

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ

Інститут (факультет) Автоматизації і комп'ютерних систем
Кафедра Інформаційних технологій, штучного інтелекту і кібербезпеки

«До захисту в ЕК»

Директор інституту (декан факультету)

Андрій Форсюк
(ім'я та прізвище)

«12» лютого 2024р.

«До захисту допущено»

Завідувач кафедри

Сергій Грибков
(ім'я та прізвище)

«12» лютого 2024р.

КВАЛІФІКАЦІЙНА РОБОТА
НА ЗДОБУТТЯ ОСВІТНЬОГО СТУПЕНЯ МАГІСТРА

зі спеціальності 122 «Комп'ютерні науки»

(код та назва спеціальності)

освітньо-професійної програми Інформаційні управляючі системи та технології

на тему: «Дослідження та розробка інформаційно – аналітичної системи для підтримки діяльності ТОВ «Світ каменю»»

Виконав: здобувач 2 курсу, групи ІС-2-3М

Скиба Владислав Олегович

(прізвище, ім'я, по батькові повністю)

(підпис)

Керівник М'якшило Олена Михайлівна

(прізвище, ім'я та по батькові повністю)

(підпис)

Консультанти Олена Харкянен

(ім'я та прізвище)

(підпис)

Консультанти

(ім'я та прізвище)

(підпис)

Рецензент Александр Пупіна

(ім'я та прізвище)

(підпис)

Я як здобувач(ка) Національного університету харчових технологій розумію і підтримую політику університету з академічної доброчесності. Я не надавав(-ла) і не одержував(-ла) незарядженої допомоги під час підготовки цієї роботи. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Здобувач

(підпис)

Київ - 2024

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ

Інститут (факультет) автоматизації і комп'ютерних систем
 Кафедра Інформаційних технологій, штучного інтелекту і кібербезпеки
 Освітній ступінь магістра
 Спеціальність 122 «Комп'ютерні науки»
(код і назва)
 Освітньо-професійна програма Інформаційні управляючі системи та технології
(назва)

ЗАТВЕРДЖУЮ

Завідувач

Кафедри Інформаційних технологій

штучного інтелекту і кібербезпеки

Грибков С. В.

«19»

Грудня

2024 року

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧА

Скиба Владислав Олегович

(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження та розробка інформаційно – аналітичної системи для підтримки діяльності ТОВ «Світ каменю»

керівник роботи М'якшило Олена Михайлівна,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом закладу вищої освіти від «19» Грудня 2023 р. № 10006-КС

2. Строк подання здобувачем роботи: 02.02.2024

3. Вихідні дані до роботи: Інформація про підприємство ТОВ «Світ каменю», метод кластерного аналізу, алгоритми інтелектуального аналізу даних

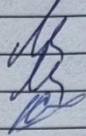
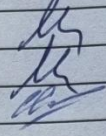
4. Зміст пояснювальної записки (перелік питань, які потрібно розробити):

Дослідження діяльності підприємства ТОВ «Світ каменю» та визначення задач планування асортименту, огляд методів інтелектуального аналізу, реалізація кластерного аналізу

5. Перелік графічного матеріалу:

Графічне представлення результатів кластерного аналізу.

6. Консультанти розділів роботи:

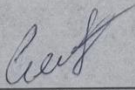
Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	М'якшило О.М., доцент		
2	М'якшило О.М., доцент		
3	Харкянен О.В., доцент		

7. Дата видачі завдання: 19.12.23

КАЛЕНДАРНИЙ ПЛАН


№	Назва етапів виконання кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Дослідження діяльності підприємства ТОВ «Світ каменю»	20.12.2023	Виконано
2	Огляд існуючих методів інтелектуального аналізу даних	03.01.2024	Виконано
3	Генерація вітрини даних для подальшого аналізу	14.01.2024	Виконано
4	Проведення кластерного аналізу на основі інформації з вітрини даних	19.01.2024	Виконано
5	Формування рекомендацій на основі кластерного аналізу	21.01.2024	Виконано
6	Оформлення роботи	22.01.2024	Виконано
7	Розробка презентації	23.01.2024	Виконано
8	Передзахист	24.01.2024	Виконано
9	Оформлення автореферату	31.01.2024	Виконано
10	Подати до захисту	02.02.2024	Виконано

Здобувач


 (підпис)

 Скиба В.О.
 (прізвище та ініціали)

Керівник роботи


 (підпис)

 М'якшило О.М.
 (прізвище та ініціали)

АНОТАЦІЯ

Метою магістерського дослідження полягає у застосуванні інтелектуального аналізу даних для формування рекомендаційних стратегій асортиментного планування для підприємства у складі інформаційно-аналітичної системи

Етапи інтелектуального аналізу даних поділити на два пункти.

Перший це проаналізувати діяльність ТОВ «Світ каменю»

Провести кластерний аналіз даних, на основі якого можна сформувати рекомендаційну стратегію планування асортименту.

Кваліфікаційна робота містить 71 сторінок, 3 таблиць, 20 рисунків і 15 літературних джерела.

КЛЮЧОВІ СЛОВА: КЛАСТЕРНИЙ АНАЛІЗ, ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, ВІТРИНА ДАНИХ, MS SQL SERVER, ПІДПРИЄМСТВО, ДОСЛІДЖЕННЯ, АНАЛІЗ.

ABSTRACT

The purpose of the master's research is to apply intelligent data analysis to form recommendation strategies of assortment planning for the enterprise as part of the information and analytical system

The stages of intelligent data analysis can be divided into two points.

The first is to analyze the activities of Svit Kamenyu LLC

Conduct a cluster analysis of data, on the basis of which it is possible to form a recommended strategy for assortment planning.

The qualification work contains 71 pages, 3 tables, 20 figures and 15 literary sources.

KEY WORDS: CLUSTER ANALYSIS, INTELLECTUAL DATA ANALYSIS, DATA WAREHOUSE, MS SQL SERVER, ENTERPRISE, RESEARCH, ANALYSIS.

ЗМІСТ

ВСТУП	8
РОЗДІЛ 1. ДОСЛІДЖЕННЯ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА ТОВ «СВІТ КАМЕНЮ» ТА ВИЗНАЧЕННЯ ЗАДАЧ ПЛАНУВАННЯ АСОРТИМЕНТУ..	11
1.1 Загальна характеристика ТОВ «Світ каменю»	11
1.2 Організаційна структура підприємства	12
1.3 Стан автоматизації підприємства	17
1.4 Розробка функціональної моделі існуючих бізнес - процесів	18
1.4.1. Виявлені проблеми в результаті моделювання	21
1.5 Огляд систем та інструментів інтелектуального аналізу даних	23
1.6 Постановка задачі дослідження	30
Висновок до розділу 1	31
РОЗДІЛ 2. ОГЛЯД МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ...	32
2.1. Поняття інформаційно-аналітичної діяльності	32
2.2 Огляд існуючих методів інтелектуального аналізу	35
2.2.1 Кластерний аналіз	38
2.3 Алгоритм k-means	42
2.4 Огляд ПП «Orange»	44
2.5 Загальні методи	47
2.6 Огляд методів інтелектуального аналізу даних	49
2.6.1 Прогнозування	50
2.6.2 Аналіз часових рядів	53
2.7 Методи прогнозування витрат	54
Висновок до розділу 2	57

РОЗДІЛ 3. ЗАСТОСУВАННЯ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ПЛАНУВАННЯ АСОРТИМЕНТУ	58
3.1 Розробка вітрини даних	58
3.2 Кластерний аналіз даних	61
Висновок до розділу 3	67
ВИСНОВКИ	69
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	70

ВСТУП

Актуальність теми. У сучасному світі, де технології постійно розвиваються, ефективне керування компаніями неможливе без застосування новітніх комп'ютерних систем. Вибір правильного програмного забезпечення стає вирішальним фактором для успішності автоматизації бізнес-процесів. В умовах інтенсивної конкуренції та обмежених ресурсів, важливим є впровадження інноваційних, більш продуктивних і менш витратних методів управління. Це стає актуальним як для управління компанією в цілому, так і для окремих відділів, які відповідають за робочі процеси.

Інтелектуальний аналіз даних (ІАД, англ. Data Mining) є фундаментальним у виявленні інсайдів з великих масивів даних, використовуючи перехресні підходи машинного навчання, статистики та баз даних для ідентифікації та тенденцій. Основна його мета полягає не лише у вилученні ключової інформації, але й у перетворенні цієї інформації на інтуїтивно зрозумілий формат. Процес включає ряд етапів: від управління даними і їх оптимізації, до розробки моделей, детального аналізу отриманих даних і їх візуального представлення. Крім того, ІАД включає в себе такі важливі елементи, як кластеризація. Кластеризація дозволяє групувати схожі об'єкти на основі їхніх характеристик, сприяючи більш ефективному аналізу даних.

