажені локально для забезпечення свіжості продукту. У багатьох випадках компанії адаптують ступінь обсмажування до локальних смаків споживачів. Зерна або мелену каву фасують у пакети різного об'єму, від дрібних порцій для побутових покупців до великих упаковок для бізнесів (кав'ярні, ресторани, офіси).

Продаж кави здійснюється через офлайн-канали, такі як супермаркети та кафе, а також через онлайн-платформи, забезпечуючи клієнтам як персоналізоване обслуговування, так і зручність замовлень з доставкою. Офлайн-продажі Включають мережі супермаркетів, спеціалізовані кавові магазини, кафе та ресторани. Офлайн-продажі дозволяють споживачам отримати консультації, перевірити товар на місці та скористатися додатковими послугами, такими як дегустації.

Щодо онлайн-продажів, продаж відбувається в інтернет-магазинах та платформах електронної комерції. Онлайн-продажі останніми роками демонструють значне зростання через зручність замовлень, доставку додому та можливість вибору широкого асортименту товарів.

Оскільки процеси відстеження товару охоплюють різні етапи – від замовлень і постачання до продажів і аналізу поведінки споживачів – збір і обробка великих обсягів інформації є складними без використання сучасних інформаційних технологій. Обробка інформації вручну потребує значних ресурсів та підвищує ризики помилок, що може вплинути на ефективність управління процесами. Для вирішення цих проблем компанії використовують OLAP (Online Analytical Processing) системи. OLAP-системи дозволяють інтерактивно аналізувати великі обсяги даних з різних джерел і швидко отримувати детальну аналітику. Вони забезпечують багатовимірний аналіз, який дозволяє глибше розуміти бізнес-процеси, виявляти тенденції і закономірності, необхідні для прийняття стратегічних рішень. Використання OLAP сприяє більш ефективному управлінню постачаннями, оптимізації складів та підвищенню продуктивності за рахунок кращого розуміння даних у реальному часі.

У сучасних умовах без використання технологій автоматизації та аналітики успішне управління постачанням та продажем кави стає майже неможливим. Інформаційні технології допомагають оптимізувати процеси, зменшити витрати, підвищити ефективність і забезпечити точне прогнозування.

1.2 Аналіз існуючих рішень та їх проблем

Система підтримки прийняття рішень (СППР) – це інформаційна система, яка допомагає в ухваленні рішень за рахунок збору, обробки й аналізу даних. Вона поєднує аналітичні інструменти та доступ до великих обсягів інформації для надання рекомендацій чи допомоги в процесі прийняття рішень. СППР можуть включати моделювання сценаріїв, прогнозування

результатів та інтерактивні звіти для оптимізації рішень у бізнесі, медицині, фінансах тощо.

Під час розробки СППР для керівництва платформою з продажу кави була проаналізована система, яка використовувалася під час роботи у кав’ярні. У ній використовувалася проста система продажу, яка мала кілька суттєвих обмежень. Продажі кавових зерен не відстежувалися автоматично, а вся інформація базувалася лише на накладних. Це створювало незручності для власника, оскільки він не міг ефективно контролювати обсяги продажів або швидко реагувати на зміни попиту. Брак автоматизації ускладнював прогнозування можливих продажів і визначення, яка саме кава є найбільш популярною серед клієнтів.

Кав'ярня пропонувала різноманітні сорти кави, але з часом віддала перевагу продукції українського обсмажчика Penyora. Проте відстеження руху товарів і загальної ефективності продажів залишалося ручним процесом, що обмежувало можливості для розвитку і підвищення ефективності бізнесу.

Можна зробити висновок, що в Україні не використовуються СППР для керівництва з продажу кави. Це може бути через те, що для малих підприємств інвестиції в СППР можуть виглядати занадто дорогими, особливо якщо їхні масштаби продажів обмежені. Витрати на розробку, впровадження та обслуговування таких систем можуть бути значними, і малий бізнес може вважати, що ці інвестиції не окупляться.

Так само багато компаній, особливо в малому та середньому бізнесі, все ще використовують традиційні методи обробки даних, такі як прості бухгалтерські програми чи бази даних. Інвестиції в СППР можуть бути сприйняті як занадто дорогі або складні для впровадження.

Загалом існуючі рішення для системи підтримки і прийняття рішень у сфері продажу кави не завжди розробляються спеціально для цієї галузі, але є багато загальних СППР платформ, які можна адаптувати для потреб бізнесу з продажу кави.

Метою аналізу є визначення сильних і слабких сторін існуючих рішень, виявлення можливостей для їхнього вдосконалення, а також визначення потенційних областей для нової інформаційно-аналітичної системи, що розробляється в рамках цієї дипломної роботи.

Salesforce - популярна платформа, яка допомогає компаніям автоматизувати процеси продажу, управління відносинами з клієнтами та збором даних [3]. Ця платформа пропонує інструменти для прогнозування та аналітики даних, що дозволяє компаніям виявляти нові можливості для продажу будь-якої продукції.

Перевагами цієї платформи насамперед є широкий функціонал. Salesforce надає широкий спектр інструментів, серед яких управління взаємовідносинами з клієнтами, автоматизація маркетингу, прогнозування продажів та аналітика. Також ключовою перевагою платформи є її гнучкість та можливості кастомізації. Вона дозволяє легко адаптувати процеси під конкретні потреби бізнесу, підтримуючи розширення через AppExchange і надаючи широкі можливості для налаштування. Головною перевагою є те, що Salesforce має потужні інструменти для аналітики, включаючи візуалізацію даних, створення кастомних звітів і дашбордів.

З недоліків цієї платформи можна виділити високу вартість. Salesforce є досить дорогою платформою, особливо для малого бізнесу або стартапів. Вартість може зростати при додаванні додаткових модулів і користувачів. Проблема з продуктивністю є дуже вагомим недоліком, адже при роботі з великим обсягом даних або складними звітами можуть виникати затримки і зниження продуктивності. Що стосується зручності використання, Salesforce не завжди може забезпечити інтуїтивно зрозумілий інтерфейс, що може створювати труднощі для нових користувачів і вимагати додаткового навчання для повного освоєння функціоналу платформи.

Інтерфейс даної платформи представлений на рис. 1

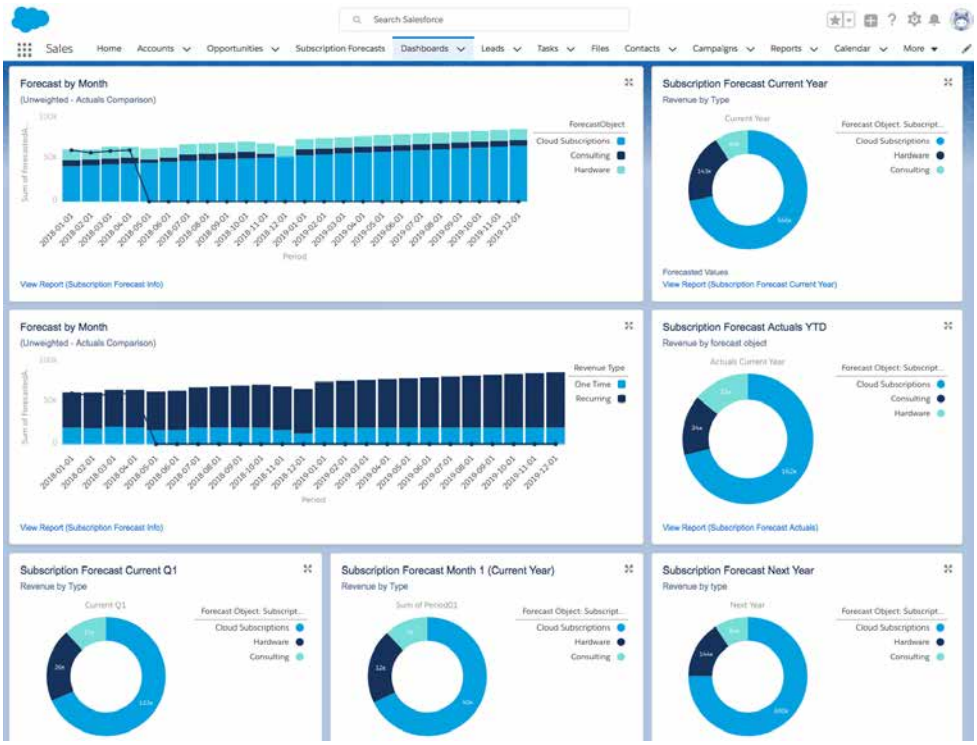


Рис. 1 Скріншот програми Salesforce

HubSpot - ця CRM-система містить функціонал для аналізу продажів, управління взаєминами з клієнтами та прогнозування, що дозволяє керівництву компанії отримувати важливу інформацію для прийняття рішень[4].

HubSpot має інтуїтивно зрозумілий і простий інтерфейс, що дозволяє легко користуватися платформою навіть новачкам. Це робить процес налаштування і використання CRM швидким і доступним. Вона поєднує в собі CRM, інструменти для маркетингу, продажів, підтримки клієнтів і управління контентом. Це дозволяє використовувати один інструмент для всіх аспектів управління бізнесом і взаємодії з клієнтами. Платформа дозволяє налаштовувати маркетингові кампанії, автоматизувати email-розсилки, налаштувати лійки продажів і відстежувати поведінку клієнтів. Це допомагає оптимізувати роботу з потенційними клієнтами та підвищити ефективність маркетингових активностей.

Недоліками даної системи є те, що у ній обмежена кількість контактів для маркетингових кампаній, обсяг автоматизації та можливість створення складних робочих процесів. HubSpot добре підходить для малого та

середнього бізнесу, але великим підприємствам з великою кількістю користувачів і контактів може бути складно масштабувати платформу через високу вартість та деякі функціональні обмеження. Інтерфейс даної системи представлений на рис. 2.

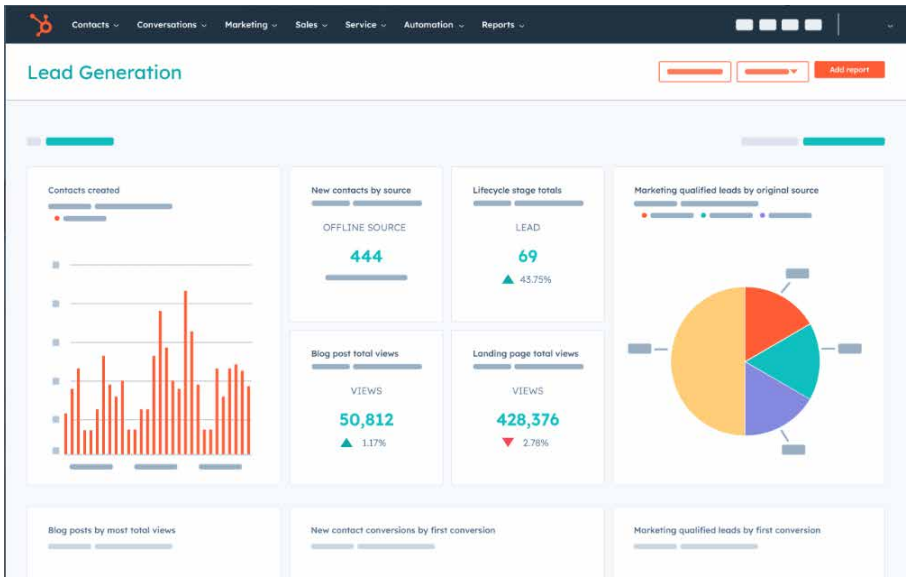


Рис. 2 Скріншот програми HubSpot

1.3 Постановка завдання для магістерської роботи

У межах магістерської роботи необхідно розробити систему для підтримки прийняття рішень у сфері продажу кави. Для цього слід провести всебічний аналіз існуючих рішень, виявити їхні переваги та недоліки, визначити можливості для вдосконалення. На основі отриманих даних буде визначено вимоги до нової системи та розроблено її функціональну архітектуру. Для досягнення цієї мети передбачено виконання кількох конкретних завдань.

Перше завдання включає в проведенні огляду та ретельного аналізу поточного стану системи продажу кави. Цей етап включає дослідження існуючих методів і підходів, що вже використовуються на практиці.

Друге завдання полягає в визначенні основних факторів, що впливають на ефективність продажу кави та їхній вплив на роботу підприємств та бізнесу.

Цей етап включає аналіз внутрішніх і зовнішніх чинників, що впливають на процес продажу, а також виявлення потенційних проблем.

Третє завдання полягає в розробці конкретних рекомендацій для підвищення якості та ефективності управління продажами кави на підприємствах і в бізнесі на основі результатів аналізу. Ця частина дослідження включає створення практичних стратегій та пропозицій для оптимізації процесу продажу кави.

Ця магістерська робота спрямована на розробку новаторського підходу до управління продажами кави, який допоможе підприємству та бізнесу підвищити ефективність управління ресурсами і посилити конкурентні позиції на ринку.

Ситема має включати оперативне джерела даних у вигляді оперативної бази даних та багатовимірного кубу.

Згідно з вищепоставленими задачами для магістерської роботи планується застосування OLAP технології для аналізу продажів кави. Ця технологія надає можливість багатовимірного дослідження даних, що є суттєвим для отримання повної картини продажів кави. Використання OLAP дає змогу аналізувати дані з різних перспектив, виявляти зв'язки та тренди, що сприятиме кращому розумінню ключових параметрів продажів і формуванню ефективних стратегій для їх покращення.

Окрім того, для досягнення цілей дослідження передбачається застосування методів інтелектуального аналізу даних Data Mining, які є важливим компонентом магістерської роботи. Використання методів Data Mining дозволить провести глибокий аналіз і моделювання даних продажів кави для виявлення різноманітних закономірностей, кореляцій і складних взаємозв'язків між різними параметрами. Ці методи допоможуть не тільки виявити загальні ознаки у даних, але й розпізнати приховані шаблони та тенденції, що сприятиме формулюванню конкретних рекомендацій для покращення якості та ефективності управління продажами кави в підприємствах і бізнесі.

2 МОДЕЛЮВАННЯ СИСТЕМИ

2.1 Загальні відомості про моделювання

Unified Modeling Language (UML) – це стандартна мова моделювання, яка використовується для опису, проектування і документування ПЗ та процесів [5]. Вона ідеально підходить для відображення складних систем у вигляді діаграм, що дозволяє розробникам, аналітикам і менеджерам розуміти й узгоджувати архітектуру та логіку роботи систем.

UML надає різні типи діаграм, які можна використовувати для моделювання різних аспектів системи. Найпопулярнішим видом діаграм UML є Діаграми прецедентів. Вони використовуються для моделювання взаємодії користувачів із системою. Вони можуть показувати, як керівники або користувачі платформи взаємодіють з функціями продажу кави. Діаграми класів моделюють структуру системи, показуючи класи, їхні атрибути, методи та зв'язки між ними. Це корисно для моделювання даних, які будуть використовуватися в платформі для продажу кави[6]. Діаграми послідовностей показують, як об'єкти системи взаємодіють один з одним під час виконання конкретних операцій, наприклад, процесу покупки кави на платформі[7]. Діаграми активності відображають процеси або потоки роботи в системі, такі як обробка замовлення на платформі для продажу кави[8].

Отже UML є потужним інструментом для моделювання систем, і його застосування в розробці СППР для керівництва платформи з продажу кави. Допомагає чітко структурувати процеси, знизити ризики помилок при розробці, а також забезпечити ефективну комунікацію серед усіх учасників проєкту. UML дозволяє детально моделювати як програмні, так і бізнес-процеси, що є важливим етапом у розробці ефективних та масштабованих інформаційних систем.

2.2 Діаграма прецедентів

Діаграма прецедентів є важливим інструментом моделювання в рамках UML, який допомагає представити динамічну поведінку системи. Вона дозволяє узагальнено описати функціональність системи шляхом включення варіантів використання, акторів та їх взаємозв'язків[9].

Актори в діаграмі прецедентів – це користувачі системи, в якій кожен актор виконує свою роль, яка називається варіантом використання, тобто сценарієм або поведінковим шаблоном, який описує взаємодію між актором та системою для досягнення певної цілі. Один варіант використання може виконуватися кількома акторами. Актором може бути людина, наприклад, клієнт, або комп'ютер, наприклад, система баз даних або сервер[10].