Кластерний аналіз, як частина ІАД, відіграє ключову роль у групуванні та сегментації даних. Цей метод дозволяє класифікувати об'єкти на основі їхніх характеристик, групуючи схожі об'єкти разом. Це дає можливість підприємству ефективно сегментувати ринки, клієнтів, продукти або будь-які інші набори даних, що сприяє прийняттю обґрунтованих стратегічних рішень. Використання кластерного аналізу допомагає краще зрозуміти структуру даних, виявити нові можливості та забезпечити більш цілеспрямовану та ефективну діяльність підприємства.

Зв'язок роботи з науковими програмами, планами, темами кафедри, університету. Наукова робота виконувалась згідно з науково-дослідною роботою на кафедрі інформаційних технологій, штучного інтелекту і кібербезпеки «Дослідження та використання сучасних інформаційних технологій для виконання функцій та завдань виробничого і організаційного управління підприємств харчової галузі» (0120U105386 2020–2025 рр.) Національного університету харчових технологій.

Об'єкт дослідження. Інформаційний ресурс, накопичений в процесі функціонування інформаційної системи підприємства ТОВ «Світ каменю».

Предмет дослідження. Застосування методів інтелектуального аналізу даних у процесі прийняття управлінських рішень в ТОВ «Світ каменю».

Мета та задачі дослідження. Мета магістерського дослідження полягає у застосуванні аналізу даних для формування рекомендацій з планування асортименту підприємства ТОВ "Світ каменю". Для досягнення цих цілей потрібно виконати ряд наступних завдань:

- Дослідити діяльність підприємства ТОВ «Світ каменю».
- Дослідити існуючі системи управління виробництвом на ТОВ «Світ каменю»;
- Провести аналіз результатів досліджень та визначити завдання для подальшого інтелектуального аналізу даних та методи їх реалізації.
- Провести кластерний аналіз даних, на основі якого можна сформулювати рекомендацію по плануванню асортименту в рамках інформаційно-аналітичної системи.

Методи дослідження. Для розробки та створення вітрини даних використовувався ERwin Data Modeler. Вітрина даних були реалізовані за допомогою MS SQL Server. Аналіз даних та підготовка управлінських рішень були здійснені за допомогою інструменту Orange.

Наукова новизна. Наукова новизна даного дослідження полягає в:

- 1) Використання інтелектуального аналізу даних для формування рекомендацій по плануванню асортименту продукції ТОВ «Світ каменю».
- 2) Вибір оптимальної моделі для проведення кластерного аналізу в середовищі Orange.

Практичне значення отриманих результатів. Полягає в тому, що вони можуть бути використані підприємством для оптимізації бізнес-процесів за допомогою інтелектуального аналізу даних та кластерного аналізу. Це дозволить підвищити ефективність управління даними, поліпшити прийняття рішень на основі даних і забезпечити більш точне планування асортименту. Результати дослідження надають практичні рекомендації щодо впровадження та використання кластерного аналізу, що можуть бути застосовані в різних індустріях, включаючи роздрібну торгівлю та виробництво.

Структура та обсяг магістерської роботи. Магістерська робота містить вступ, 3 розділи, висновки 4, список літератури з найменувань 15. Основна частина викладена на сторінках друкованого тексту 70. Робота включає ілюстрації 20 та таблиці 3.

РОЗДІЛ 1. ДОСЛІДЖЕННЯ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА ТОВ «СВІТ КАМЕНЮ» ТА ВИЗНАЧЕННЯ ЗАДАЧ ПЛАНУВАННЯ АСОРТИМЕНТУ

1.1 Загальна характеристика ТОВ «Світ каменю»

ТОВ "Світ Каменю" - це виробник продукції з природного каменю, який пройшов шлях успіху протягом понад 11 років. Історія компанії почалася більше 11 років тому, коли засновник, сповнений вірою в потужність та красу природного каменю, вирішив створити свою власну компанію.

У початковому етапі компанія зосереджувалася на виробництві підвіконь та облицювання камінів. Важливим кроком було вирішення піти на повне виробництво, що дозволило контролювати якість кожного виробу та втілювати у життя індивідуальні проекти клієнтів.

Протягом наступних років "Світ Каменю" активно розвивався. Інвестиції в сучасне високотехнологічне обладнання та залучення кваліфікованих фахівців стали ключовими факторами у забезпеченні конкурентоспроможності компанії. Розширення асортименту продукції на виробництво сходів, плиток лицювальних, стільниць та інших виробів, робило "Світ Каменю" важливим учасником ринку.

Компанія завоювала визнання завдяки своїй унікальності та креативності в дизайні. Використання різноманітних видів каменю з усього світу - від італійського мрамору до українських гранітів - дозволяє надавати проектам елегантність та статус.

Основні етапи історії та розвитку ТОВ «Світ каменю»

- Заснування було в 2007 році, підприємство ТОВ «Світ каменю» заснувалась на Волині, та почали вони своє виробництво з граніту, травертину.
- Розвиток підприємства почався у 20012 році, а саме з удосконалення високотехнологічного обладнання, що дозволяло їм швидко виконувати замовлення високої складності.

- Розширення послуг, окрім реалізації натурального каменю, ТОВ «Світ каменю» почало надавати наступні професійні послуги:
 - Виготовлення виробів з природнього каменю за індивідуальним замовленням;
 - Порізка каменю пилами діаметром від 350 мм до 3000 мм.;
 - Термооброблення виробів з каменю;
 - Фігурна обробка виробів з каменю;
 - Доставка виробів з каменю;
- Розширення асортименту продукції у 2015 році підприємство продовжило розширювати свій асортимент продукції та збільшувати свою продуктивність на ринку. Було реалізовано виробництво з закордонних матеріалів такі як: Італійський мармур, штучний кварц, Онікс, засоби для догляду.
- Розширення виробництва у 2018 році розширило своє виробництво, оскільки кількість клієнтів зростала з кожним роком, було прийняте рішення розширити виробництво і оновити високотехнологічне обладнання для більш продуктивної роботи.

В даний момент коли почалась повномасштабна війна ТОВ «Світ каменю», на деякий час призупинив виробництво, частина працівників пішли на фронт, але попри складне становище підприємство відновило виробництво та надання послуг, що допомагає розвитку економіки України.

В майбутньому ТОВ "Світу Каменю" планує впровадження інновацій, розширення асортименту та неперевершені дизайнерські рішення. Компанія готова втілювати в життя нові ідеї та надавати житлу вишуканість та неповторність.

1.2 Організаційна структура підприємства

На рисунку 1.1 наведена структура підприємства ТОВ «Світ каменю», яка формувалась під впливом багатьох факторів, такі як: диверсифікація виробництва, розширення клієнтської бази.

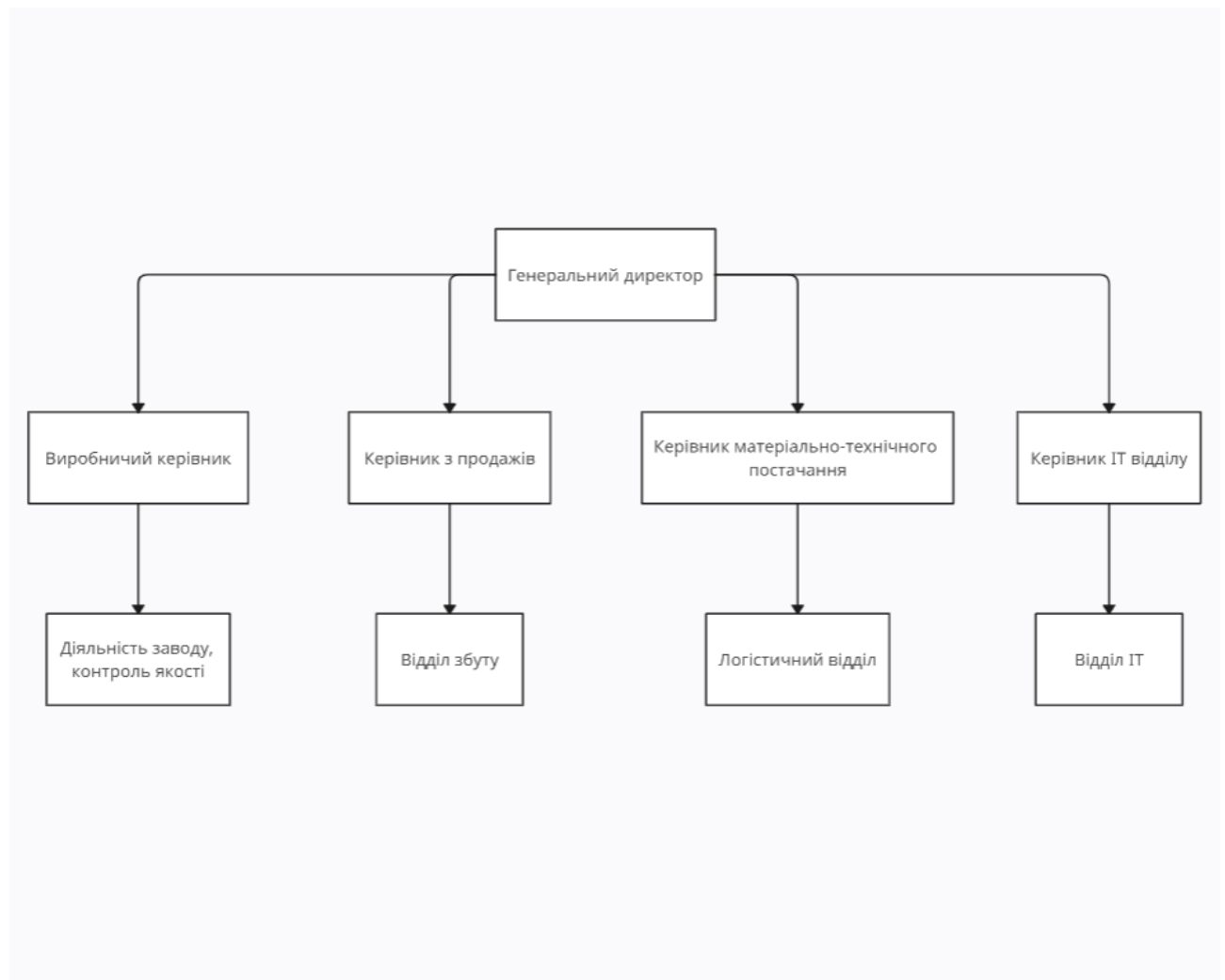


Рисунок 1.1. Організаційна структура підприємства

Підприємство не є великим, але з розвиненою організаційною структурою, яка має різні підрозділи та рівні управління.

- Генеральний директор: Відповідає за стратегічне управління компанією та реалізацію корпоративної візі. Приймає ключові стратегічні рішення та встановлює цілі підприємства.
- Виробничий Керівник: Наглядає за виробничим процесом та впровадженням технологічних інновацій. Керує діяльністю заводу та забезпечує ефективність виробництва. Відповідає за якість виробів та контроль за виробничим процесом.

- Керівник з продажів та маркетингу: Організовує стратегії продажів та розвиток ринків збуту. Керує комерційною діяльністю та взаємодіє з клієнтами. Розробляє маркетингові стратегії, веде рекламні кампанії та вивчає ринкові тенденції.
- Керівник з матеріально – технічного постачання: Організовує управління та координацію централізованого постачання сировини та матеріалів, а також відповідає за реалізацію продукції власного виробництва. Керує діяльністю логістичного відділу.
- Керівник інформаційних технологій: Відповідає за впровадження IT-стратегій та забезпечення кібербезпеки. Має під собою команду програмістів, адміністраторів та інженерів з IT-інфраструктури.

Більше детально досліджувалась діяльність IT-відділу. Структура відділу складається з керівника IT відділу, системного адміністратора та розробників програмного забезпечення.

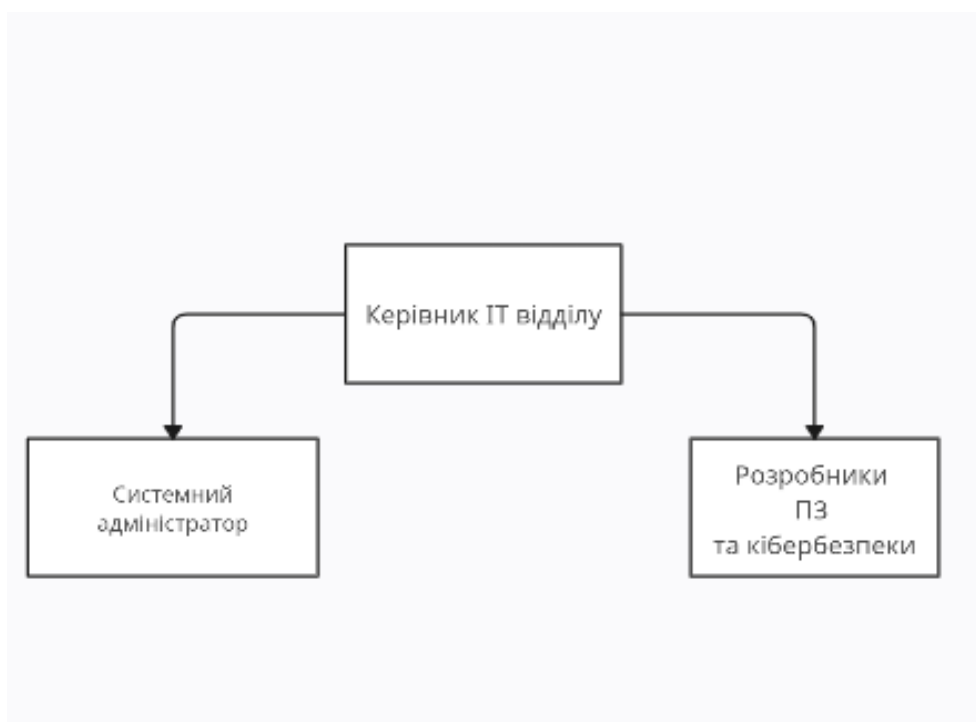


Рисунок 1.2. Організаційна структура IT відділу підприємства

- Директор з інформаційних технологій:
Відповідає за стратегічне керівництво інформаційно-технічним розвитком компанії. Він визначає і впроваджує стратегії та політики в сфері IT,

спрямовані на досягнення цілей підприємства. Директор також відповідає за забезпечення кібербезпеки та ефективного використання технічних ресурсів.

- Системний адміністратор:

Відповідає за щоденне функціонування та підтримку інформаційно-технічної інфраструктури. Йому доручено встановлення, конфігурування та підтримку серверів, мережевого обладнання, а також вирішення технічних проблем, що виникають в процесі роботи.

- Розробник програмного забезпечення та кібербезпеки:

Відповідає за створення програм та інформаційних систем, а також за захист комп'ютерних систем від потенційних загроз. Ця роль включає в себе написання коду, аналіз програмного забезпечення на предмет вразливостей та розробку стратегій кібербезпеки для захисту інформації компанії.

Ці посади в інформаційних технологій є важливими, оскільки вони визначають та забезпечують ефективне функціонування та розвиток ІТ-систем підприємства. Розглянемо кожну посаду окремо:

ІТ-директор:

- Стратегічне керівництво: Розробляє та впроваджує стратегії ІТ-розвитку, які відповідають стратегії підприємства, забезпечуючи технологічну конкурентоспроможність.
- Кібербезпека: Гарантує безпеку та захист інформації від кіберзагроз, дбаючи про конфіденційність, цілісність та доступність даних.
- Оптимізація ресурсів: Керує ефективним використанням технічних ресурсів, включаючи сервери, мережеве обладнання та інші компоненти інфраструктури.

Системний адміністратор:

- Денна експлуатація: Забезпечує неперервне функціонування ІТ-систем щоденно, вирішуючи технічні питання та підтримуючи користувачів.
- Захист інфраструктури: Відповідає за захист серверів та мережі від несанкціонованого доступу та інших загроз.

Розробник програмного забезпечення та кібербезпеки:

- Розробка програм: Відіграє ключову роль у створенні програм та додатків, що відповідають потребам підприємства та його клієнтів.
- Кібербезпека: Розробляє та впроваджує стратегії кібербезпеки для захисту від потенційних кібератак, забезпечуючи безпеку інформаційних ресурсів.

Взаємодія з іншими структурними підрозділами наведена у таблиці 1.1.

Таблиця 1.1 Взаємодія відділу IT з іншими структурними підрозділами

№	Підрозділ	Одержання	Надання
1	Генеральний директор	- Стратегічні звіти від усіх відділів. - Ринкові аналітичні дані. - Звіти про виробничі показники.	- Стратегічні напрямки розвитку. - Корпоративні візії та цілі. - Ключові рішення щодо управління компанією.
2	Виробничий керівник	- Звіти про потреби в сировині та матеріалах від відділу матеріально-технічного постачання. - IT підтримка від відділу IT. - Дані про обсяги продажів від відділу з продажу для планування виробництва.	- Звіти про ефективність виробництва та якість продукції до Генерального директора. - Інформацію про виконання планів виробництва до відділу збуту. - Вимоги до якості та термінів постачання до відділу матеріально-технічного постачання.
3	Керівник з продажу	- Звіти про виробничі можливості від виробничого керівника для планування продажів. - IT підтримка від	- Прогнози продажів та звіти про виконання до Генерального директора. - Замовлення на виробництво до

		відділу ІТ для CRM та інших систем.	виробничого керівника для планування обсягів виробництва.
4	Керівник матеріально-технічного постачання	- Звіти про потреби в матеріалах та сировині від виробничого керівника. - Звіти про статус інвентаризації та потреби в оновленні обладнання від відділу ІТ.	- Забезпечення сировиною та матеріалами для виробництва до виробничого керівника. - Інформацію про закупівлі та запаси до Генерального директора.