У результаті аналізу було створено діаграму, зображено на рис. 3

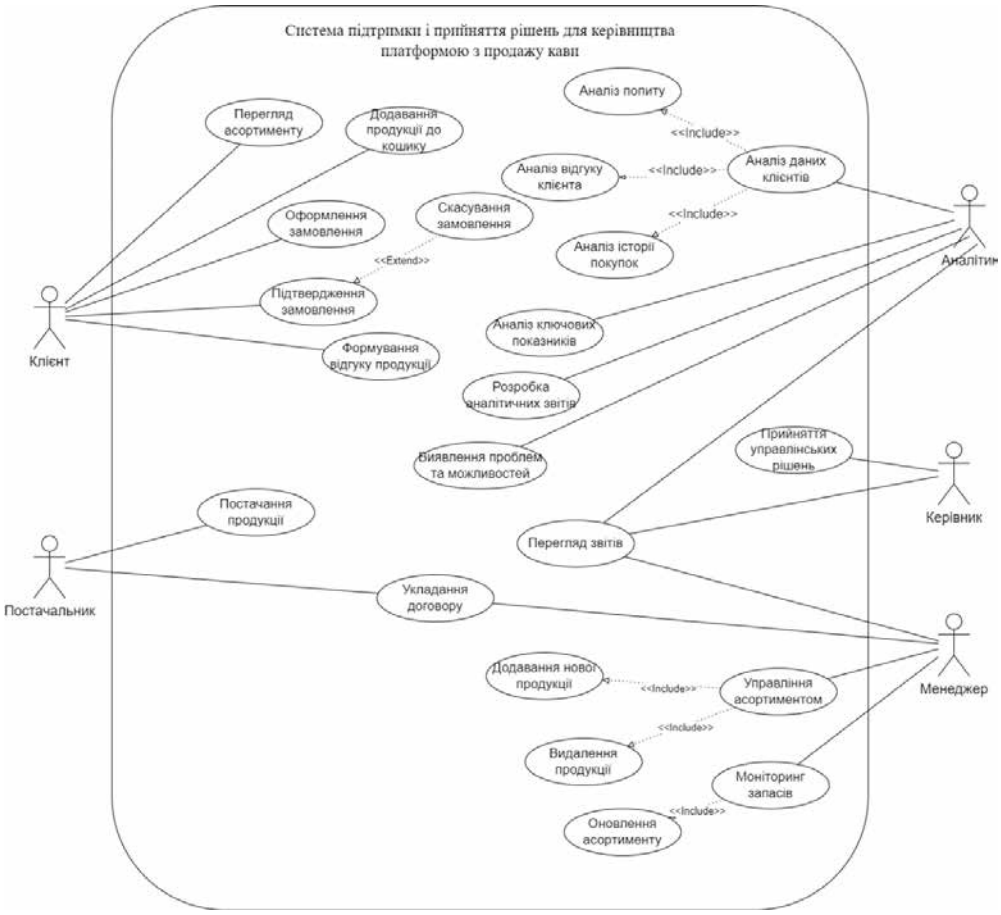


Рис. 3 Діаграма прецедентів

Актори, такі як «Клієнт», «Постачальник», «Керівник», «Менеджер» представляють собою основні ролі в системі. Кожен актор має свої власні прецеденти, функціональні операції, які він може виконувати. Для прикладу

візьмемо актора «Клієнт», який може переглядати асортимент та створювати замовлення, «Постачальник» може постачати продукцію та укладати договір, а «Керівник» в свою приймає управлінські рішення та переглядає звіти.

Діаграма також зображує, що система пітримує аналітичні можливості, представлені «Аналітиком». Він може аналізувати дані клієнтів, історії покупок, попит та відгуки клієнтів на той чи інший товар.

Отже, діаграма прецедентів у розробці системи слугує для моделювання вимог користувачів та їхньої взаємодії із системою. Вона допомагає зрозуміти, які функції системи необхідні для різних типів користувачів, і визначити сценарії їх використання. У даному дослідженні будемо використовувати діаграму прецедентів для аналізу та проектування системи. Це спрощує процес проектування та дає можливість розробити більш ефективну систему з точки зору функціональності та користувацької взаємодії.

2.2 Архітектура системи

Для системи, яка розробляється було прийнято рішення обрати тип архітектури, яка називається клієнт-серверна.

У клієнт-серверній архітектурі кожен комп'ютер або процес у мережі виконує роль клієнта або сервера. Сервери зазвичай є потужними комп'ютерами, спеціально призначеними для надання ресурсів, таких як друкарні, дискові накопичувачі та обробка мережевого трафіку. Клієнти, з свого боку, працюють на робочих станціях або персональних комп'ютерах і виконують програми для взаємодії з серверами.

Клієнти і сервери спілкуються між собою через комунікаційний канал, який може бути фізичною мережею передачі даних, такою як Інтернет або локальна мережа. Клієнти звертаються до серверів для доступу до ресурсів, отримання послуг або обробки даних, а сервери відповідають на їх запити та надають необхідні ресурси.

Ця архітектура дозволяє розділити функціональність та завдання між клієнтами та серверами, забезпечуючи більшу масштабованість, безпеку та

ефективність системи. Клієнтська частина зазвичай відповідає за інтерфейс користувача та взаємодію з користувачем, тоді як серверна частина забезпечує обробку даних, логіку бізнес-процесів та надання ресурсів[11].

На рис. 4 зображено архітектуру системи, вона має такі вузли:

1. Робочі станції клієнта, аналітики, менеджера та керівника: Ці компоненти представляють робочі місця користувачів, які взаємодіють з системою.
2. Два сервера з базою даних та сховищем даних: Ці сервери відповідають за зберігання та обробку інформації. База даних містить структуровану інформацію, а сховище даних в свою чергу – неструктуровану інформацію.

Зв’язки між компонентами:

1. Робоча станція менеджера підключена до сервера за допомогою мережі. Менеджер використовує робоче місце для внесення даних до сервера.
2. Сервер підключений до робочого місця аналітика за допомогою мережі. Аналітик використовує робоче місце для прогнозів продажу кави та генерування звітів.

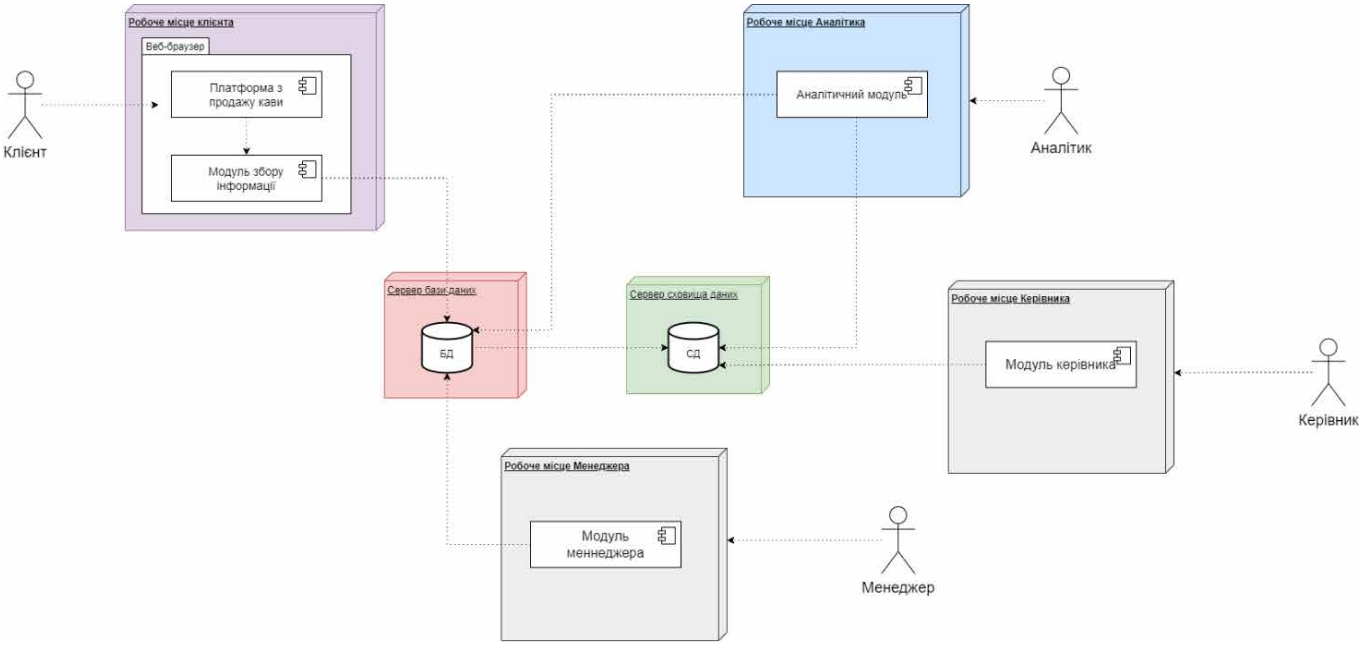


Рис. 4 Архітектура системи

3 РОЗРОБКА СИСТЕМИ

3.1 Структура джерела інформації для проведення інтелектуального аналізу

3.1.1 Структура сховища даних.

Сховище даних СППР для керівництва платформою з продажу кави є основним джерелом даних для виконання аналізу. Сховище даних — це централізоване сховище, яке використовується для зберігання великих обсягів структурованих даних, що походять з різних джерел, таких як транзакційні системи, ERP, CRM, тощо [12]. Воно використовується як оперативне джерело, оскільки в ньому зберігаються актуальні дані, що регулярно оновлюються і можуть бути використані для аналізу, звітності та прийняття рішень в реальному часі. Завдяки інтеграції з різними джерелами даних воно забезпечує централізований доступ до необхідної інформації з усіх важливих бізнес-процесів.

В основі концепції сховища даних (СД) лежить розподіл інформації, що використовують в системах оперативної обробки даних (OLTP) і в системах підтримки прийняття рішень (СППР). Такий розподіл дозволяє оптимізувати як структури даних оперативного зберігання для виконання операцій введення, модифікації, знищення та пошуку, так і структури даних, що використовуються для аналізу. В СППР ці два типи даних називаються відповідно оперативними джерелами даних (ОДД) та сховищем даних.

Сховища даних — основа для побудови систем підтримки прийняття рішень. Основна мета створення сховища в тому, щоб зробити усі значимі для управління бізнесом дані доступними в стандартизованій формі, придатними для аналізу та отримання необхідних звітів. Для досягнення цього потрібно отримати дані із існуючих внутрішніх та зовнішніх, доступних для комп'ютера, джерел. Незважаючи на відмінності в підходах та реалізаціях, усім сховищам даних властиві такі спільні риси: предметна орієнтованість, інтегрованість, прив'язка до часу, незмінність.

Перш ніж потрапити до сховища даних оперативні дані перевіряють, очищують та певним чином агрегують.

Під час розробки системи було створено сховище даних, яке дозволить проводити аналіз у різних розрізах. Структура сховища даних зображена на рис. 5. Скрипти запитів створення всіх таблиць СД наведено у додатку А.

Для збереження необхідних даних у СД були розроблені такі таблиці:

- ProductDim – вимір, що містить інформацію про продукт з такими полями:
  - IdProduct – ідентифікатор продукту.
  - IdProductCategory – ідентифікатор категорії продукту.
  - name – назва продукту.
  - description – опис продукту.
  - price – ціна продукту.
  - weight – вага продукту.
- ProductCategoryDim – вимір, що містить інформацію про категорію продукту з такими полями:
  - IdProductCategory – ідентифікатор категорії продукту.
  - sort – сорт кави.
  - roasting – обсмаження кави.
  - form – форма кави.
  - type – тип пакування.
- RegionDim – вимір, що містить інформацію про регіон кави з такими полями:
  - IdRegion - ідентифікатор регіону продукту;
  - name – назва регіону;
  - country – країна;
- SupplierDim - вимір, що містить інформацію про постачальника кави з такими полями:
  - IdSupplier - ідентифікатор постачальника продукції;
  - nameCompany - назва компанії;

- contactPerson – контактна особа;
- address – адреса;
- phone – номер телефону;

- DateDim - вимір, що зберігає часовий проміжок:
  - IdDate – ідентифікатор дати;
  - year – рік;
  - month – місяць;
  - day – день;
- SoldProductFact - таблиця фактів про зафіксовану кількість проданої продукції та суму з такими полями:
  - IdDate - ідентифікатор дати;
  - IdProduct - ідентифікатор продукту;
  - IdRegion - ідентифікатор регіону продукту;
  - IdSupplier - ідентифікатор постачальника продукції;
  - number_of_units – кількість проданої продукції;
  - suma – сума проданої продукції;

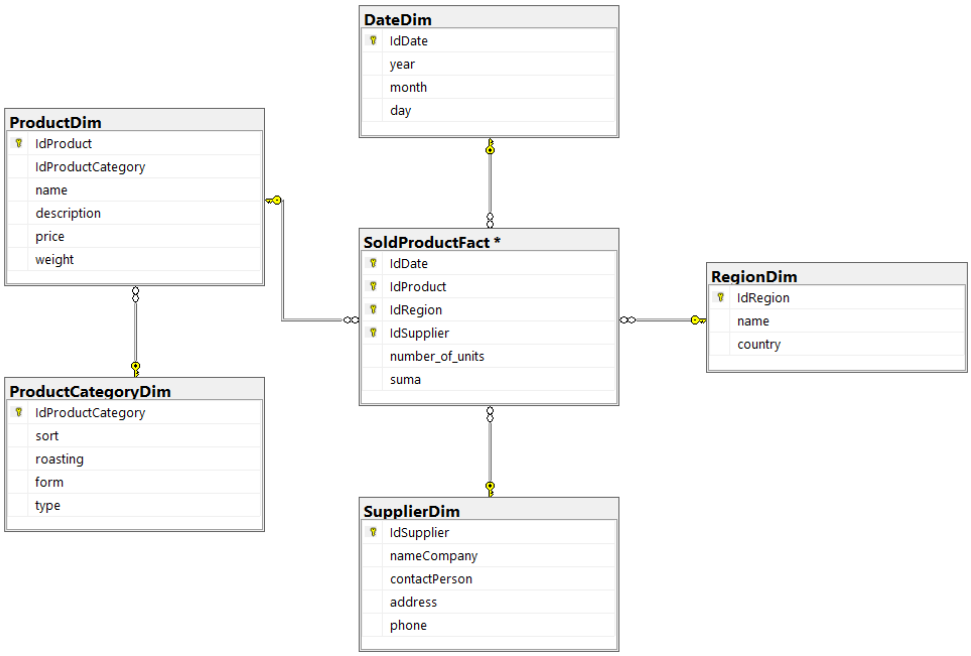


Рис. 5 СД системи підтримки прийняття рішень керівництва з продажу кави

Так як інформація у таблицях-вимірах є відносно постійною, тому дані для цих таблиць заповнювались за допомогою SQL запитів, один з яких наведений на рис. 6

```
use SellingCoffee
go
--inserting SupplierDim
BEGIN TRANSACTION
INSERT INTO SupplierDim(nameCompany,contactPerson,address, phone) VALUES
('Coffee Roasters','Іван Шевченко','м. Харків, вул. Героїв Харкова 273к','+380503665100'),
('Leader Coffee','Марія Адамчук','м. Одеса, вул. Генерала Цветаєва 3/5','+380677350555'),
('OCoffee','Тарас Балюх','м. Київ, вул. Антоновича, 17','+380931709217'),
('Art Coffee Shop', 'Денис Малюк','м.Запоріжжя, вул.Фортечна, 83а','+380501743330')
GO
COMMIT
```

Рис.6 Внесення умовно-постійної інформації про постачальників

Результат виконання внесення інформації у таблицю «Постачальники» наведено на рис.7

	IdSupplier	nameCom...	contactPers...	address	phone
▶	1	Coffee Roas...	Іван Шевче...	м. Харків, в...	+38050366...
	2	Leader Coff...	Марія Ада...	м. Одеса, в...	+38067735...
	3	OCoffee	Тарас Балюх	м. Київ, вул....	+38093170...
	4	Art Coffee S...	Денис Мал...	м.Запоріж...	+38050174...
*	NULL	NULL	NULL	NULL	NULL

Рис. 7 Результат виконання внесення інформації у таблицю

3.1.2 Загальні поняття з напрямку OLAP-технології.

Основне призначення OLAP-систем - підтримка аналітичної діяльності, довільних запитів користувачів-аналітиків[13]. OLAP є аналітичним інструментом і спочатку ґрунтувався на багатовимірних базах даних (ББД). Вони сконструйовані спеціально для підтримки аналізу кількісних даних з численною кількістю вимірювань, містять дані у багатовимірному вигляді.