1.3 Стан автоматизації підприємства

ТОВ «Світ каменю» використовує наступне програмне забезпечення (ПЗ):

- ERP система в якій знаходиться інформація про продажі, логістику, приходи, витрати в систему, програма - генерує результат, а власник бізнесу на підставі цієї інформації приймає управлінські рішення. Також використовують базу даних MS SQL Sever Manager Studio яка зберігає дані з ERP системи, ця платформа надійна та потужна для обробки даних підприємства і використовується у всіх відділах.
- Підприємство використовує інформаційні системи для управління ресурсами, такими як обладнання, залишки та матеріали.
- API системи для реалізації зв'язку та обміну даними між різними компонентами системи, що в свою чергу спрощує комунікацію з клієнтами.
- Lotus Notes - це система організації корпоративного документообороту. Вона призначена для роботи з інформацією і документами.
- Система штрих-кодування Bar Code призначена для автоматизації логістики на складі, контролю якості відвантаження продукції та

обміну документами відвантаження з SAP, тому на підприємстві де часто є дефекти матеріалу це є хорошим рішенням.

Інформаційні системи використовуються для локальної обробки та роботи з інформацією, що дозволяє співробітникам ефективно взаємодіяти з даними та виконувати завдання безпосередньо на своїх робочих станціях.

1.4 Розробка функціональної моделі існуючих бізнес - процесів

Наведена на рисунку 1.3 функціональна схема описує бізнес – процес діяльності відділу ІТ підприємства коли поступає запит на розробку програмного забезпечення.

Запит підприємства:

- Підприємство дає запит директору з інформаційних технологій, які можуть стосуватися вирішення проблем, вимог до нового функціоналу, оновлення програм тощо.

Потреби підприємства:

- Директор з інформаційних технологій співпрацює з замовником, щоб уточнити потреби та вимоги до проекту(Обсяг проекту, мета, ціль, функціонал)

Планування програмного забезпечення:

- Керівник відділу ІТ співпрацює з розробниками програмного забезпечення, щоб визначити технології які будуть використовуватись та інструменти на основі вимог замовника.

Розробка програмного забезпечення:

- Розробники програмного забезпечення приступають до реалізації створення програмного забезпечення відповідно до вимог та дизайну.

Тестування програмного забезпечення:

- Розробники тестують програмне забезпечення, щоб ПЗ працювало коректно та перевіряють чи відповідає програмне забезпечення вимогам замовника.

Заходи захисту даних:

- Відділ кібербезпеки вживає комплекс заходів для забезпечення безпеки даних. Це включає в себе моніторинг мережі для виявлення потенційних загроз, встановлення та підтримку сучасних систем шифрування, регулярне оновлення програмного забезпечення для усунення вразливостей.

Обслуговування та підтримка програмного забезпечення:

- Системні адміністратори відповідають за ефективне управління і підтримку програмного забезпечення. Вони встановлюють, налаштовують та вдосконалюють апаратне та програмне забезпечення, забезпечують надійність мережі та систем зберігання даних.

Готове програмне забезпечення:

- Замовнику передають готове програмне забезпечення.

Було побудовано діаграму діяльності ІТ відділу підприємства ТОВ «Світ каменю».

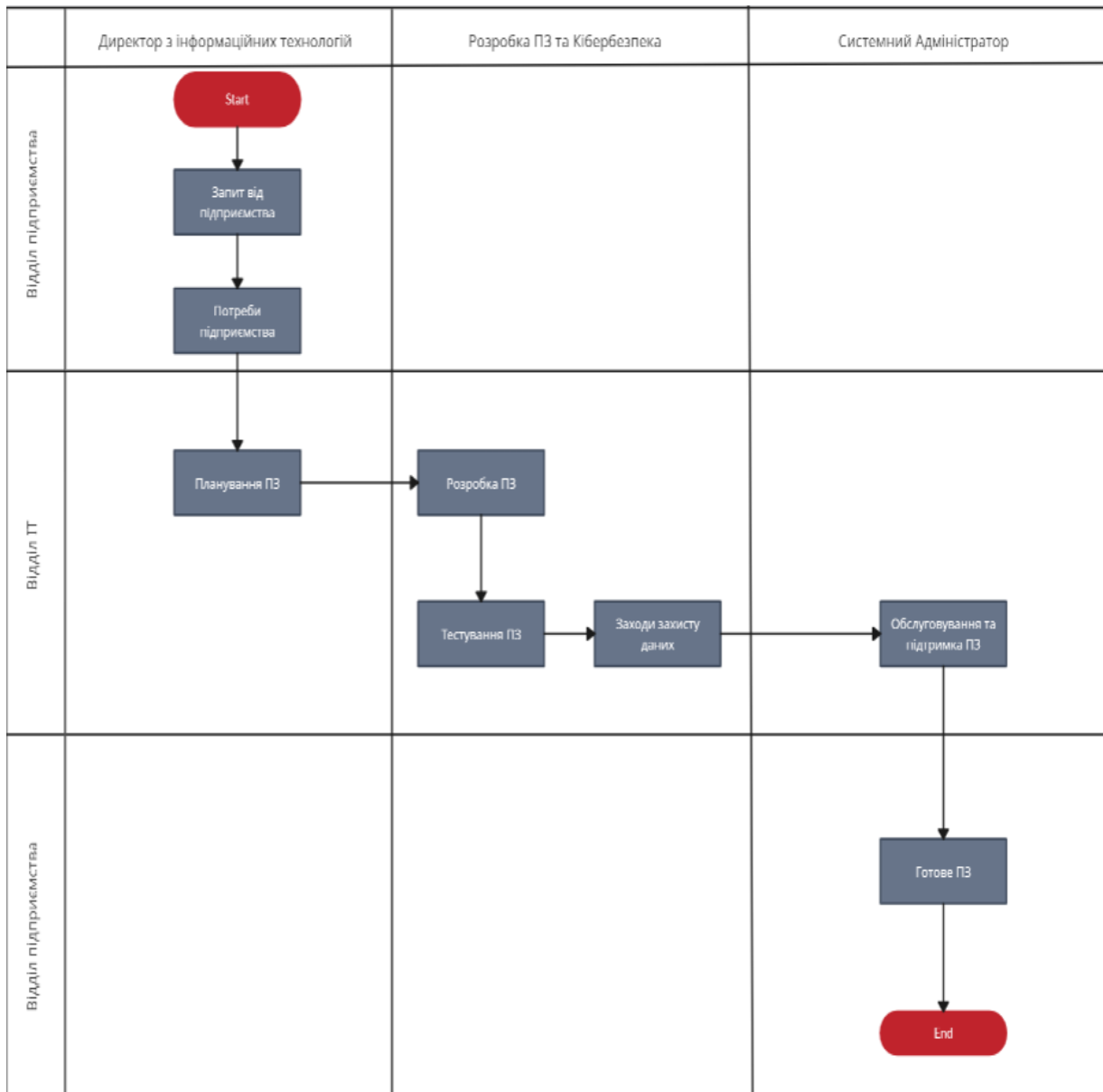


Рисунок 1.3 діаграма діяльності відділу IT підприємства.

Змодельюємо бізнес-процес, що відбувається у виробничому відділі за допомогою діаграми прецедентів (рисунок 1.4)

Діаграма прецедентів демонструє механізм реалізації попередньо визначених процесів у межах системи. Це інструмент прикладного аналізу, який детально відтворює цілісний ланцюг дій, що виконуються окремими учасниками для здійснення конкретних операцій [3].