On-Line Analytical Processing (OLAP) - технологія оперативної аналітичної обробки даних, що використовує методи і засоби для збору, зберігання та аналізу багатовимірних даних з метою підтримки процесів прийняття рішень. OLAP дає змогу організувати вимірювання у вигляді ієрархії. Дані представлені у вигляді гіперкубів (кубів) - логічних і фізичних моделей показників, що спільно використовують вимірювання, а також

ієрархії у цих вимірюваннях. Деякі дані заздалегідь агреговані в БД, інші розраховуються відразу.

OLAP-куб містить базові дані та інформацію про вимірювання (агрегати). Куб потенційно містить всю інформацію, потрібну для відповідей на будь-які запити.

Засоби OLAP дають змогу досліджувати дані за різними вимірюваннями. Користувачі можуть вибирати, які показники аналізувати, які вимірювання і як відображати в крос-таб-лиці, поміняти рядки і стовпці pivoting, потім робити зрізи, щоб концентруватися на певній комбінації розмірностей. Можна змінювати деталізацію даних, рухаючись рівнями за допомогою деталізації та збільшення, а також крос-деталізацію через інші вимірювання.

Для підтримки ББД використовуються OLAP-сервери, оптимізовані для багатовимірного аналізу і які поставляються з аналітичними можливостями.

Виокремлюють і використовують три основні способи реалізації OLAP-сервера для реалізації багатовимірної моделі:

1) MOLAP - багатовимірні БД;

2) ROLAP - реляційні БД;

3) HOLAP - багатовимірні і реляційні БД.

Для розробки курсового проєкту було використано архітектуру MOLAP.

MOLAP-сервери використовують для зберігання та управління даними багатовимірними БД. MOLAP використовує БД, що показує результуючі дані, спеціальний варіант процесора просторових БД. Дані зберігаються у вигляді впорядкованих багатовимірних масивів. Такі масиви поділяються на гіперкуби і полікуби.

У гіперкубі всі комірки, що зберігаються в БД, мають однакову розмірність, тобто знаходяться у якнайбільшому базисі вимірювань.

У полікубі кожна комірка зберігається із власним набором вимірювань, і пов'язані з цим труднощі обробки перекладаються на внутрішні механізми системи.

Фізичні дані, подані у багатовимірному вигляді, зберігаються у двовимірних файлах. Куб представляється у вигляді однієї пласкої таблиці, в яку за рядками вписуються усі комбінації членів усіх вимірювань з відповідними їх значенням мір

MOLAP найкраще підходить для невеликих наборів даних, він швидко розраховує агрегати і повертає відповіді, але при цьому генеруються величезні обсяги даних[14].

3.1.3 Механізм вилучення, обробки і передачі даних

Для розробки кубу було використано середовище Microsoft Visual Studio [15] з розширенням SSAS[16]. На першому етапі необхідно визначити джерело даних – база даних OLAP або сховище даних. На основі визначеного джерела даних будуть імпортуватись необхідні дані. На рис. 8 зображено формування підключення джерела даних за допомогою модуля Data Source Wizard, де ми обираємо попередньо створене сховище даних.

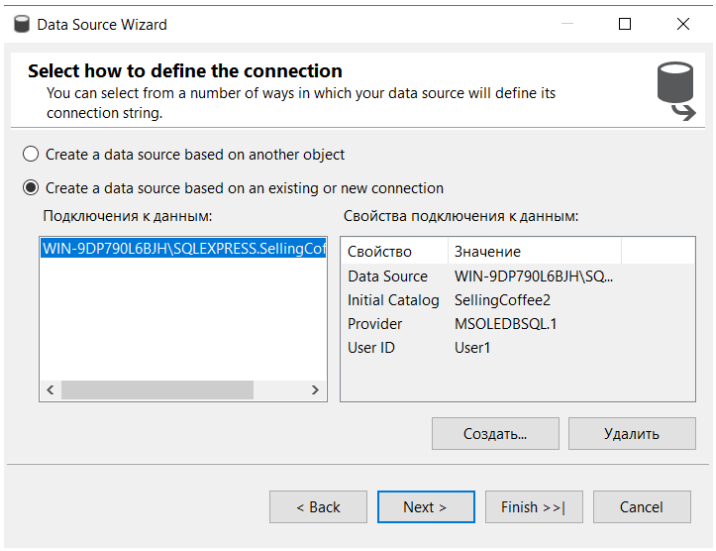


Рис. 8 Додавання джерела даних

Наступним кроком у розробці куба є створення уявлення джерела даних (DSV), яке є абстракцією реляційного джерела даних. DSV стає основою для кубів і вимірів, які будуть створені в багатовимірному проекті. Головною метою DSV є надання контролю над структурами даних у проекті та можливість працювати незалежно від базових джерел даних. Наприклад, ви

можете перейменувати або об'єднати стовпці без прямого впливу на вихідне джерело даних.

У процесі розробки може бути створено кілька представлень джерел даних у проекті або базі даних Analysis Services. Кожне з цих представлень може бути налаштоване так, щоб відповідати конкретним вимогам рішення, що розглядається. Такий підхід дозволяє розробникам створювати різні відображення даних, щоб задовольнити вимоги різних рішень чи забезпечити різноманітні погляди на дані залежно від потреб проекту.

На рис. 9 зображено додавання уявлення за допомогою Data Source View Wizard. У розроблюваному проєкті уявленням виступає створене СД.

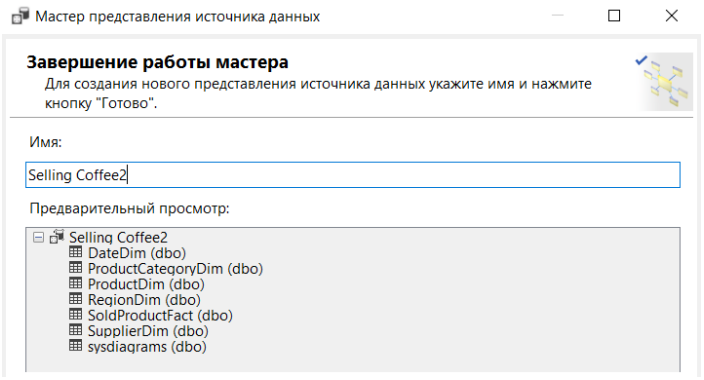


Рис. 9 Створення уявлення на основі СД SellingCoffe2

На основі створеного джерела та уявлення джерелах даних необхідно створити виміри для OLAP куба. Дивлячись на структуру створеного сховища у пункті 2.3 куб матиме 5 виміри.

На рис. 10 зображено процес створення часового виміру, яке буде дозволяти проводити аналіз у розрізі часу: року, місяця, дня.

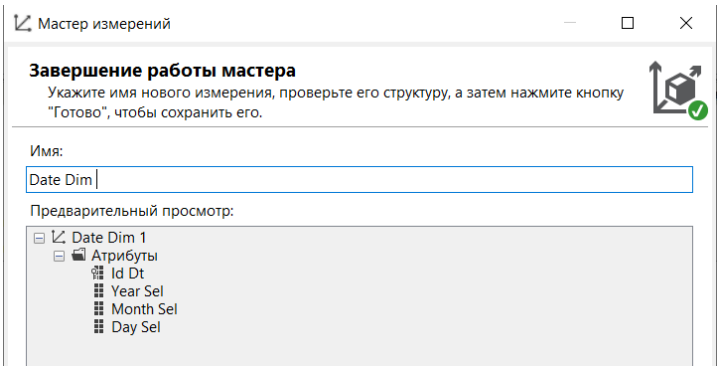


Рис. 10 Створення часового виміру (DateDim)

Наступним виміром є вимір Продукції, який є одним з ключових вимірів. У подальшому можна проводити аналіз у розрізі продукції та категорії продукції, до якої відносяться певні товари.

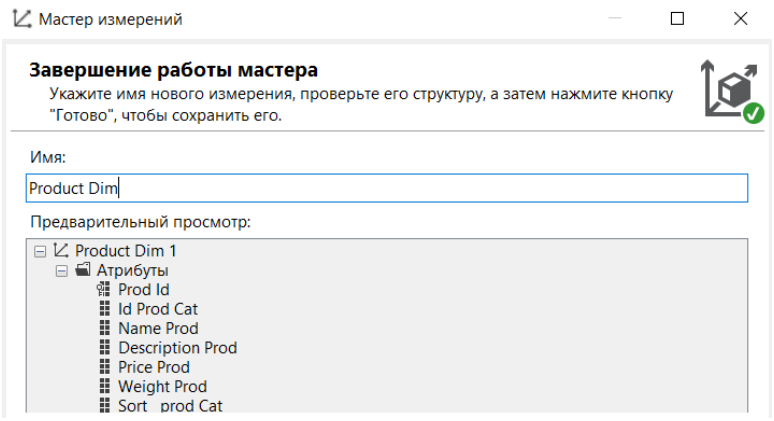


Рис. 11 Створення виміру товару (ProductDim)

Наступним виміром є Категорія продукції. У подальшому можна проводити аналіз у розрізі категорії продукції, до якої відносяться певні товари.

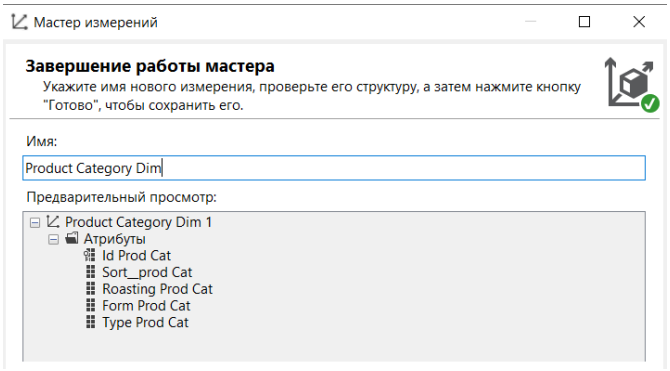


Рис. 12 Створення виміру категорія товару (ProductCategoryDim)

Наступним виміром є Регіон. У подальшому можна проводити аналіз у розрізі регіону продукції, до якого відносяться товари.

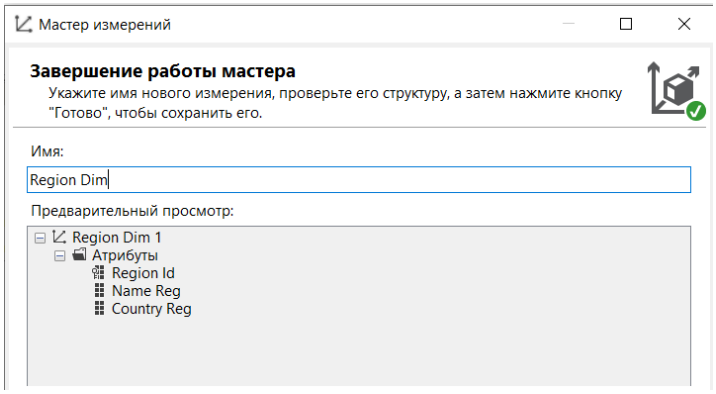


Рис. 13 Створення виміру Регіон (RegionDim)

Наступним виміром є Постачальники. У подальшому можна проводити аналіз у розрізі постачальників продукції, до якого відносяться товари.

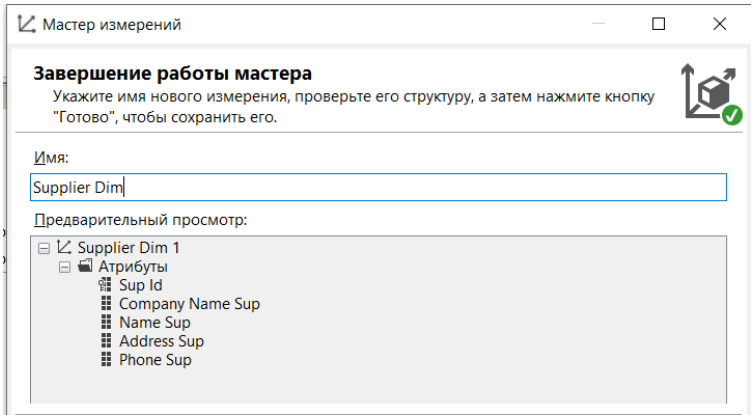


Рис. 14 Створення виміру Постачальник(SupplierDim)

Після створення всіх необхідних вимірів наступним кроком є створення кубу OLAP за допомогою майстра кубів (Cube Wizard). На рис. 15 зображено створення кубу. На першому етапі необхідно обрати таблицю фактів, яка виступає основою для створення кубу.

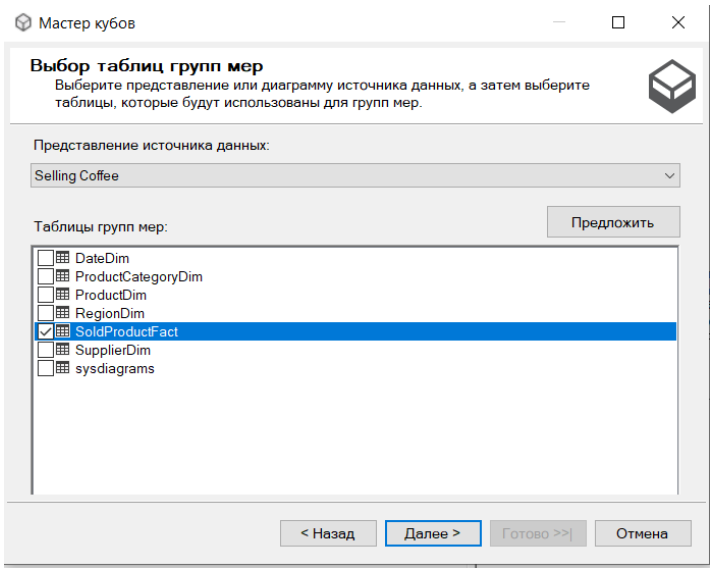


Рис. 15 Вибір таблиці для створення кубу

Наступним етапом іде вибір полів та вимірів, які формуватимуть куб. Вибір полів та вимірів представлено на рис. 16.

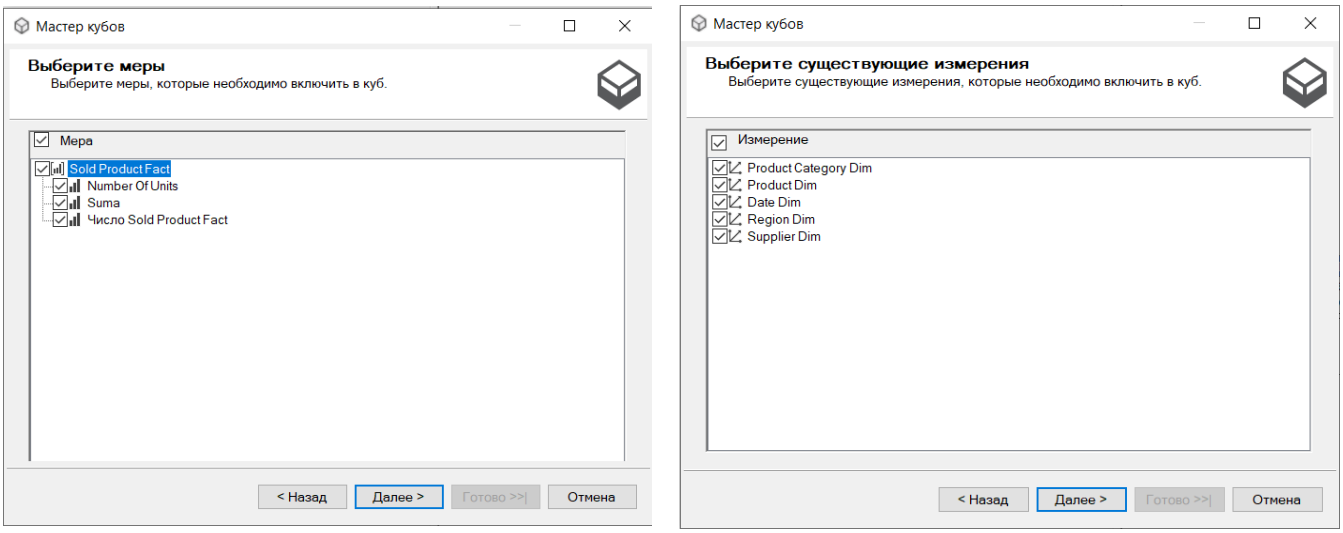


Рис. 16 Вибір полів та вимірів

Після виконання всіх попередніх етапів результатом є створений SellingCoffee2, що зображений на рис. 17.