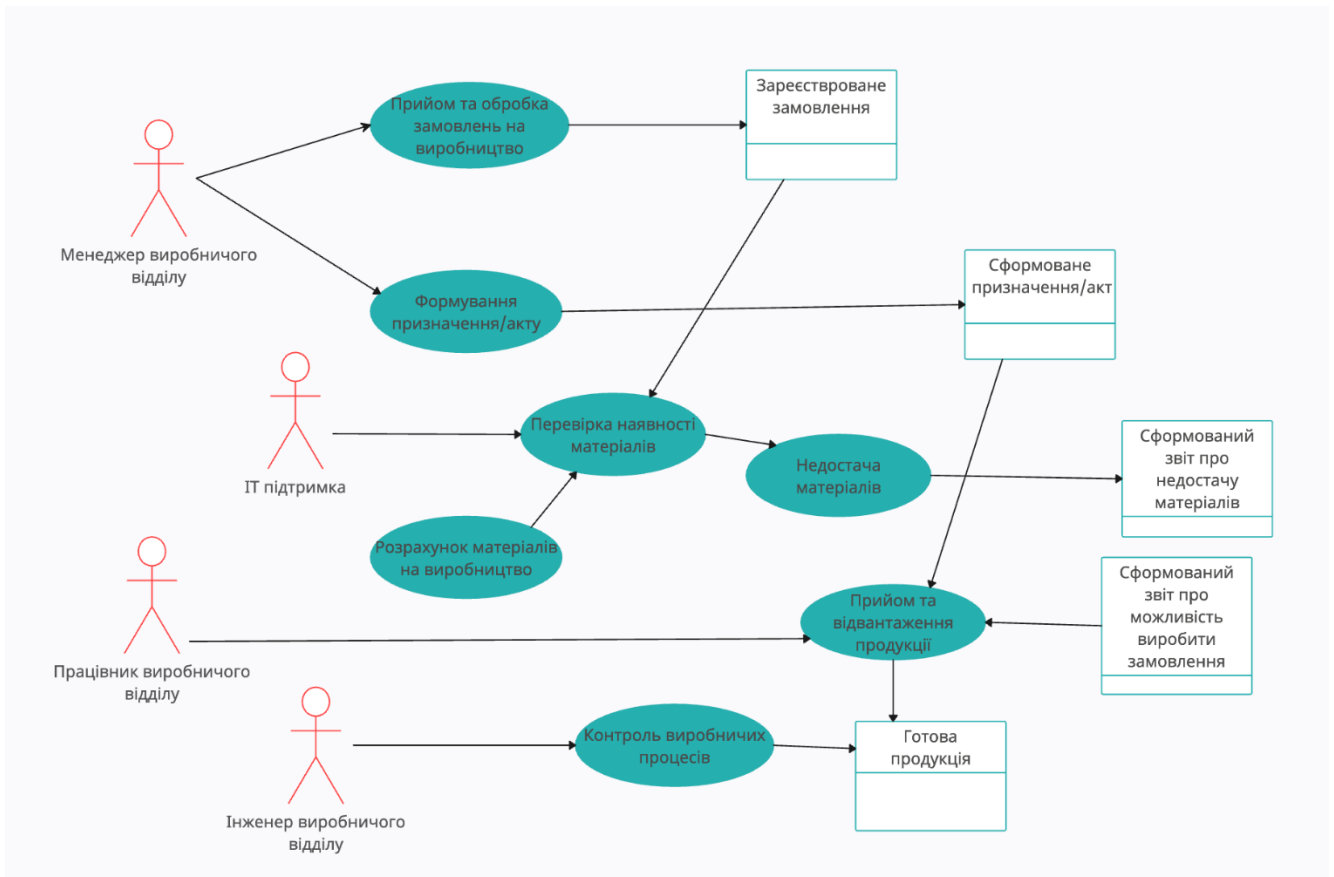


Рисунок 1.4 Діаграма прецедентів для виробничого відділу ТОВ «Світ каменю».

1.4.1. Виявлені проблеми в результаті моделювання

У результаті аналізу стану інформаційних процесів виявлено низку проблем:

- Після розрахунку матеріалів на виробництво, було перевірено чи наявні матеріали на складі, після чого було виявлено нестачу матеріалів.
- Погане планування асортименту на що вказує нестача матеріалів.

Діаграма станів показує різні стани, через які проходить об'єкт або система, та переходи між цими станами, часто у відповідь на визначені події або умови. Метою діаграми станів є надати зрозуміле і точне уявлення про поведінку об'єкта у відповідь на зовнішні впливи, спростити процес розробки та аналізу комплексних систем, та забезпечити чітке розуміння можливих станів, яких може досягти система. Діаграма станів (Рисунок 1.5) демонструє результат розрахунку, при перевірці наявності на складі готової продукції і результат розрахунку недостатньої кількості продукції [2].

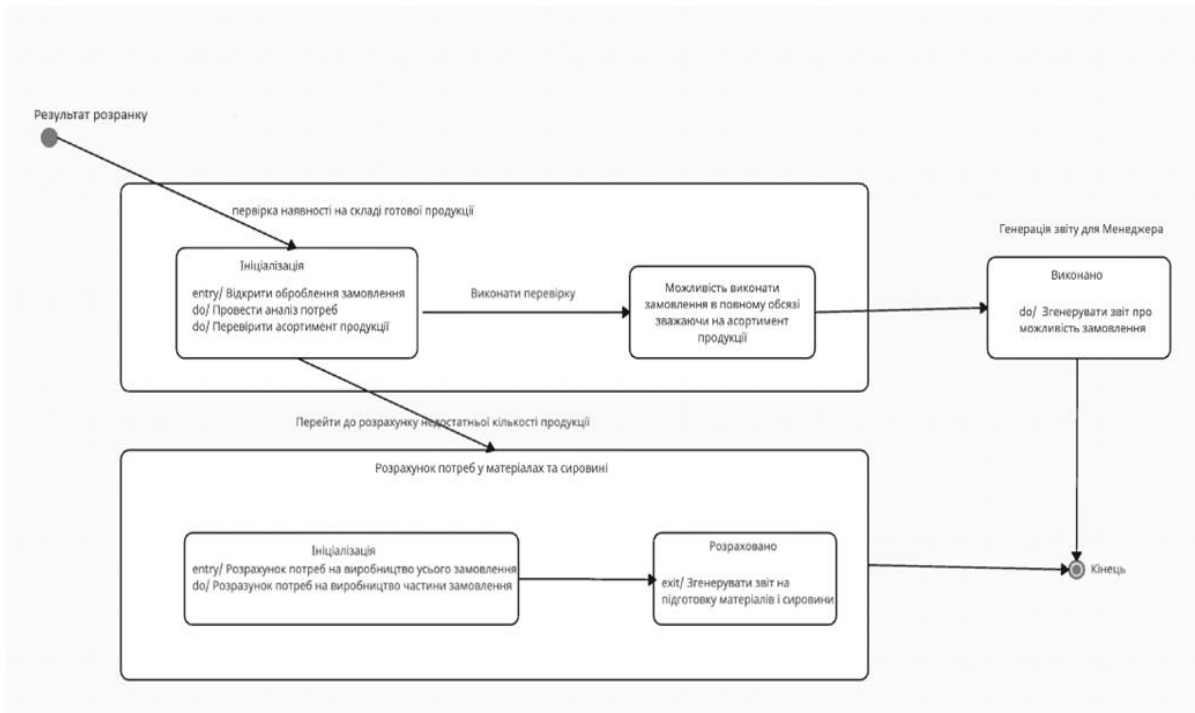


Рисунок 1.5. Діаграма станів для класу «Результат розрахунку».

Діаграма активності для варіантів використання представляє блок-схему, яка ілюструє процес роботи менеджера виробничого відділу з інформаційною системою під час обробки замовлення (рисунок 1.6).

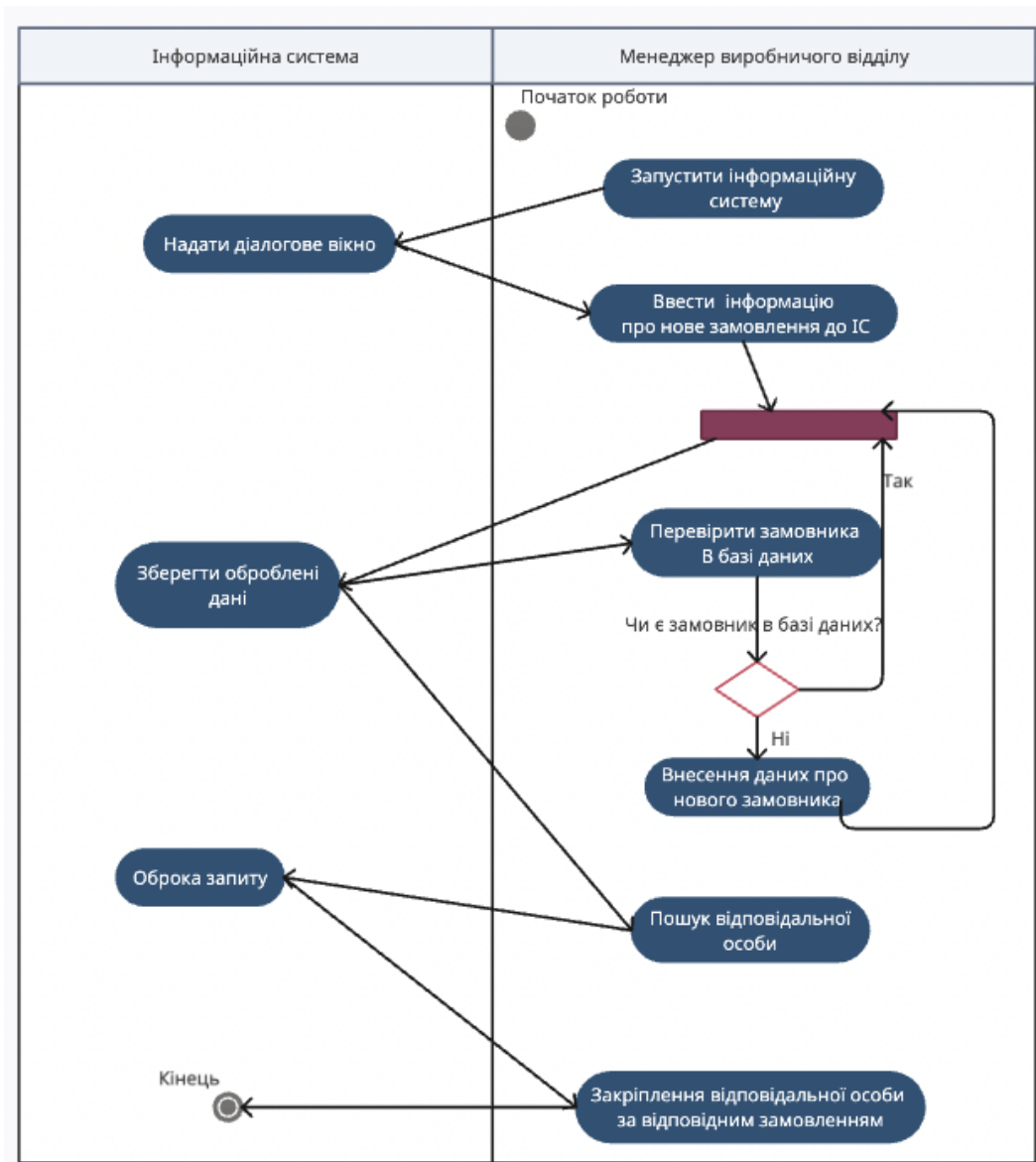


Рисунок 1.6. Діаграма активності для варіанту використання «Прийом та обробка замовлення на виробництво».

1.5 Огляд систем та інструментів інтелектуального аналізу даних

Сучасний бізнес-ландшафт практично не можна уявити без використання автоматизованих інформаційних систем, які сприяють підтримці процесу прийняття рішень. Втім, автоматизація часто обмежується лише окремими аспектами управління на рівні підприємства або відділу і рідко включає в себе передові засоби, зокрема методи інтелектуального аналізу даних, які могли б суттєво підсилити ефективність цих систем [4].

Планування асортименту є ключовим елементом для успіху будь-якого комерційного бізнесу, адже відповідний вибір товарів, які будуть пропонуватись

покупцям, впливає на задоволеність клієнтів, обороти продажів та прибутковість. Ефективне планування асортименту включає аналіз ринкових трендів, попиту споживачів та здатність прогнозувати майбутні зміни у споживацьких вподобаннях.

Застосування інтелектуального аналізу даних дозволяє глибше зрозуміти поведінку покупців, виявити приховані зв'язки між покупками та адаптувати асортимент товарів згідно з цією інформацією. Це може включати оптимізацію запасів, персоналізацію пропозицій та краще управління запасами, зменшуючи надлишки чи дефіцит товарів. В результаті, компанії здатні не лише підвищувати задоволеність клієнтів, але й оптимізувати витрати та підвищити рентабельність.

З самого початку розвитку зарубіжних інформаційних систем, їх фундамент складала принципи ефективності та оптимізації ресурсів, включно з матеріальними та фінансовими. Це відображено в таких системах як IBM InfoSphere Warehouse та IBM SPSS, які зосереджені на інтелектуальному аналізі даних та управлінні ними, а також в IC (управління запасами), MRP (планування потреби в матеріалах), MRP II (планування виробничих ресурсів), та ERP (планування ресурсів корпорації). Це не означає, що розробники ігнорували важливість бухгалтерського обліку в цих системах. Навпаки, бухгалтерський модуль тісно пов'язаний з управлінськими завданнями.

Цей підхід до розробки інформаційних систем відображає глибоке розуміння динамічних вимог сучасного бізнес-середовища у західних країнах, де акцент робиться не тільки на ефективності, але й на адаптивності та інтеграції різних бізнес-процесів. Системи, такі як ERP, демонструють гнучкість у плануванні та координації різних аспектів корпоративної діяльності - від виробництва до фінансів та логістики, що в свою чергу дозволяє компаніям швидше реагувати на зміни ринку та підвищувати свою конкурентоспроможність. Важливо також відзначити, що ця інтеграція і адаптація йде рука об руку з забезпеченням високого рівня даних та аналітичних можливостей, які дозволяють керівництву компанії приймати обґрунтовані та своєчасні рішення [5].

IBM SPSS відомий як високоефективний інструмент для статистичного аналізу та обробки даних, який надає можливість провести інтелектуальний аналіз даних, базуючись на попередніх тенденціях, та забезпечує простоту в використанні.

IBM SPSS, з його розширеними аналітичними можливостями, став невід'ємною частиною для багатьох компаній, які прагнуть глибше зрозуміти свої дані. Завдяки гнучкості в аналізі та легкості в інтерпретації результатів, цей інструмент допомагає організаціям ідентифікувати ключові шаблони та зв'язки в своїх даних, що має величезне значення для прийняття стратегічних рішень. Від ринкових досліджень до фінансового ринку, IBM SPSS надає можливість для комплексного аналізу, підвищуючи точність та ефективність у прийнятті рішень на різних рівнях управління.

Для того, щоб зібрати різноманітні кількісні індикатори та отримати точні висновки з обробки даних, необхідно опрацьовувати значні масиви інформації. У цьому контексті використовуються різні програмні платформи, серед яких SPSS Statistics (Superior Performance Software System) займає провідну позицію. SPSS є провідним продуктом на ринку комерційного статистичного аналізу, виступаючи універсальним інструментом для статистичного та інтелектуального аналізу даних та їх управління.

Пакет SPSS дозволяє здійснювати широкий спектр аналізу даних, і його нові версії ефективно використовуються в різних наукових та бізнес-сферах. Завдяки своєму гнучкому та інтуїтивно зрозумілому інтерфейсу, SPSS робить процес введення даних і їх статистичного аналізу доступним для користувачів різного рівня. Його редактор даних надає зручність для введення та редагування інформації в табличній формі. SPSS також забезпечує можливість створення різноманітних високоякісних графіків та діаграм. З його допомогою можна аналізувати об'ємні файли з тисячами змінних, використовуючи прості меню, таблиці та діалогові вікна, і все це без необхідності використання програмування для запису команд [6].

Функції SPSS дозволяють виконувати ряд завдань: управління даними; конвертацію даних та створення нових змінних; аналітичну обробку даних;

збереження та аналіз даних з опитувань, маркетингових досліджень та продажів. Програмний пакет також сприяє автоматизації процесів створення, збереження та обробки баз даних для різних видів інформації.

Аналітичний процес у SPSS включає етапи: планування; збір даних; надання доступу до даних; підготовка даних для аналізу; проведення аналізу; створення звітів; поширення результатів; автоматизація обробки даних та їх інтерпретація.

Крім роботи з власними типами даних, SPSS може читати дані з майже будь-якого типу файлів для створення звітів у вигляді таблиць, графіків і діаграм, а також для виконання описового статистичного аналізу, складніших статистичних досліджень та моделювання.

Програмний пакет має модульну будову, де кожен модуль представляє інтегрований набір програмних продуктів, що забезпечують повний спектр досліджень – від планування до управління даними, проведення аналізу і представлення результатів.

- Редактор даних: це гнучка система, схожа на електронну таблицю, яка дозволяє визначити, вводити, редагувати і переглядати дані.
- Засіб перегляду результатів: він полегшує перегляд результатів, дозволяючи відображати або приховувати окремі частини виведення, налаштовувати їх порядок, а також переміщати готові для презентації таблиці і діаграмами в іншій програмі чи імпортувати їх з них.
- Багатовимірні таблиці: Використовуються для представлення результатів аналізу. Ці таблиці дозволяють глибше аналізувати дані, змінюючи розташування рядків, стовбців, і шарів, що допомагає виявити важливі висновки, які можуть бути непомітними в стандартних звітах. Також вони дозволяють порівнювати групи, розбиваючи таблиці таким чином, що кожна група відображається окремо.
- Високоякісна графіка: Цей інструмент дозволяє створювати різні види повнокольорових, високодеталізованих діаграм, таких як кругові,

лінійні, гістограми, діаграми розсіювання, тривимірні діаграми та багато інших.