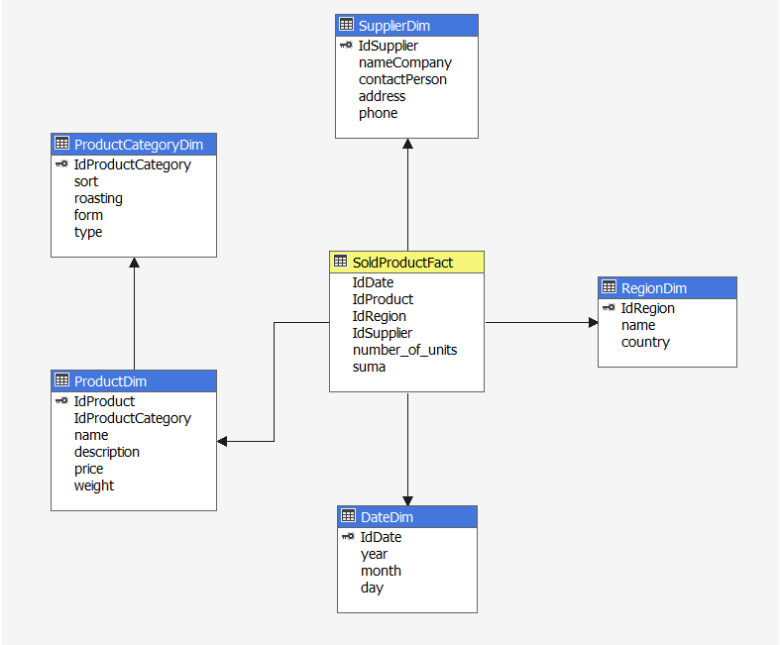


Рис. 17 Створений куб SellingCoffee2

Процес передачі даних було реалізовано за допомогою служби SQL Server Integration Services. В інструменті SSIS є служба Data Flow, за допомогою якої було проведено заповнення таблиць вимірів та фактів.

Наповнення пустого СД відбувається на основі заповненого сховища даних і поділено на 3 етапи (рис. 18).

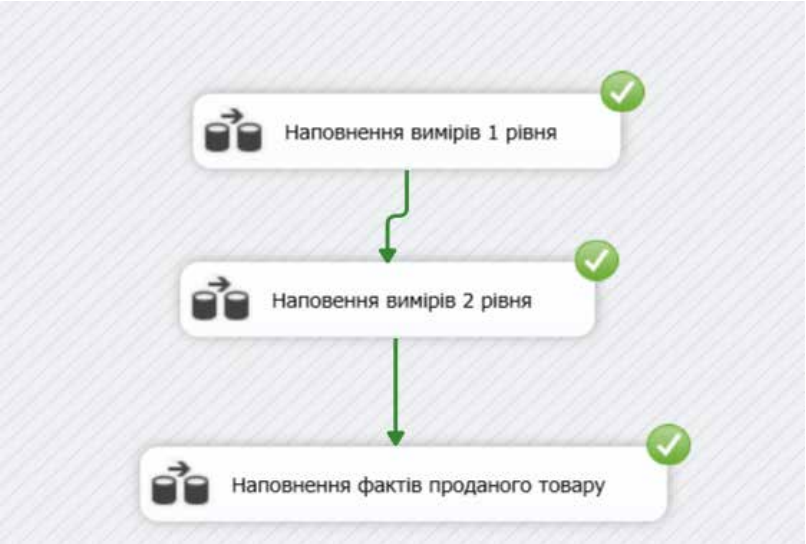


Рис. 18 Потоки даних для наповнення СД

На першому етапі ми заповнюємо таблиці вимірів 1 рівня(батьківскього), дані для яких беруться, як було зазначено вище, з наповненого СД. На рис. 19 зображено потоки, які реалізують передачу даних з БД у СД.

Рис. 19 Перший крок наповнення (таблиць-вимірів 1 рівня)

На прикладі буде продемонстровано наповнення виміру ProductCategoryDim.

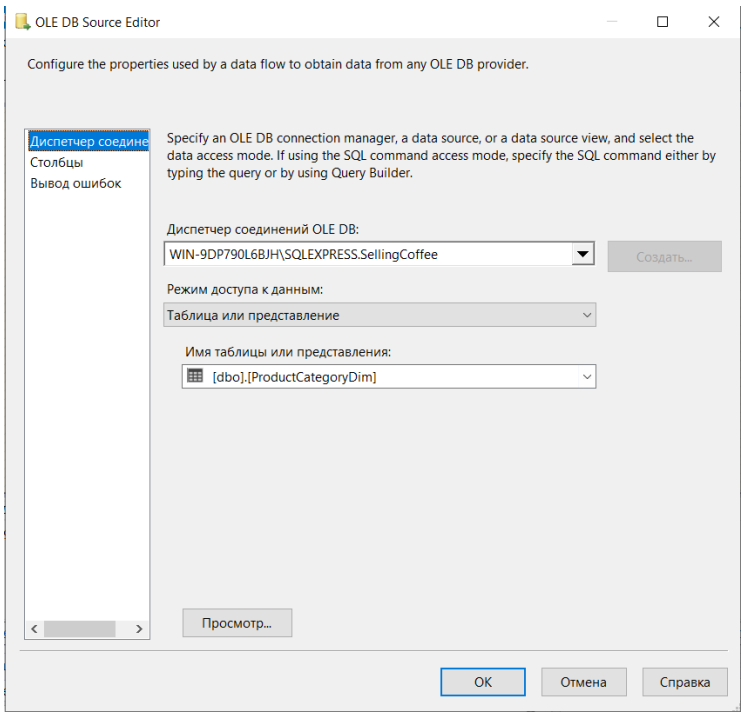


Рис. 20 Вибір потрібного виміру з списку

Процес передачі даних на основі сформованої вибірки представлено на рис. 21

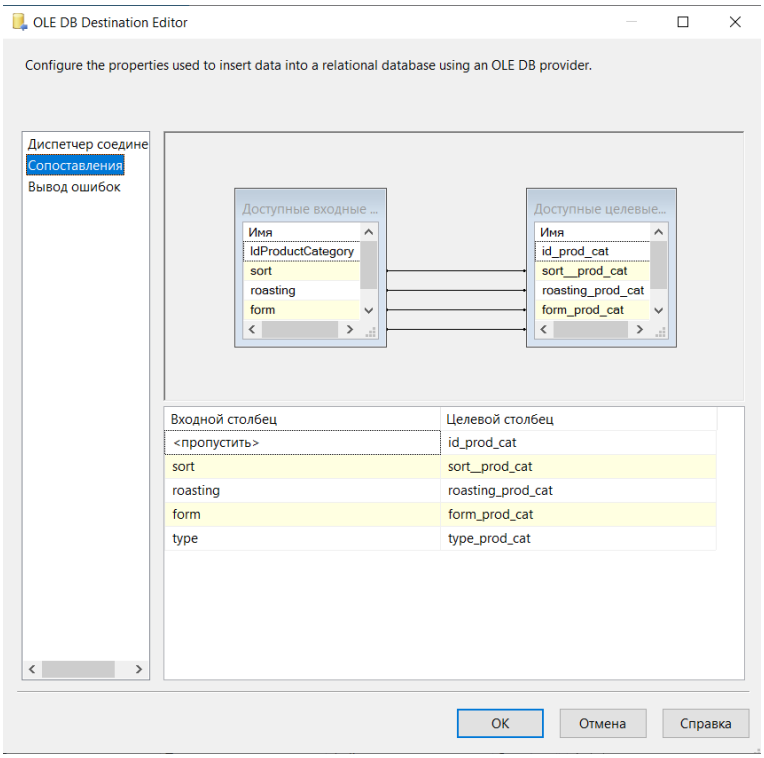


Рис. 21 Процес передачі даних на основі сформованої вибірки

На другому етапі ми заповнюємо таблиці вимірів 2 рівня(дочірнього), дані для яких беруться, як було зазначено вище, з наповненого СД. На рис. 22 зображено потоки, які реалізують передачу даних з БД у СД.

Рис. 22 Другий крок наповнення (таблиць-вимірів 2 рівня)

На прикладі буде продемонстровано наповнення виміру ProductDim.

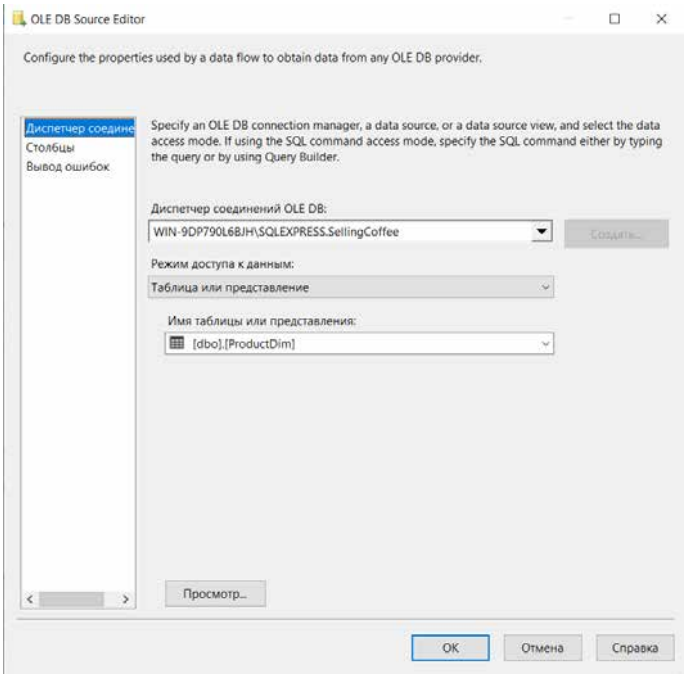


Рис. 23 Вибір потрібного виміру з списку(Джерело OLE DB)

Процес передачі даних на основі сформованої вибірки представлено на рис. 24.

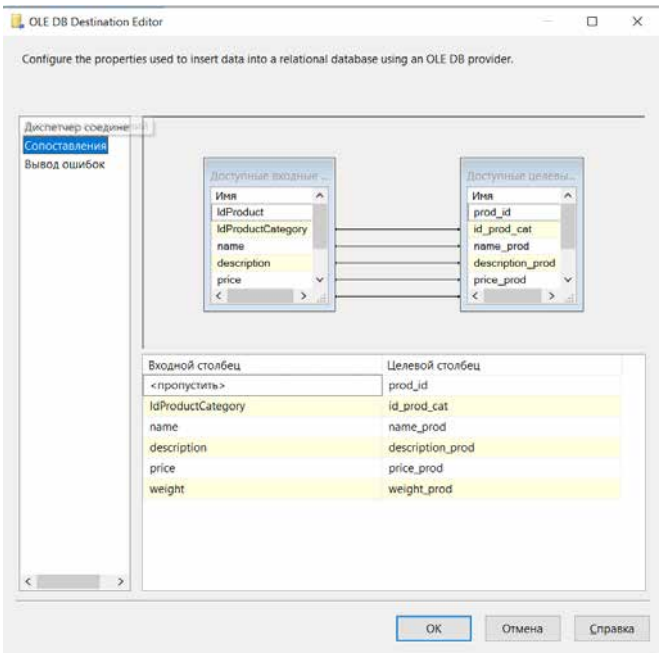


Рис. 24 Процес передачі даних на основі сформованої вибірки

На третьому етапі проводиться заповнення таблиці фактів проданого товару. Запит, наведений нижче проводить проводить вибірку з вимірів і рахує суму проданого товару та суми цього проданого товару із наповненого СД.

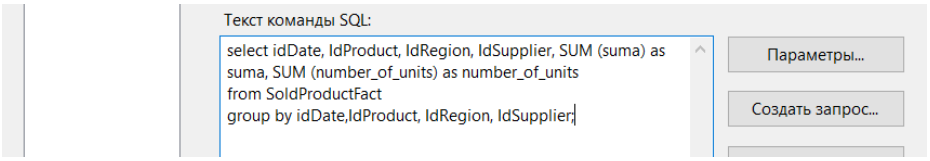


Рис.25 SQL запит для вибірки

Успішне завершення потокі даних свідчить про те, що інтеграція даних у проекті пройшла без помилок і дані були успішно завантажені в відповідні таблиці.

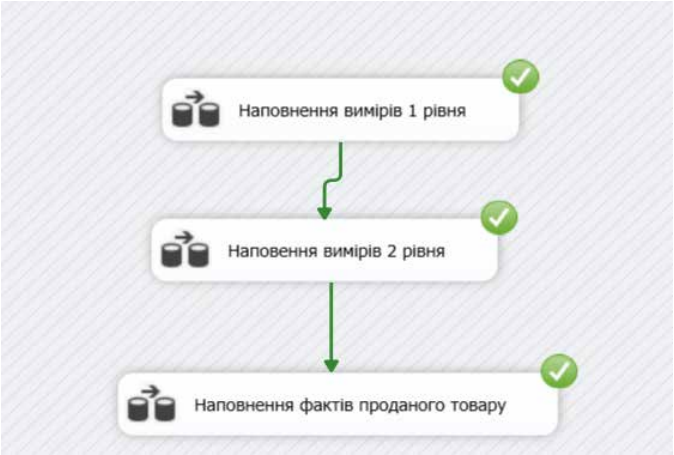


Рис. 26 Успішна інтеграція даних

3.2 Загальні поняття технології Data Mining

Технологію Data Mining[17] можна розуміти як процес дослідження даних за допомогою очищення, пошуку закономірностей, розробки моделей і створення тестів. Data Mining включає в себе поняття машинного навчання, статистики та управління базами даних. Як наслідок, часто легко сплутати інтелектуальний аналіз даних з аналітикою даних, наукою про дані або іншими процесами роботи з даними.

Data Mining (видобування даних) є важливою технологією в системах підтримки прийняття рішень, особливо в контексті керування платформою з продажу кави. Ця технологія дозволяє витягувати корисну інформацію з

великих масивів даних, допомагаючи керівникам приймати обґрунтовані рішення.

Data Mining дозволяє аналізувати історію покупок клієнтів, щоб визначити їхні переваги та поведінкові патерни. Це може включати аналіз частоти покупок, визначення, як часто клієнти купують каву та які сорти користуються найбільшою популярністю, а також сегментацію клієнтів за схожими характеристиками та поведінкою для більш таргетованого маркетингу. Застосування методів прогнозування, таких як регресійний аналіз, дозволяє визначати майбутні тенденції попиту на різні види кави, що дає можливість ефективніше планувати запаси та знижувати ризики недостатньої або надмірної кількості товару на складі.

Аналіз даних про продажі допомагає визначити, які товари найчастіше купують разом, що дозволяє формувати більш привабливі товарні набори та акційні пропозиції. Крім того, Data Mining дозволяє аналізувати дані про програми лояльності, щоб визначити найефективніші стратегії утримання клієнтів. Це може включати аналіз ефективності бонусних програм та надання індивідуальних пропозицій на основі попередніх покупок клієнта.

Використання технологій Data Mining у системі підтримки прийняття рішень для платформи з продажу кави дозволяє значно підвищити ефективність управління, оптимізувати маркетингові стратегії та покращити обслуговування клієнтів. Це забезпечує конкурентну перевагу на ринку та сприяє сталому розвитку бізнесу.

3.3 Огляд інструментів для реалізації завдань Data Mining

В даній роботі для реалізації задач Data Mining було використано Mining Structure[18]. Mining Structure є важливою концепцією в Data Mining і використовується для зберігання даних та визначення взаємозв'язків між цими даними. Вона відіграє ключову роль у побудові моделей Data Mining та аналізі даних.

Mining Structure – це організаційна одиниця, яка визначає формат та джерела даних для аналізу. Вона містить метадані про типи даних, їхні властивості та взаємозв'язки, але не містить самих даних. Mining Structure використовується для побудови однієї або більше моделей Data Mining, що аналізують дані, які вона описує. Основні компоненти Mining Structure включають джерела даних, атрибути, взаємозв'язки та метадані. Джерела даних вказують на таблиці, бази даних або інші джерела, які містять необхідні для аналізу дані. Атрибути – це стовпці даних, які будуть використовуватися у моделях. Вони можуть включати числові дані, текстові дані, дати тощо. Взаємозв'язки визначають, як різні атрибути взаємопов'язані між собою, а метадані описують типи даних та інші властивості атрибутів, такі як максимальне та мінімальне значення, середнє значення, стандартне відхилення тощо.