- Доступ до бази даних: Ця функція полегшує завантаження бази даних з будь-якого джерела, використовуючи простий управлінський інтерфейс, що мінімізує зусилля на завантаження.
- Перетворення даних: Ця можливість допомагає підготувати дані до аналізу, даючи можливість виділяти підмножини, об'єднувати категорії, додавати, групувати, зливати, розділяти і транспонувати файли, а також виконувати інші типи перетворень.

Командна мова SPSS. Хоча більшість завдань у SPSS можна виконувати за допомогою інтерфейсу з мишею та діалогових вікон, програма також має потужний командний інтерфейс. Цей інтерфейс дозволяє записувати та автоматизувати рутинні завдання. Крім того, командний інтерфейс надає доступ до деяких функцій, які не доступні через графічний інтерфейс користувача. Повний опис командної мови інтегрований у систему довідки SPSS і також доступний у вигляді окремого PDF-документа "Керівництво по синтаксису", яке можна знайти у розділі «Довідка» [3].

Структура програмного пакета охоплює команди для визначення та перетворення даних, а також вибірку об'єктів. Пакет включає різноманітні методики статистичного аналізу даних, такі як:

- Підсумкові статистики для індивідуальних змінних;
- Частотні аналізи, сумарні статистики та візуалізації для будь-якої кількості змінних;
- Розробка багатовимірних таблиць спряженості з обчисленням мір зв'язку, обчислення середніх, стандартних відхилень та сум за групами;
- Аналіз розсіювання та множинне порівняння;
- Кореляційний та дискримінантний аналіз, однофакторний аналіз розсіювання;
- Загальна лінійна модель (GLM) для аналізу розсіювання;

- Факторний аналіз;
- Кластерний аналіз та ієрархічний кластерний аналіз;
- Ієрархічний лог-лінійний аналіз;
- Багатовимірний аналіз розсіяння, непараметричні тести, множинна регресія;
- Методи оптимального шкалювання та інші.

IBM InfoSphere Warehouse (DB2)

InfoSphere Warehouse, також відоме як DB2, є програмним рішенням, створеним спеціально для потреб бізнес-організацій та відділів середнього розміру. Це програмне забезпечення пропонує широкий спектр можливостей та гнучкості в сферах продуктивності та складного аналізу даних. Продуктивність DB2 досягається за рахунок архітектури обробки даних паралельно на багатьох процесорах (Massively Parallel Processing) і включає в себе розширені можливості, такі як багатовимірна кластеризація та матеріалізовані таблиці запитів, що сприяє швидкій обробці запитів від користувачів [4].

Інструменти IBM InfoSphere Warehouse DB2 охоплюють:

- Масову паралельну обробку даних, що є ключовим фактором для підвищення продуктивності програмного забезпечення. Система ефективно працює на декількох серверах і процесорах одночасно, зменшуючи час, необхідний для обробки операцій;
- Розділення даних на таблиці, що дозволяє ефективно сортувати великі обсяги даних на відокремлені сегменти, сприяючи швидшій обробці;
- Систему компресії даних. Навіть з великими обсягами пам'яті сучасних жорстких дисків, завантаження великих даних може займати чимало часу. InfoSphere Warehouse використовує ефективну технологію стиснення даних, що дозволяє заощаджувати до 80% дискового простору;

- Багатовимірну кластеризацію, яка є ще одним способом покращення продуктивності програмного забезпечення;
- Функції агрегації, що є особливо корисними для завдань, які потребують частих повторень, спрощуючи цей процес.

Функція аналізу даних викликається як процедура, що зберігається, і створює в базі даних модель XML. Цю модель можна використовувати для оцінки нових даних або витягувати інформацію з моделі в таблицю і використовувати такі таблиці в Cognos. Користувач може викликати функцію аналізу даних інтерактивно, звертаючись до відповідної процедури, що зберігається бази даних зі звіту Cognos (рис. 2.0.).

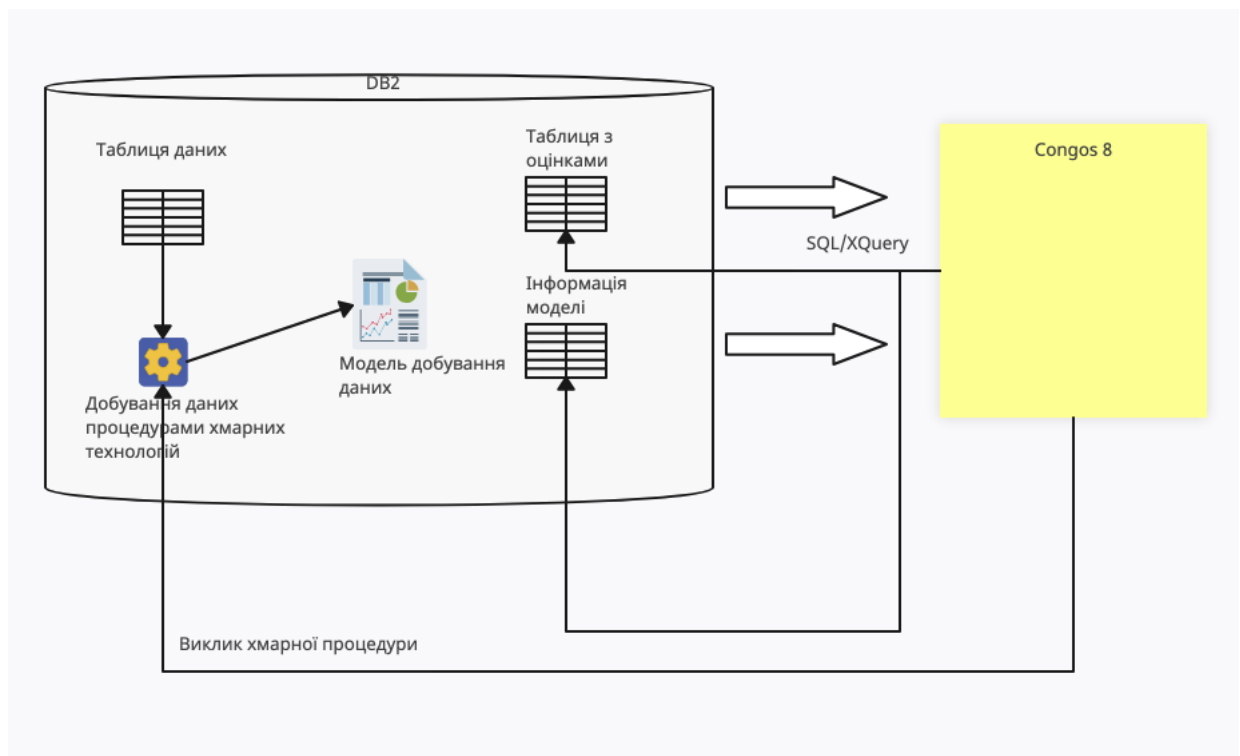


Рисунок 1.7 Функціональна схема DB2.

IBM InfoSphere Warehouse містить ряд ключових аналітичних компонентів, що є важливими для сховищ даних:

- Аналітична обробка в реальному часі, що є стандартною практикою у сфері аналізу даних, дозволяє здійснювати багатовимірний доступ до оброблюваних даних;
- Розширений аналіз даних доповнює стандартні засоби збору даних, включаючи аналіз асоціацій, кластеризацію, класифікацію та прогнозування;
- Швидка і ефективна інтеграція з численними джерелами даних.

IBM InfoSphere Warehouse вирізняється гнучкістю і легкістю масштабування:

- Терабайтне ціноутворення сприяє зниженню вартості за умови значного стиснення даних;
- Віртуальний образ DB2 спрощує і прискорює процес розгортання;
- Ефективний захист даних забезпечує безпеку, наприклад, шляхом надання різних рівнів доступу для різних типів користувачів;
- Автономне управління, що підвищує продуктивність і зменшує витрати;
- Висока доступність, пропускна здатність і неперервність роботи;
- Інтеграція різноманітного вмісту і джерел даних за допомогою технології InfoSphere Federation Server;
- Підтримка XML-файлів з можливістю зберігання, обробки та управління XML-даними;
- SQL-сумісність, яка уникла проблем із впровадженням функцій Oracle і Sybase у сховищах даних;
- Графічний інтерфейс користувача для проектування та моделювання;
- Підтримка управління базами даних з використанням адміністратора баз даних Optim - інструменту, що допомагає мінімізувати ризик помилок при впровадженні складних змін у бази даних.