Для створення Mining Structure використовуються різні інструменти та платформи. Наприклад, Microsoft SQL Server Analysis Services (SSAS) надає інтерфейс для визначення структури даних, включаючи вибір атрибутів, визначення взаємозв'язків та налаштування метаданих. Після створення Mining Structure її можна використовувати для побудови моделей Data Mining. Кожна модель аналізує дані згідно з визначеними у Mining Structure правилами та зв'язками. Наприклад, можна створити модель класифікації для прогнозування поведінки клієнтів або модель кластеризації для сегментації ринку.

Використання Mining Structure має кілька переваг. По-перше, одна Mining Structure може використовуватися для створення декількох моделей Data Mining, що дозволяє ефективніше використовувати ресурси та дані. По-друге, визначені одного разу структури можуть бути повторно використані в різних проектах, що знижує час і витрати на підготовку даних. По-третє, використання визначених взаємозв'язків та метаданих забезпечує цілісність та узгодженість даних у всіх моделях. Нарешті, Mining Structure дозволяє

визначати складні взаємозв'язки між атрибутами, що робить аналіз більш точним та детальним.

Існує кілька інструментів та платформ, які підтримують створення та управління Mining Structure. Microsoft SQL Server Analysis Services (SSAS) є одним з найпопулярніших інструментів для створення та використання Mining Structure. SSAS надає зручний інтерфейс для визначення структур даних та побудови моделей. IBM SPSS Modeler пропонує широкий спектр можливостей для Data Mining, включаючи створення Mining Structure та моделей аналізу. SAS Enterprise Miner є потужним інструментом для аналізу даних, який підтримує визначення Mining Structure та побудову різноманітних моделей Data Mining. RapidMiner – відкрите програмне забезпечення для аналізу даних, що підтримує створення Mining Structure та реалізацію моделей Data Mining. KNIME – відкрите аналітичне програмне забезпечення, яке дозволяє створювати та використовувати Mining Structure для різних завдань Data Mining.

Mining Structure є фундаментальною складовою Data Mining, що дозволяє організовувати та аналізувати великі обсяги даних ефективно та зручно. Використання Mining Structure забезпечує гнучкість, повторне використання та цілісність даних, що значно підвищує ефективність процесів Data Mining та прийняття рішень.

3.4 Аналіз даних

Для дослідження були використані дані, які були взяті з накладних, які надалися з місця роботи. Ці дані були відфільтровані та взяті за 2023-2024 роки. Вони надають можливість для детального аналізу та виявлення корисних інсайтів, що сприятимуть поліпшенню бізнес-процесів та прийняттю обґрунтованих рішень.

Однією з найважливіших складових дослідження стали дані про товари, постачальників та клієнтів. Загалом опрацьовано 914 запитів, що представляє

собою значну кількість даних для подальшого аналізу та використання методів Data Mining.

Аналіз даних з використанням Power BI [19] дозволяє отримати глибоке розуміння змін у продажах товарів за різними регіонами. Завдяки зручному інтерфейсу Power BI, можна створювати діаграми, що наочно відображають кількість проданих одиниць за різні періоди, наприклад, 2023-2024 роки. Вибір саме стовпчастої діаграми для візуалізації допомагає наочно порівняти продажі у різних регіонах, таких як Бразилія, Гватемала та Сальвадор, і виявити значущі тренди, як зростання продажів KIMBO AMICA TRADIZIONE або високі результати Lavazza Super Crema у Сальвадорі в 2024 році. Це відображено на рис. 27.

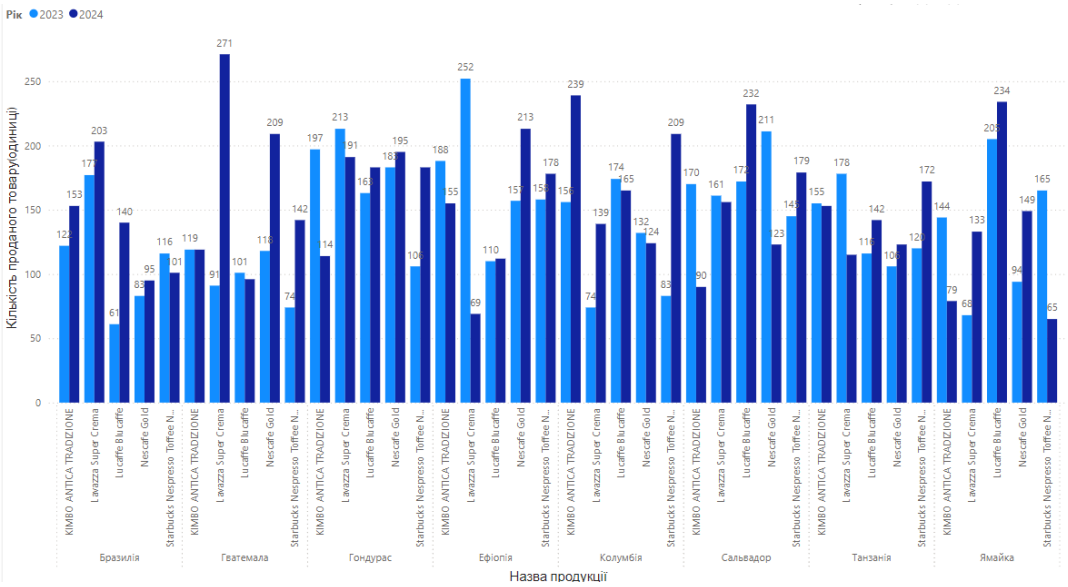


Рис. 27 Стовпчаста діаграма кількості проданого товару за 2023-2024 рік за категорією Регіон

Power BI є чудовим інструментом для такого аналізу, оскільки він дозволяє інтерактивно досліджувати дані, легко налаштовувати різні візуалізації та надавати доступ до звітів у реальному часі для всіх зацікавлених осіб. Зручність у використанні, інтеграція з іншими джерелами даних, а також здатність обробляти великі об'єми інформації роблять Power BI ідеальним вибором для таких завдань, адже він дозволяє швидко отримати аналітику, приймати рішення на основі даних і коригувати стратегії для покращення бізнес-результатів.

Також було проаналізовано дані за допомогою KPI. Ключові показники ефективності (KPI, Key Performance Indicators) — це метрики, які використовуються для вимірювання продуктивності та результативності діяльності компанії[21]. Вони допомагають оцінити, наскільки успішно виконуються стратегічні цілі і завдання. Основне призначення KPI — вимірювати прогрес у досягненні цілей, вчасно виявляти проблеми, аналізувати та усувати їх, а також приймати управлінські рішення. Існують різні KPI, залежно від конкретної сфери діяльності, але загальні категорії включають фінансові показники (наприклад, прибуток, оборот, маржа), показники якості (наприклад, рівень задоволеності клієнтів, якість продукту чи послуги), маркетингові (коефіцієнт клікабельності, вартість ліда тощо), показники продуктивності (наприклад, час, необхідний для виконання конкретних завдань) тощо.

Класифікація KPI:

1. Фінансові KPI: відображають фінансові результати, такі як прибуток, рентабельність, обсяг продажів, ROI (рентабельність інвестицій).
2. Нефінансові KPI: оцінюють аспекти, що не пов'язані безпосередньо з фінансами, наприклад, рівень задоволеності клієнтів, ефективність операційних процесів, утримання персоналу.
3. Внутрішні та зовнішні KPI: внутрішні відображають внутрішні процеси компанії, тоді як зовнішні орієнтовані на взаємодію з клієнтами та ринком.

Характеристики ефективного KPI:

- Специфічність: пов'язаність із конкретною ціллю.
- Вимірюваність: можливість кількісного відображення.
- Досяжність: має бути реалістичним і піддаватися впливу.
- Актуальність: відповідність стратегічним цілям.
- Обмеженість у часі: повинен бути орієнтований на певний часовий період.

Застосування KPI допомагає організаціям моніторити прогрес, вчасно виявляти проблеми й коригувати стратегії. Наприклад, аналіз KPI у сфері продажів дає можливість відстежувати зростання або спад обсягів продажу, визначати ефективність маркетингових кампаній та оптимізувати процеси для досягнення вищих результатів.

У ході аналізу даних було визначено KPI, розміщені на рис. 28. А саме:

1. KPI_SoldProduct1_2023_LavazzaSuperCrema_Brazil - визначення кількості проданої продукції LavazzaSuperCrema за 2023 рік, регіон – Бразилія.
2. KPI_SoldProduct1_2024_LavazzaSuperCrema_Brazil - визначення кількості проданої продукції LavazzaSuperCrema за 2024 рік, регіон – Бразилія.
3. KPI_SoldProduct2_2023_LavazzaSuperCrema_Ethiopia - визначення кількості проданої продукції LavazzaSuperCrema за 2023 рік, регіон – Ефіопія.
4. KPI_SoldProduct2_2024_LavazzaSuperCrema_Ethiopia - визначення кількості проданої продукції LavazzaSuperCrema за 2024 рік, регіон – Ефіопія.
5. KPI_SoldProduct1_Suma_2023_LavazzaSuperCrema_ Brazil – сума проданої продукції LavazzaSuperCrema за 2023 рік, регіон – Бразилія.
6. KPI_SoldProduct1_Suma_2024_LavazzaSuperCrema_ Brazil – сума проданої продукції LavazzaSuperCrema за 2024 рік, регіон – Бразилія.

Отобразить структуру	Значение	Цель	Состояние	Тренд	Вес
KPI_SoldProduct1_2023_LavazzaSuperCrema_Brazil	177	100		↑	
KPI_SoldProduct1_2024_LavazzaSuperCrema_Brazil	203	100		↑	
KPI_SoldProduct1_2023_LavazzaSuperCrema_Ethiopia	252	100		↑	
KPI_SoldProduct1_2024_LavazzaSuperCrema_Ethiopia	69	100		↓	
KPI_SoldProduct1_Suma_2023_LavazzaSuperCrema_Brazil	111864	100000		↑	
KPI_SoldProduct1_Suma_2024_LavazzaSuperCrema_Brazil	213616	100000		↑	

Рис. 28 KPI

У висновку можна зазначити, що аналіз KPI за 2023 і 2024 роки виявив різноспрямовані тенденції у продажах продукту Lavazza Super Crema в різних

регіонах. У Бразилії спостерігається стійкий позитивний тренд: кількість проданих одиниць зросла з 177 у 2023 році до 203 у 2024 році, що спричинило збільшення суми продажів з 111,684 грн до 213,616 грн. Це свідчить про підвищення попиту та ефективність стратегії просування продукту на цьому ринку.

Водночас в Ефіопії зафіксовано різке зниження продажів — з 252 одиниць у 2023 році до 60 одиниць у 2024 році, що вказує на можливі проблеми, пов'язані зі зміною споживчих вподобань чи посиленням конкуренції. Також це різке зниження могло відбутися через неякісну обсмажку або використання неякісних зерен, що негативно вплинуло на сприйняття продукту споживачами. Ці дані є важливими для подальшої оцінки ринкових стратегій , що сприятиме підвищенню продажів та задоволеності споживачів.

4 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

4.1 Дослідження використання задач класифікації

4.1.1 Використання 1-Rule для класифікації.

Алгоритм класифікації «1-Rule»[22] - це простий, але ефективний підхід, який використовується в інтелектуальному аналізі даних для задач класифікації. Він працює шляхом вибору одного правила на основі найчастішого класу (або результату) в наборі даних. Ось як це працює:

Потрібно вибрати найпоширеніший клас: Алгоритм аналізує цільову змінну (змінну для прогнозування) в наборі даних і визначає, який клас зустрічається найчастіше.

Другим кроком треба створіти правило: Після того, як визначено найчастіший клас, алгоритм створює правило, яке передбачає цей клас для всіх екземплярів у наборі даних.

Третім кроком є застосування правила: Створене правило застосовується до нових даних або спостережень, щоб віднести їх до передбачуваного класу.

Наприклад, розглянемо набір даних із цільовою змінною, яка вказує, чи є імейл спамом, чи ні. Якщо більшість імейлів у наборі даних класифікуються як неспам, алгоритм 1-правило згенерує правило, яке визначить, що всі імейли не є спамом. Це правило буде застосовуватися для класифікації нових імейлів як неспам.

Хоча алгоритм 1-го правила простий і легкий у застосуванні, він не завжди дає найточніші результати, особливо в наборах даних з незбалансованим розподілом класів або складними взаємозв'язками між ознаками та цільовою змінною. Однак він може слугувати базовою або відправною точкою для більш досконалих алгоритмів класифікації і може бути корисним для швидкого розуміння найпоширенішого класу в наборі даних.

Для розв'язання задачі класифікації було використано Microsoft Visual Studio та дані з СД, що містяться в Microsoft SQL Management Studio[23]. Було написано програму та використано алгоритм 1R.

На рис. 29 показано запит для отримання перших 160 записів та їх назв з таблиці SoldProductFact для продуктів, регіонів та постачальників:

```
private void LoadData()
{
    try
    {
        using (SqlConnection connection = new SqlConnection(connectionString))
        {
            connection.Open();

            // Запит для отримання перших 160 записів та їх назв для продукції
            string productQuery = @"SELECT TOP 160
                                    pd.name AS Name,
                                    spf.IdProduct,
                                    spf.number_of_units
                                FROM SoldProductFact spf
                                INNER JOIN ProductDim pd ON spf.IdProduct = pd.IdProduct";

            SqlDataAdapter productAdapter = new SqlDataAdapter(productQuery, connection);
            DataTable productTable = new DataTable();
            productAdapter.Fill(productTable);

            // Запит для отримання перших 160 записів та їх назв для регіонів
            string regionQuery = @"SELECT TOP 160
                                    rd.name AS Name,
                                    spf.IdRegion,
                                    spf.number_of_units
                                FROM SoldProductFact spf
                                INNER JOIN RegionDim rd ON spf.IdRegion = rd.IdRegion";

            SqlDataAdapter regionAdapter = new SqlDataAdapter(regionQuery, connection);
            DataTable regionTable = new DataTable();
            regionAdapter.Fill(regionTable);

            // Запит для отримання перших 160 записів та їх назв для постачальників
            string supplierQuery = @"SELECT TOP 160
                                    sd.nameCompany AS Name,
                                    spf.IdSupplier,
                                    spf.number_of_units
                                FROM SoldProductFact spf
                                INNER JOIN SupplierDim sd ON spf.IdSupplier = sd.IdSupplier";
```

Рис. 29 Запит для отримання перших 160 записів та їхніх назв з таблиці SoldProductFact

Середнє значення кількості проданих товарів - 10. Далі ділимо дані на два класи - з високою та низькою ефективністю. На рис. 30 підраховано кількість зустрічей для кожного товару, регіону та постачальника. Якщо кількість_одиниць_продажу >= 10, це низька ефективність, якщо < 10 - висока.

```
            // Підрахунок кількості зустрічей для кожного товару (продукції)
            var productCounts = productTable.AsEnumerable()
                                .GroupBy(row => row.Field<int>("IdProduct"))
                                .Select(group => new
                                {
                                    Name = group.First()["Name"].ToString(),
                                    LowSales = $"{group.Count(item => item.Field<int>("number_of_units") >= 10)}/{group.Count()}",
                                    HighSales = $"{group.Count(item => item.Field<int>("number_of_units") < 10)}/{group.Count()}",


                                });
        //Підрахунок кількості зустрічей для кожного регіону
        var regionCounts = regionTable.AsEnumerable()
                           .GroupBy(row => row.Field<int>("IdRegion"))
                           .Select(group => new
                           {
                               Name = group.First()["Name"].ToString(),
                               LowSales = $"{group.Count(item => item.Field<int>("number_of_units") >= 10)}/{group.Count()}",
                               HighSales = $"{group.Count(item => item.Field<int>("number_of_units") < 10)}/{group.Count()}"
                           });
```

```
//Підрахунок кількості зустрічей для кожного постачальника
var supplierCounts = supplierTable.AsEnumerable()
                        .GroupBy(row => row.Field<int>("IdSupplier"))
                        .Select(group => new
                        {
                            Name = group.First()["Name"].ToString(),
                            LowSales = $"{group.Count(item => item.Field<int>("number_of_units") >= 10)}/{group.Count()}",
                            HighSales = $"{group.Count(item => item.Field<int>("number_of_units") < 10)}/{group.Count()}"
                        });
```

Рис. 30 Підрахунок кількості зустрічей для кожного продукту, регіону та постачальника

На рис. 31 відображено результат класифікації за продуктами, регіонами та постачальниками.

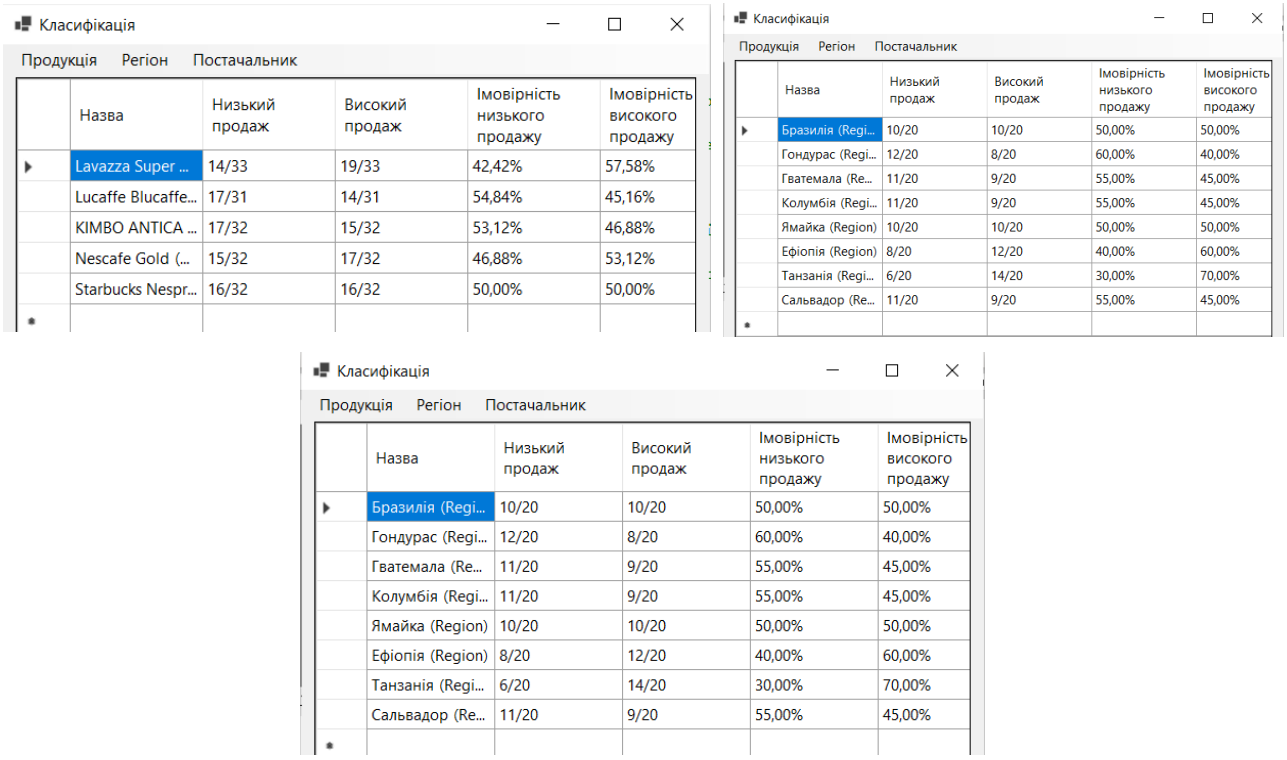


Рис. 31 Результат класифікації за продуктом, регіоном та постачальником

Дані розраховані за період 01.10.2023-01.11.2023

Середнє значення продажу для всіх проданих товарів - 10.

Дані розділені на два класи:

- Низькі продажі

- Високі продажі

Згідно з результатами за назвами продуктів, можна зробити наступний висновок:

Lavazza Super Crema має високу ймовірність продажу, яка становить 57,58%, що може свідчити про популярність цього продукту серед покупців. Однак інші продукти, такі як Lucaffe Blucaffe та KIMBO ANTICA

TRADIZIONE, мають вищу ймовірність низького рівня продажів, що варто враховувати при плануванні товарних запасів та маркетингових стратегій.

Nescafe Gold та Starbucks Nespresso Toffee Nut Latte мають приблизно однакову ймовірність продажу, що може свідчити про схожу популярність цих продуктів серед споживачів.

Згідно з результатами за регіонами, можна зробити наступні висновки:

1. Гондурас та Сальвадор мають вищі ймовірності низьких продажів, ніж високих, що може свідчити про потенційні труднощі з реалізацією продукції з цих країн.

2. Ефіопія та Танзанія, навпаки, мають вищі ймовірності високих продажів, що свідчить про можливі переваги в експорті їхньої продукції.

3. Бразилія, Ямайка, Гватемала та Колумбія мають приблизно рівні ймовірності низьких та високих продажів, що може свідчити про стабільний або прогнозований рівень попиту на їхню продукцію.

Згідно з результатами за постачальниками, можна зробити наступний висновок:

Art Coffee Shop користується високим попитом, а Coffee Roasters - низьким.

Метод 1-Rule ефективний завдяки своїй простоті та здатності швидко створювати правила для класифікації. Однією з головних переваг є швидкість виконання: оскільки метод використовує лише одну змінну для прийняття рішень, навчання моделі відбувається дуже швидко, що є корисним для великих наборів даних або в умовах обмеженого часу.

Ще однією важливою перевагою є інтерпретованість. Правила, які генерує 1-Rule, легко зрозуміти, що робить їх зручними для пояснення результатів моделі як технічним, так і нетехнічним користувачам.

1-Rule також може бути ефективним на невеликих наборах даних або в задачах, де складні моделі можуть виявитися зайвими. У таких випадках цей метод дозволяє досягти хороших результатів без зайвого ускладнення моделі.

Завдяки тому, що метод використовує лише одну змінну для класифікації, він знижує ризик перенавчання, що часто стає проблемою в більш складних алгоритмах. Таким чином, 1-Rule надає просте та швидке рішення для задач, де немає потреби у високій складності.

4.1.2 Використання методу наївного Баєса

Наївні байєсівські класифікатори- це набір алгоритмів класифікації, заснованих на теоремі Байєса[24]. Це не один алгоритм, а сімейство алгоритмів, де всі вони мають спільний принцип, тобто кожна пара ознак, що класифікуються, є незалежною одна від одної. Для початку розглянемо набір даних.

Один з найпростіших та найефективніших алгоритмів класифікації, класифікатор Naïve Bayes, допомагає у швидкій розробці моделей машинного навчання з можливостями швидкого прогнозування.

Наївний алгоритм Байєса використовується для задач класифікації. Він широко використовується в класифікації текстів. У задачах класифікації тексту дані містять високу розмірність (оскільки кожне слово представляє одну ознаку в даних). Він використовується для фільтрації спаму, виявлення настроїв, класифікації рейтингів тощо. Перевагою використання наївного Байєса є його швидкість. Він швидкий, і робити прогнози легко при великій розмірності даних.

Ця модель прогнозує ймовірність того, що екземпляр належить до класу із заданим набором значень ознак. Це імовірнісний класифікатор. Це тому, що він припускає, що одна ознака в моделі не залежить від існування іншої ознаки. Іншими словами, кожна ознака робить свій внесок у передбачення, не маючи зв'язку між собою. У реальному світі ця умова виконується рідко. В алгоритмі навчання та прогнозування використовується теорема Байєса.

Фундаментальне припущення наївного Байєса полягає в тому, що кожна ознака вносить свій внесок:

- Незалежність ознак: Ознаки даних є умовно незалежними одна від одної, враховуючи мітку класу.

- Неперервні ознаки є нормально розподіленими: Якщо ознака неперервна, то вважається, що вона нормально розподілена в межах кожного класу.

- Дискретні ознаки мають мультиноміальний розподіл: Якщо ознака дискретна, то вважається, що вона має мультиноміальний розподіл в межах кожного класу.

- Ознаки однаково важливі: Вважається, що всі ознаки роблять однаковий внесок у прогнозування мітки класу.

- Відсутність пропущених даних: Дані не повинні містити жодних пропущених значень.

Стосовно нашого набору даних це поняття можна зрозуміти так:

- Ми припускаємо, що жодна пара ознак не є залежною. Наприклад, температура «Спекотно» не має нічого спільного з вологістю, або прогноз «Дощитиме» не впливає на вітер. Таким чином, ознаки вважаються незалежними.

- По-друге, кожній ознаці надається однакова вага (або важливість). Наприклад, знаючи лише температуру та вологість, неможливо точно передбачити результат. Жодна з ознак не є несуттєвою і вважається, що вона в рівній мірі впливає на результат.

На рисунку 32 показано структуру куба (побудованого за допомогою SSAS), на основі якого в подальшому буде виконуватися інтелектуальний аналіз даних. Цей куб побудований для вирішення завдань аналізу проданих товарів.

Програмна реалізація алгоритму: Алгоритм було реалізовано за допомогою засобів SQL та C#.

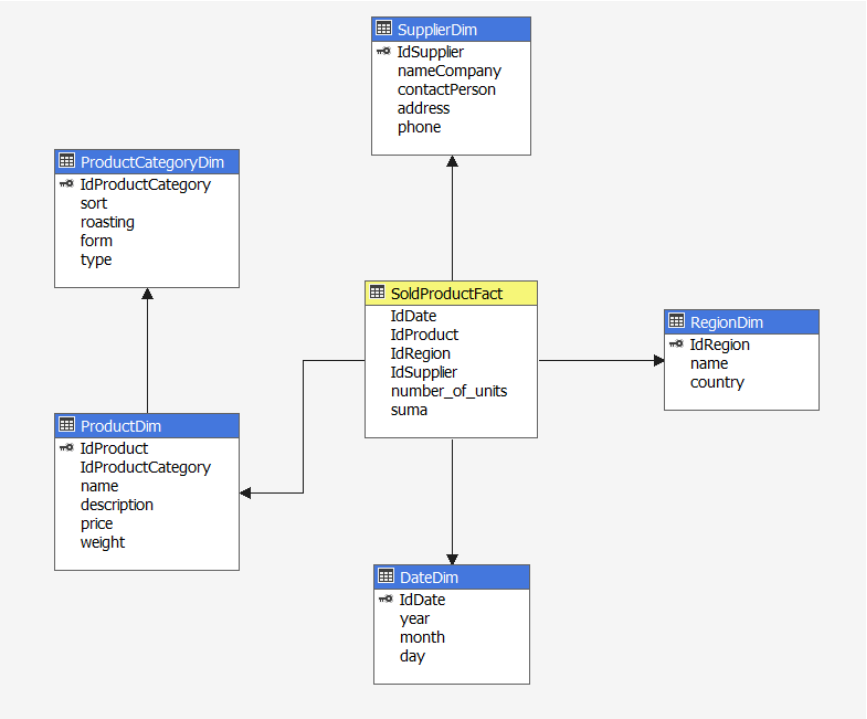
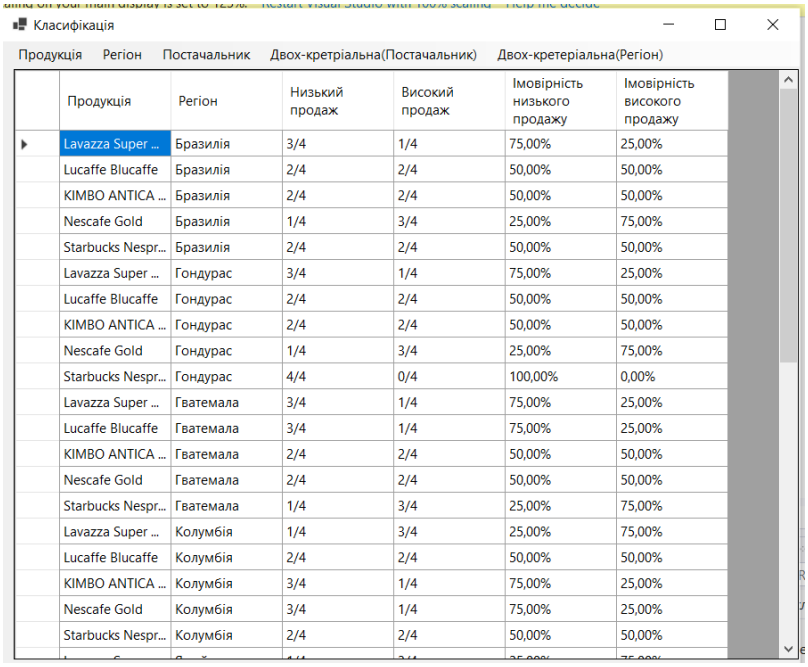


Рис. 32 Структура куба

На рис. 33 представлено результат роботи алгоритму для продуктів та регіону. Згідно з результатом, три продукти мають ймовірність високих продажів, яка досягає 100% - KIMBO ANTICA TRADIZIONE (Танзанія), Nescafe Gold (Танзанія) та Starbucks Nespresso Toffee Nut Latte (Ефіопія).



Класифікація

Продукція | Регіон | Постачальник | Двох-кретріальна(Постачальник) | Двох-кретеріальна(Регіон)

Продукція	Регіон	Низький продаж	Високий продаж	Імовірність низького продажу	Імовірність високого продажу
Lavazza Super ...	Бразилія	3/4	1/4	75,00%	25,00%
Lucaffe Blucaffe	Бразилія	2/4	2/4	50,00%	50,00%
KIMBO ANTICA ...	Бразилія	2/4	2/4	50,00%	50,00%
Nescafe Gold	Бразилія	1/4	3/4	25,00%	75,00%
Starbucks Nespr...	Бразилія	2/4	2/4	50,00%	50,00%
Lavazza Super ...	Гондурас	3/4	1/4	75,00%	25,00%
Lucaffe Blucaffe	Гондурас	2/4	2/4	50,00%	50,00%
KIMBO ANTICA ...	Гондурас	2/4	2/4	50,00%	50,00%
Nescafe Gold	Гондурас	1/4	3/4	25,00%	75,00%
Starbucks Nespr...	Гондурас	4/4	0/4	100,00%	0,00%
Lavazza Super ...	Гватемала	3/4	1/4	75,00%	25,00%
Lucaffe Blucaffe	Гватемала	3/4	1/4	75,00%	25,00%
KIMBO ANTICA ...	Гватемала	2/4	2/4	50,00%	50,00%
Nescafe Gold	Гватемала	2/4	2/4	50,00%	50,00%
Starbucks Nespr...	Гватемала	1/4	3/4	25,00%	75,00%
Lavazza Super ...	Колумбія	1/4	3/4	25,00%	75,00%
Lucaffe Blucaffe	Колумбія	2/4	2/4	50,00%	50,00%
KIMBO ANTICA ...	Колумбія	3/4	1/4	75,00%	25,00%
Nescafe Gold	Колумбія	3/4	1/4	75,00%	25,00%
Starbucks Nespr...	Колумбія	2/4	2/4	50,00%	50,00%

Рис 33 Результат роботи алгоритму для товарів та регіону

На рис. 34 показано, що три продукти є найпродуктивнішими з усіх. Це насамперед «Starbucks Nespresso Toffee Nut Latte», регіон – Ефіопія, «KIMBO

ANTICA TRADIZIONE», регіон – Танзанія, та «Nescafe Gold» регіон – Танзанія.

Starbucks Nespr...	Ефіопія	0/4	4/4	0,00%	100,00%
KIMBO ANTICA ...	Танзанія	0/4	4/4	0,00%	100,00%
Nescafe Gold	Танзанія	0/4	4/4	0,00%	100,00%
Starbucks Nespr...	Танзанія	3/4	1/4	75,00%	25,00%
Lavazza Super ...	Сальвадор	1/4	3/4	25,00%	75,00%
Lucaffe Blucaffe	Сальвадор	2/4	2/4	50,00%	50,00%
KIMBO ANTICA ...	Сальвадор	3/4	1/4	75,00%	25,00%
Nescafe Gold	Сальвадор	3/4	1/4	75,00%	25,00%
Starbucks Nespr...	Сальвадор	2/4	2/4	50,00%	50,00%

Рис. 34 Три продукти є найпродуктивнішими з усіх.

Реалізація розв'язання задачі класифікації методом наївного Байєса з використанням інструментів SSAS.

1. Створення інтелектуальної структури аналізу за допомогою майстра Data Mining Wizard[20].
2. Вибрати таблицю вимірювань, за якою буде показано і спрогнозовано зміну фактичних показників. У задачі, що розв'язується, такою таблицею буде ProductDim, в якій зберігається інформація про існуючі продукти з їх характеристиками:
3. Виберіть поле «IdProduct» в якості ключового поля.

Далі BI пропонує вибрати атрибути, які будуть відображатися в інтелектуальній моделі аналізу на додаток до ключового поля. Оскільки об'єкти потрібно класифікувати за показниками, обираємо обчислювальну міру, яка розраховує середнє значення кількості проданих продуктів.

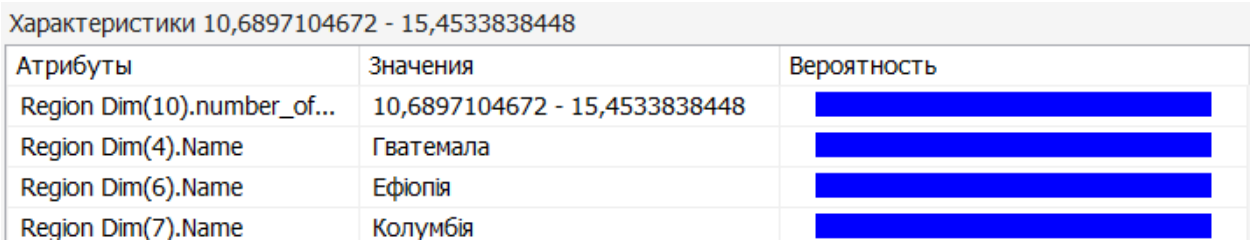
Далі потрібно визначити налаштування для обраних атрибутів

● Вхід - вхідна зміна, яка суттєво впливає на перебіг досліджуваного процесу (Продукт, Назва);

● Прогнозування - зміна, значення якої буде прогнозуватися (Середня кількість проданих товарів).

Оскільки необхідно визначити залежність результативності від двох критеріїв, ми додаємо до моделі аналізу таблицю RegionDim, в якій зберігаються дані про регіон.

Рис. 35 демонструє результат методу Наївного Баєса в Mining Structure:


Характеристики 10,6897104672 - 15,4533838448

Атрибуты	Значения	Вероятность
Region Dim(10).number_of...	10,6897104672 - 15,4533838448	
Region Dim(4).Name	Гватемала	
Region Dim(6).Name	Ефіопія	
Region Dim(7).Name	Колумбія	

Рис. 35 Результат методу Наївного Баєса в Mining Structure

Підсумовуючи, можна сказати наступне:

Згідно з отриманими результатами, можна зробити висновок, що SSAS підтверджує правила, які були отримані за допомогою власного розробленого алгоритму. У вхідних даних було 8 регіонів - Бразилія, Гондурас, Гватемала, Ефіопія, Колумбія, Ямайка, Танзанія та Сальвадор. Дізнавшись це можна сказати, що кава з Гватемали, Ефіопії та Колумбії буде продаватися найбільше з ймовірністю 100%. Продукт буде продаватися по 10-15 одиниць.

За допомогою методу Баєса можна визначити, які товари, якого регіону найкраще продавати конкретним клієнтам для отримання високих обсягів продажів.

4.2 Дослідження використання методу асоціативних правил

Вивчення використання майнінгу на основі асоціативних правил25] передбачає вивчення застосування таких алгоритмів, як Apriori або FP-зростання, для виявлення закономірностей у наборах даних. Ось як ви можете підійти до вивчення цієї теми:

- Розуміння теорії: Почніть з вивчення теоретичних основ асоціативного пошуку правил. Дізнайтеся про такі ключові поняття, як підтримка, впевненість і підйом, які використовуються для оцінки сили і значущості асоціативних правил.

- Вивчення алгоритмів: Зануртеся в алгоритми, які зазвичай використовуються для видобутку асоціативних правил, такі як Apriori та FP-зростання. Зрозумійте, як працюють ці алгоритми, їхні сильні та слабкі сторони, а також обчислювальну складність.

- Попередня обробка даних: Дізнайтеся про важливість попередньої обробки даних в асоціативному видобутку правил. Вивчіть методи очищення, перетворення та дискретизації даних для підготовки набору даних до видобування правил.

- Генерація правил: Вивчіть, як генеруються асоціативні правила на основі транзакційних або реляційних наборів даних. Зрозумійте процес виявлення частих наборів елементів і створення правил асоціації на основі цих наборів елементів.

- Метрики оцінювання: Вивчіть метрики оцінювання, які використовуються для оцінки якості правил асоціацій, включаючи підтримку, впевненість і підйом. Дізнайтеся, як інтерпретувати ці метрики і використовувати їх для фільтрації або ранжування правил асоціацій.

- Застосування: Вивчіть реальні приклади застосування інтелектуального аналізу асоціативних правил у різних галузях, таких як роздрібна торгівля, охорона здоров'я та телекомунікації. Зрозумієте, як асоціативні правила використовуються для вилучення дієвих ідей та керування процесами прийняття рішень.

- Поглиблені теми: Вивчіть поглиблені теми в галузі інтелектуального аналізу асоціативних правил, такі як видобуток послідовних шаблонів, періодичних шаблонів або наборів даних високої розмірності. Розгляньте дослідження останніх розробок і методів у цій галузі.

- Тематичні дослідження: Проаналізуйте тематичні дослідження або наукові роботи, які демонструють практичне використання методу видобування асоціативних правил для вирішення конкретних проблем або вирішення бізнес-завдань. Зверніть увагу на використані методології та отримані знання.

- Інструменти та бібліотеки: Ознайомтеся з популярними інструментами та бібліотеками для асоціативного видобутку правил, такими як Weka, R або бібліотеки Python, такі як Mlxtend. Експериментуйте з цими інструментами, щоб отримати практичний досвід у видобутку правил.

- Експериментальний дизайн: Якщо ви проводите дослідження в цій галузі, розробіть експерименти, щоб оцінити продуктивність різних алгоритмів інтелектуального аналізу або варіацій існуючих алгоритмів. Враховуйте такі фактори, як характеристики набору даних, налаштування параметрів і метрики оцінювання.

Ретельно вивчивши ці аспекти асоціативного видобутку правил, можна розвинету всебічне розуміння методів, застосувань і наслідків цього важливого методу інтелектуального аналізу даних.

Щоб реалізувати пошук асоціативних правил засобами SSAS перш за все треба вибрати метод пошуку асоціативних правил зі списку доступних технологій інтелектуального аналізу.

Наступним кроком є визначення розмірності, яка буде використовуватися для пошуку правил. В даному випадку ми визначаємо для пошуку вимір «ProductDim».

Для пошуку правил необхідно вказати ключ у структурі виміру. За замовчуванням встановлено ключ розмірності, але можна вибрати будь-який інший атрибут. В даному випадку влаштовує ключ виміру за замовчуванням «Product Id»

Далі можна побачити вибрані атрибути виміру, які є цікавими для даної системи, як поля введення для пошуку правил: Назва, Ціна, Форма, Обсмажування, Сортування, Тип, Вага та кількість одиниць виміру.

Потрібно додати вкладену таблицю - Регіон.

Для пошуку правил необхідно вказати ключ у структурі вимірювань. За замовчуванням встановлено ключ вимірювання, але можна вибрати будь-який інший атрибут. В даному випадку влаштовує ключ виміру за замовчуванням «Id Region».

Далі потрібно вибрати атрибути виміру, які цікавлять, як поля введення для пошуку правил: Назва та кількість одиниць виміру.

На рис. 36 показано вікно для графічного відображення зв'язків між множинами елементів.

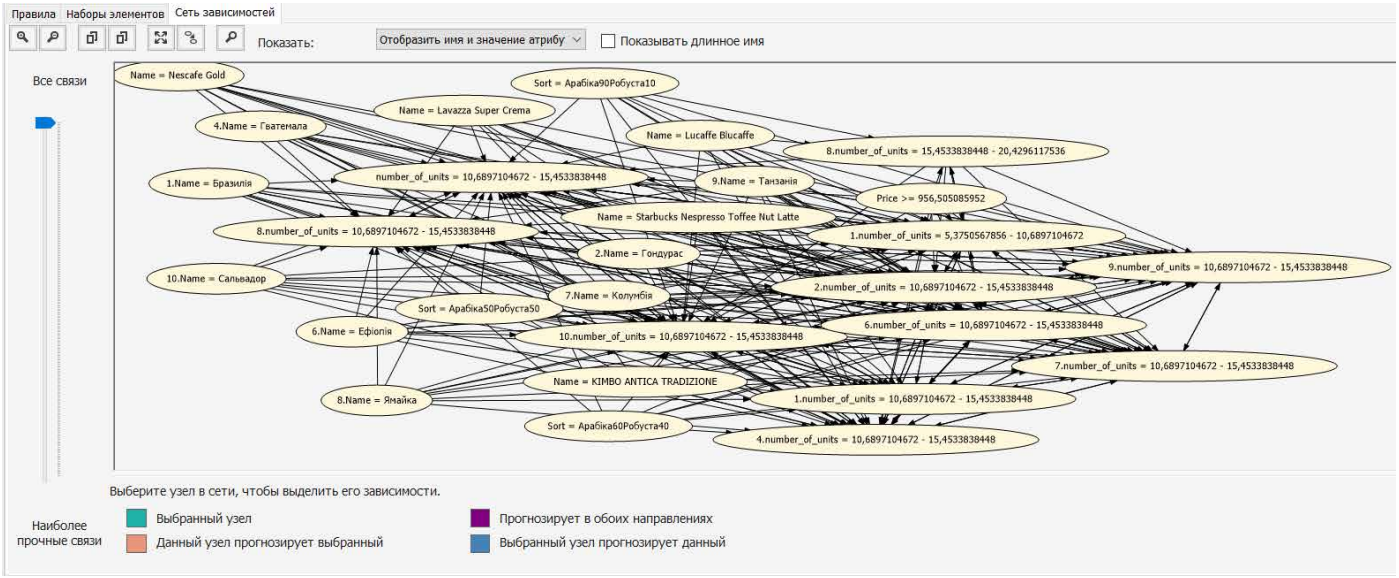


Рис. 36 Вікно для графічного відображення зв'язків між множинами елементів

В результаті було створено правила, які описують дані і показані на рис 37.

В...	Важность	Правило
1,000	0,806	Name = KIMBO ANTICA TRADIZIONE, 2 = Гондурас -> number_of_units = 10,6897104672 - 15,4533838448
1,000	1,505	Name = Lucaffe Blucaffe -> 8 = 15,4533838448 - 20,4296117536
1,000	0,806	Name = KIMBO ANTICA TRADIZIONE, 1 = Бразилія -> number_of_units = 10,6897104672 - 15,4533838448
1,000	1,204	8 = 15,4533838448 - 20,4296117536 -> 1 = 5,3750567856 - 10,6897104672
1,000	0,806	Name = KIMBO ANTICA TRADIZIONE, Weight = 250 - 280 -> number_of_units = 10,6897104672 - 15,4533838448
1,000	0,806	8 = 15,4533838448 - 20,4296117536 -> 10 = 10,6897104672 - 15,4533838448
1,000	0,806	8 = 15,4533838448 - 20,4296117536 -> 9 = 10,6897104672 - 15,4533838448
1,000	0,806	8 = 15,4533838448 - 20,4296117536 -> 7 = 10,6897104672 - 15,4533838448
1,000	0,806	8 = 15,4533838448 - 20,4296117536 -> 6 = 10,6897104672 - 15,4533838448
1,000	0,806	8 = 15,4533838448 - 20,4296117536 -> 4 = 10,6897104672 - 15,4533838448
1,000	0,806	8 = 15,4533838448 - 20,4296117536 -> 2 = 10,6897104672 - 15,4533838448
1,000	0,806	8 = 15,4533838448 - 20,4296117536 -> number_of_units = 10,6897104672 - 15,4533838448
1,000	1,505	Price >= 956,505085952, Name = Lucaffe Blucaffe -> 8 = 15,4533838448 - 20,4296117536
1,000	1,505	Price >= 956,505085952, 1 = 5,3750567856 - 10,6897104672 -> 8 = 15,4533838448 - 20,4296117536
1,000	1,204	8 = 15,4533838448 - 20,4296117536, Price >= 956,505085952 -> 1 = 5,3750567856 - 10,6897104672
1,000	1,505	Price >= 956,505085952, Sort = Арабіка90Робуста10 -> 8 = 15,4533838448 - 20,4296117536
1,000	1,505	Price >= 956,505085952, 10 = 10,6897104672 - 15,4533838448 -> 8 = 15,4533838448 - 20,4296117536
1,000	0,806	8 = 15,4533838448 - 20,4296117536, Price >= 956,505085952 -> 10 = 10,6897104672 - 15,4533838448
1,000	1,505	Price >= 956,505085952, 9 = 10,6897104672 - 15,4533838448 -> 8 = 15,4533838448 - 20,4296117536
1,000	0,806	8 = 15,4533838448 - 20,4296117536, Price >= 956,505085952 -> 9 = 10,6897104672 - 15,4533838448
1,000	1,505	Price >= 956,505085952, 7 = 10,6897104672 - 15,4533838448 -> 8 = 15,4533838448 - 20,4296117536

Рис. 37 Створені правила

Таким чином, найважливішим правилом було те, що з назвою Lucaffe Blucaffe і регіоном Ямайка (за Id Region Jamaica = 8) кількість проданих продуктів буде від 15 до 20.

Це також показано на графічному відображенні взаємозв'язків, яке відображено на рис. 38.

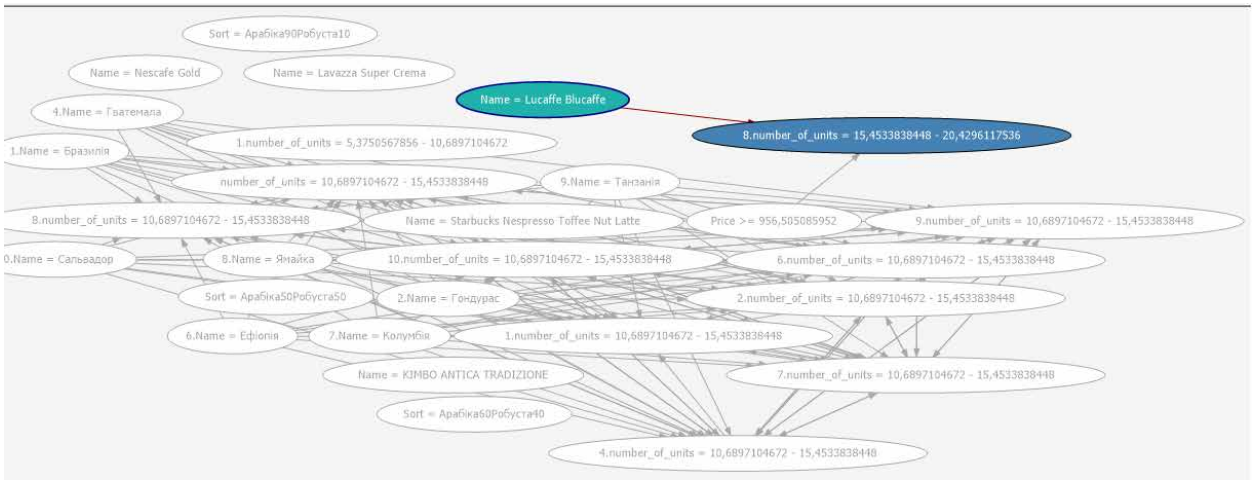


Рис. 38 Графічне відображення взаємозв'язків

Отже метод асоціативних правил дозволяє виявляти залежності або закономірності між різними елементами в наборі даних. Основна мета цього методу — знайти зв’язки між об'єктами, які часто зустрічаються разом, що допомагає зрозуміти приховані взаємозв'язки в даних.

Метод асоціативних правил є корисним для аналізу великих масивів даних і знаходження прихованих взаємозв'язків, які не завжди очевидні на перший погляд.

4.3 Дослідження використання алгоритмів кластеризації

Вивчення алгоритмів кластеризації передбачає вивчення застосування різних методів групування точок даних у кластери на основі схожості[26]. Нижче наведено структурований підхід до вивчення цієї теми:

- Розуміння кластеризації: Почніть з отримання чіткого розуміння того, що таке що таке кластеризація і чому вона важлива для аналізу даних. Дізнайтеся про різні типи алгоритмів кластеризації, такі як розбиття на частини, ієрархічна кластеризація, кластеризація на основі щільності та кластеризація на основі моделей.

- Вивчаємо алгоритми: Пориньте в деталі загальновживаних алгоритмів кластеризації включаючи K-середні, ієрархічну кластеризацію (агломеративну

та дивізіональну), DBSCAN, OPTICS та моделі гауссової суміші (GMM). Ви зрозумієте, як працює кожен алгоритм, їх переваги, обмеження та придатність для різних типів даних і застосувань.

- Метрики оцінювання: Дізнайтеся про метрики оцінювання, які використовуються для оцінки якості

результатів кластеризації, такі як оцінка силуету, індекс Девіса-Болдіна та індекс Данна. Зрозумійте, як ці показники вимірюють компактність і відокремленість кластерів та як інтерпретувати їхні значення.

- Попередня обробка даних: Вивчіть методи попередньої обробки даних та перед застосуванням алгоритмів кластеризації. Це може включати обробку відсутніх значень, масштабування або нормалізацію ознак та зменшення розмірності за допомогою таких методів, як PCA або t-SNE.

- Застосування: Дослідіть реальні застосування кластеризації в різних областях, таких як сегментація клієнтів, сегментація зображень, виявлення аномалій та рекомендаційні системи. Зрозумієте, як кластеризація використовується для виявлення закономірностей, структури та інсайтів у даних.

- Поглиблені теми: Вивчіть поглиблені теми кластеризації, такі як кластеризація кластеризація з обмеженнями, напівкерована кластеризація, ансамблева кластеризація та кластеризація у багатовимірних просторах. Розглянемо дослідження останніх розробок і методів у цій галузі.

- Тематичні дослідження: Проаналізуйте конкретні приклади або наукові роботи, які демонструють практичне використання алгоритмів кластеризації для вирішення конкретних проблем або вирішення бізнес-завдань. Зверніть увагу на застосовані методології, використані набори даних та отримані висновки.

- Інструменти та бібліотеки: Ознайомтеся з популярними інструментами та бібліотеками для кластеризації, такими як scikit-learn, MATLAB, R та бібліотека scipy для Python. Експериментуйте з цими інструментами, щоб

отримати практичний досвід застосування алгоритмів кластеризації до реальних наборів даних.

- Експериментальний дизайн: Якщо ви проводите дослідження в цій галузі, сплануйте експерименти для оцінки продуктивності різних алгоритмів кластеризації або варіацій існуючих алгоритмів. Враховуйте такі фактори, як параметри алгоритму, характеристики набору даних та метрики оцінювання.

Ретельно вивчивши ці аспекти алгоритмів кластеризації, ви розвинете всебічне розуміння методів, застосувань та наслідків цього важливого методу аналізу даних.

Створення моделі Mining Structure за допомогою методу кластеризації SASS. Потрібно вибрати необхідну таблицю розмірностей, ключове поле та атрибути. Ці атрибути відображені на рисунку 39.

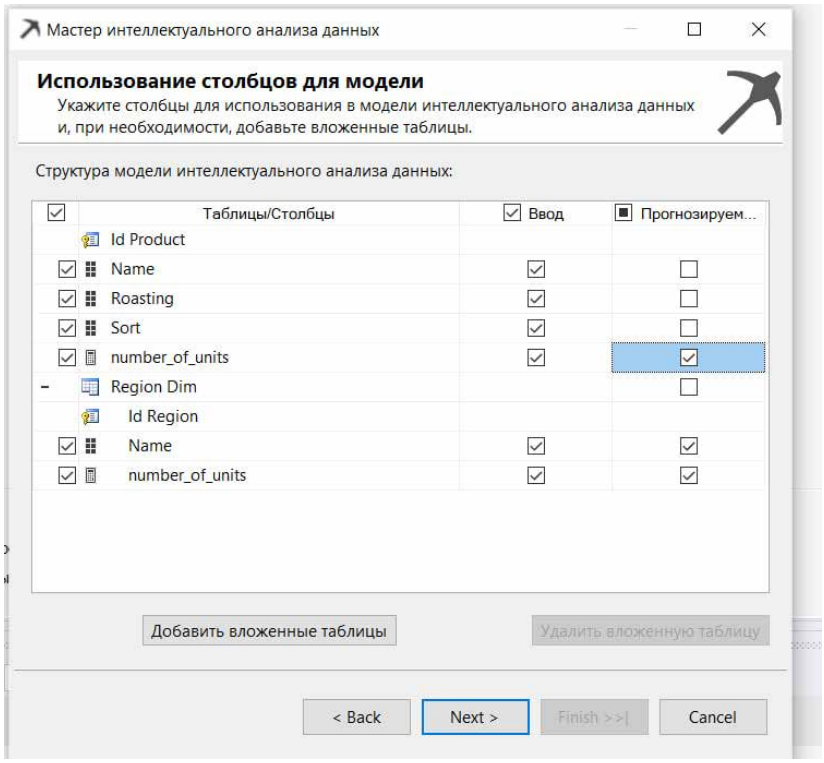


Рис. 39 Атрибути

Після обробки даних куба ми отримуємо результати кластеризації, які відображені на рис. 40.

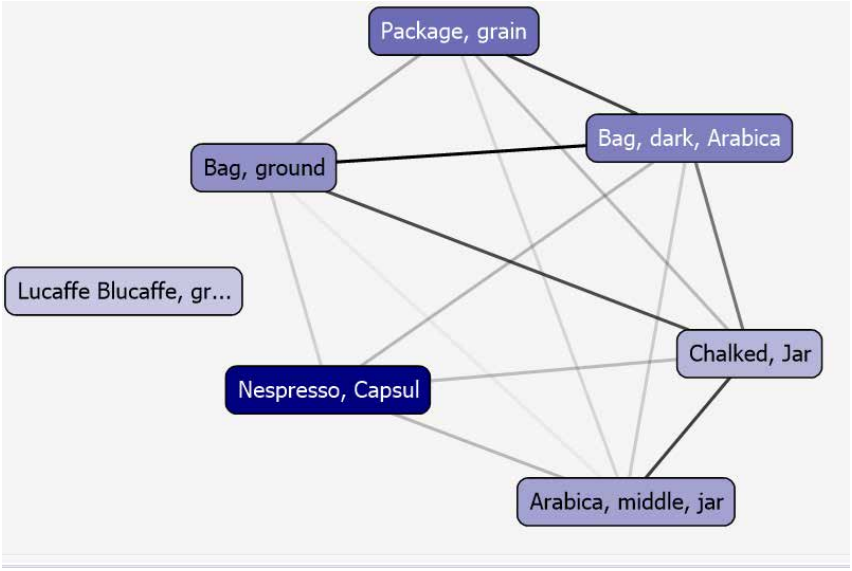


Рис. 40 Результати кластеризації

Результати показують, що 7-й кластер зі значенням 12,88 (максимальне значення 10,37) має значення - Lucaffe Blucaffe, вид - мелена, регіон - Сальвадор, обсмаження - середнє, тип - фасована. Це означає, що саме ця кава буде продаватися найбільше згідно з методом кластеризації. Це також можна побачити в характеристиках кластерів на рис. 41.

Характеристики Lucaffe Blucaffe, ground, middle, Arabica30Robusta70, bag

Переменные	Значения	Вероятность
Name	Lucaffe Blucaffe	
Form	Мелена	
Sort	Арабіка30Робуста70	
Type	Пакет	
Roasting	Темна	
Price	508,9 - 1 000,7	
Region Dim(10).Name	Сальвадор	

Рис. 41 Характеристики кластерів

Алгоритм кластеризації ефективний завдяки своїй здатності групувати дані на основі схожості, що дозволяє виявляти структуру в даних і знаходити закономірності. По-перше, кластеризація дає змогу автоматично знаходити групи об'єктів, які поділяють подібні характеристики, навіть якщо ці групи не були визначені заздалегідь. Це дуже корисно для аналізу даних, коли немає початкової інформації про можливі зв'язки між об'єктами.

ВИСНОВКИ

Була побудована СППР для керівництва платформи з продажу кави, що допомагає керівництву аналізувати дані про продажі, клієнтів і ринок, щоб приймати оптимальні рішення щодо асортименту, маркетингових стратегій, ціноутворення та управління запасами. Це дозволяє максимально ефективно використовувати ресурси компанії, знижуючи ризики надлишку чи нестачі товарів, забезпечуючи безперервність постачання і підвищуючи задоволеність клієнтів.

Використано дві технології OLAP(аналіз даних в режимі реального часу) і Data Mining інтелектуальний аналіз даних.

Data Mining – дозволяє знайти нові закономірності, тоді як OLAP – їх підтвердити або спростувати. В цій роботі, за допомогою Data Mining було виявлено кілька важливих тенденцій. За допомогою методу 1-Rule було виявлено, що Lavazza Super Crema має ймовірність продажу 57,58%, що свідчить про високий попит на цей продукт. Для Lucaffe Blucaffe та KIMBO ANTICA TRADIZIONE ймовірність низького рівня продажів вища, що вказує на низьку популярність цих продуктів. Продукти Nescafe Gold та Starbucks Nespresso Toffee Nut Latte мають схожу ймовірність продажу, що вказує на подібний рівень попиту. За допомогою методу Наївного Байєса було виявлено що Starbucks Nespresso Toffee Nut Latte, KIMBO ANTICA TRADIZIONE та Nescafe Gold є найбільш продуктивними продуктами, що підтверджує високу ймовірність їхнього успіху на ринку. Відповідно до результатів методу Наївного Байєса, Lavazza Super Crema також не є серед найбільш продуктивних, але це може бути через різницю в моделях чи характеристиках даних, що використовуються у цьому методі. Метод асоціативних правил дозволив визначити, що найважливішим правилом є те, що Lucaffe Blucaffe в регіоні Ямайка демонструє продажі від 15 до 20 одиниць (Id Region Jamaica = 8). Це вказує на те, що в Ямайці попит на цей продукт стабільний, і його можна очікувати в таких кількостях. А метод кластеризації

показав, що 7-й кластер (з максимальним значенням 12,88) включає Lucaffe Blucaffe, вид — мелена, регіон — Сальвадор, обсмаження — середнє, тип — фасована. Це означає, що ця кава найбільше продається в групі з подібними характеристиками.

Використання технології OLAP також підтвердила такі закономірності. Продукція Lavazza Super Crema в Бразилії та Сальвадорі показує стійкий тренд зростання продажів у 2023-2024 роках. Це може підтвердити висновки, отримані за допомогою 1-Rule та асоціативних правил, що цей продукт має високий попит. OLAP також дозволяє порівнювати продуктивність різних постачальників, наприклад, для Art Coffee Shop і Coffee Roasters, і бачити, як їхні продукти продаються в різних регіонах і за різні періоди часу.

В результаті виконання магістерської роботи була розроблена система підтримки прийняття рішень (СППР) для керівництва платформи з продажу кави, що використовує OLAP-технології для аналізу великого обсягу даних та дозволяє реалізовувати потоки даних з СД у пусте СД та аналізувати великий обсяг інформації для досконалого вивчення всіх особливостей продажу кави.

Перший етап роботи включав опис предметної області, аналіз існуючої системи обліку продажів і управління товарними запасами на підприємстві, а також огляд наявних рішень і виявлених недоліків.

На другому етапі було проведено детальний аналіз предметної області за допомогою діаграми прецедентів, сформовано список акторів, визначено вимоги до майбутньої системи та описано її топологію.

Третій етап передбачав детальне описання структури джерела інформації — сховища даних. З'ясовано поняття про OLAP, спроектовано сховище даних, яке наповнюється з уже існуючої системи даних (СД). Також було розроблено звітну інформацію та розраховано ключові показники ефективності (KPI).

На четвертому етапі було створено і розгорнуто гіперкуб на основі сховища даних, реалізовано передачу даних із СД за допомогою служби Data Flow. За допомогою Mining Structure і його методів було проведено аналіз

даних з використанням технік Data Mining для отримання цінних інсайтів і прогнозів.

Під час виконання магістерської роботи було використано СУБД SQL Server Management Studio для створення та наповнення СД, середовище розробки Visual Studio з інтегрованими службами Integration Services для аналізу даних та Microsoft Power BI - для візуалізації та аналізу даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Системи підтримки і прийняття рішень (DSS) – [Електронний ресурс]. - Режим доступу: https://1694631.site123.me
2. Що таке data mining (аналіз даних)? – [Електронний ресурс]. - Режим доступу: https://futurenow.com.ua/shho-take-data-mining-analiz-danyh/
3. Salesforce – [Електронний ресурс]. - Режим доступу: https://www.salesforce.com
4. HubSpot – [Електронний ресурс]. - Режим доступу: https://www.hubspot.com
5. UML Distilled: A Brief Guide to the Standard Object Modeling Language – [Електронний ресурс]. - Режим доступу: https://pja.mykhi.org/0sem/MAS/books/Addison%20Wesley%20-%20UML%20Distilled,%202nd%20Edition.pdf
6. Class Diagram | Unified Modeling Language (UML) – [Електронний ресурс]. - Режим доступу: https://www.geeksforgeeks.org/unified-modeling-language-uml-class-diagrams/
7. Sequence Diagrams – Unified Modeling Language (UML) – [Електронний ресурс]. - Режим доступу: https://www.geeksforgeeks.org/unified-modeling-language-uml-sequence-diagrams/
8. Activity Diagrams – Unified Modeling Language (UML) [Електронний ресурс]. - Режим доступу: https://www.geeksforgeeks.org/unified-modeling-language-uml-activity-diagrams/
9. UML - Use Case Diagrams – [Електронний ресурс]. - Режим доступу: https://www.tutorialspoint.com/ uml/uml_use_case_diagram.htm
10. Primary and Secondary Actors (Use Case Diagram) – [Електронний ресурс]. - Режим доступу: https://www.softwareideas.net/primary-and-secondary-actor-use-case-diagram

11. Client Server Architecture – Detailed Explanation – [Електронний ресурс]. - Режим доступу: https://www . interviewbit.com/blog/client-server-architecture/
12. Сховище даних – [Електронний ресурс]. - Режим доступу: https://pidru4niki.com/74246/informatika/shovische_danih
13. OLAP-системи. – [Електронний ресурс]. - Режим доступу: https://pidru4niki.com/1670032447784/informatika/olap-sistemi
14. Що таке MOLAP (багатовимірний OLAP) у сховищі даних? – [Електронний ресурс]. – Режим доступу: https://www.guru99.com/uk/multidimensional-online-analytical processing.html
15. Microsoft for Developers Blog - [Електронний ресурс]. – Режим доступу: https://visualstudio.microsoft.com
16. Multidimensional Model Databases (SSAS) - [Електронний ресурс]. – Режим доступу: https://learn.microsoft.com/uk-ua/analysis-services/multidimensional-models/multidimensional-model-databases-ssas?view=asallproducts-allversions
17. What is Data Mining? Key Concepts, How Does it Work? - [Електронний ресурс]. – Режим доступу: https://www.upgrad.com/blog/what-is-datamining-key-concepts-how-does-it-work/
18. Mining Structures (Analysis Services - Data Mining) - [Електронний ресурс]. – Режим доступу: https://learn.microsoft.com/en-us/analysis-services/data-mining/mining-structures-analysis-services-data-mining?view=asallproducts-allversions
19. Microsoft Power BI Desktop - [Електронний ресурс]. – Режим доступу: https://www.microsoft.com/uk-ua/download/details.aspx?id=58494
20. Data Mining Wizard (Analysis Services - Data Mining) - [Електронний ресурс]. – Режим доступу: https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-wizard-analysis-services-data-mining?view=asallproducts-allversions

21. Ключові показники ефективності (KPI) для торгівельного бізнесу - [Електронний ресурс]. – Режим доступу: https://keepincrm.com/kpi-calculation
22. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann Publishers.
23. Download SQL Server Management Studio (SSMS) - [Електронний ресурс]. – Режим доступу: https://learn.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver16
24. Naive Bayes Classifiers - [Електронний ресурс]. – Режим доступу: https://www.geeksforgeeks.org/naive-bayes-classifiers/
25. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22(2), 207-216.
26. Cluster Analysis Guide with Examples - [Electronic resource]. - Access mode: https://www.resonio.com/market-research/cluster-analysis/