1.6 Постановка задачі дослідження

Реляційна база даних, яка застосовується в корпоративній інформаційній системі, вимагає значного часу для обробки запитів, перетворюючи зібрані

підприємством дані в аналітичну інформацію. Через те, що початкові дані розподілені між численними таблицями, їх агрегація вимагає виконання складних операцій об'єднання, що споживає значну обчислювальну потужність та може призводити до перевантаження бухгалтерської системи компанії. Слід провести інтелектуальний аналіз даних, тому що він дасть значні переваги для підприємства, яке спеціалізується на виробі з каменю, оскільки цей метод дозволяє глибше зрозуміти бізнес-процеси, збільшити ефективність, виявити нові можливості та вдосконалити прийняття рішень.

Завдяки сучасним методам інтелектуального аналізу даних, можна забезпечити корпоративні ресурси зручними та гнучкими інструментами. Ці інструменти допомагають систематизувати накопичені дані та своєчасно надавати їх у формі, яка є необхідною для ухвалення управлінських рішень.

Для аналізу даних в даний час широко застосовують методи та засоби штучного інтелекту, зокрема ІАД, нейронні мережі, нечіткі моделі, дерева рішень, пошук асоціацій, методи регресійного аналізу та ін.

Інтелектуальний аналіз даних – одна з найбільш складних і в той же час найбільш актуальний вид аналітичної роботи з даними. Складність процесу ІАД пов'язана з необхідністю аналізу і оцінювання великих обсягів даних, ускладненням методів і появою нових концепцій та підходів до аналізу. Тому, на сьогоднішній день, стан розвитку методів аналізу тісно пов'язаний з розвитком інформаційних технологій.

У підприємстві "Світ каменю", яка виробляє продукції таких видів каменю: граніт, мрамур, онікс, травертин та штучний кварц. Для ефективного аналізу даних необхідно враховувати та використовувати техніки інтелектуального аналізу даних під час роботи з вітриною даних в якій знаходиться статистичні дані про збут за 2023 рік.

Висновок до розділу 1

У цьому розділі проаналізовано діяльність ТОВ "Світ каменю". На основі аналізу автоматизації бізнес-процесів підприємства, запропоновано створення інформаційної системи для інформаційної підтримки формування рекомендацій щодо планування асортименту продукції, що дозволить покращити процес прийняття управлінських рішень.

З метою більш раціонального планування виробництва продукції ТОВ "Світ каменю" рекомендується використати методи інтелектуального аналізу даних.

Використання методів ІАД надає можливість:

- проведення аналізу статистичних даних підприємства;
- вибору моделі для планування асортименту на наступні періоди, що буде основою для планування виробничої програми;
- побудова моделі кластерного аналізу для реалізації продукції певного виду на майбутній період.

РОЗДІЛ 2. ОГЛЯД МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

2.1. Поняття інформаційно-аналітичної діяльності

Інформаційно-аналітичні системи визначаються як особливий клас інформаційних систем, призначених для аналітичної обробки даних, а не для автоматизації повсякденної діяльності організації. Інформаційно-аналітичні системи об'єднують, аналізують і зберігають як єдине ціле інформацію, яка вилучена як з баз даних організації, так і із зовнішніх джерел. Сховища даних, що входять до складу інформаційно-аналітичних систем, забезпечують перетворення великих об'ємів деталізованих даних в узагальнену вивірену інформацію, яка придатна для прийняття обґрунтованих рішень. На відміну від звичайних баз даних сховища містять оброблене, впорядковане і зрозуміле керівникам представлення даних.; вони стають складальним конвеєром по підготовці інформації в інтегрованому, несуперечливому, наочному вигляді для підтримки прийняття управлінських рішень.

Задачами ІАС є ефективно зберігання, обробка та аналіз даних. Технологічна платформа ІАС дозволяє підприємству (організації) здійснювати інтеграцію та координацію його бізнес-процесів.

Типова ІАС встановлюється в одному або кількох дата центрах, в яких виконується відповідне програмне забезпечення даного підприємства (організації). ІАС може включати в себе додаткові застосунки, що працюють в рамках організаційної структури підприємства (організації) (наприклад, для цілей управління).

ІАС забезпечує для підприємства єдиний інформаційний простір і гарантує, що ця інформація буде доступна на всіх функціональних рівнях ієрархії та управління.

Інформаційно-аналітичні системи будуються на основі оперативних даних, одержуваних у режимі реального часу з оперативних систем, що автоматизують основні види діяльності організації, а також інших доступних джерел даних, які можуть знадобитися при прийнятті стратегічних рішень. Базовий комплекс інформаційно-аналітичних систем торкається всієї управлінської вертикалі: корпоративної звітності, аналіз бізнесу, фінансово-економічне й стратегічне планування.

Інформаційно-аналітичні системи є надбудовою над уже функціонуючими на підприємстві інформаційними додатками, не роблячи особливого впливу на їхнє функціонування й не вимагаючи їхньої заміни. Ключовою функцією цих систем є акумулювання даних по всіх видах діяльності компанії - від стану складів до фінансової й бухгалтерської звітності й усіляких відомостей про клієнтів і їх смаків.

За способами застосування інформаційно-аналітичні системи діляться на системи для рішення тактичних і стратегічних завдань

Інформаційно-аналітичні системи підрозділів припускають більшу деталізацію й більш складну аналітичну обробку. Ці системи допомагають підготувати

інформацію для прийняття рішень в області збуту, продуктової пропозиції, фінансового планування і т. д.

Інформаційно-аналітичні системи верхнього рівня призначені для керівників верхньої ланки й служать для прийняття стратегічних рішень

До типових завдань, розв'язуваних за допомогою інформаційно-аналітичних систем, відносяться:

- Складання консолідованої звітності й надання зведеної інформації про діяльність підприємства.
- Аналіз діяльності дочірніх підприємств, філій і підрозділів компанії.
- Аналіз фінансової діяльності, оптимізація фінансових потоків, реальна оцінка собівартості продукції;
- Проведення комплексної оцінки діяльності підприємства, заснованої на постійному контролі найбільш істотних її аспектів: фінансів, відносин з партнерами й клієнтами, внутрішнього стану компанії, темпів розвитку;
- Аналіз збутових процесів (маркетингові компанії, складання плану, контроль виконання розпоряджень, розрахунки за відвантаженою продукцією, прогноз надходження засобів, планування асортименту).

Переваги, одержувані підприємством після впровадження корпоративної інформаційно-аналітичної системи:

- Наявність у керівників ефективних інструментів оцінки стану бізнесу на основі єдиного джерела ключових показників, підтримуваного в режимі реального часу;
- Можливість оцінки перспектив розвитку;
- Підвищення обґрунтованості прийняття управлінських рішень;
- Можливість узгодження оперативних планів і бюджетів зі стратегічними цілями;

- Розширені можливості аналітики, надані сховищем даних, засобами багатомірного аналізу даних, прогнозування й моделювання різних ситуацій.
- Розширені можливості стратегічного керування на основі потужних засобів аналізу й звітності.
- Виключення проблем, пов'язаних з оцінкою ситуації на основі показників, одержуваних на основі не інтегрованих рішень.

2.2 Огляд існуючих методів інтелектуального аналізу

У сучасному світі рішення складних завдань та ефективне управління інформацією вимагає широкого застосування різноманітних методів інтелектуального аналізу даних (Data Mining). Ці методи дозволяють виявляти значущі шаблони та відкриття в великих обсягах даних, що є критично важливим для розробки стратегій, оптимізації процесів та прийняття обґрунтованих рішень у бізнесі та науці.

Інтелектуальний аналіз даних (Data Mining) – це комплексний процес аналізу великих обсягів даних для виявлення прихованих закономірностей, зв'язків, аномалій та інших значимих шаблонів, які не можуть бути виявлені при традиційному підході до аналізу даних. Цей процес включає в себе використання статистичних методів, штучного інтелекту та машинного навчання для вивчення та інтерпретації даних з метою отримання корисної інформації.

Основною метою інтелектуального аналізу даних є перетворення сирих даних у зрозумілу форму, роблячи їх корисними для прийняття обґрунтованих рішень. Процес включає декілька етапів: підготовку даних, вибір методу аналізу, оцінку отриманих результатів і їх інтерпретацію. Підготовка даних може включати очищення, інтеграцію, вибірку та трансформацію даних. Після підготовки, наступний крок – вибір відповідного алгоритму або техніки для аналізу, що може бути класифікацією, кластеризацією, прогнозуванням, виявленням аномалій тощо [7].

- **Прогнозування ризиків:** Моделі інтелектуального аналізу даних дозволяють компаніям оцінювати й управляти різними видами ризиків, які можуть виникати в їхній діяльності. Це може включати фінансові ризики, ризики пов'язані з кредитуванням, а також оперативні ризики.
- **Оптимізація ланцюга постачання:** Інтелектуальний аналіз даних може допомогти покращити ефективність ланцюга постачання, аналізуючи попит, прогножуючи запаси та оптимізуючи логістичні процеси.
- **Оптимізація виробничих процесів:** Використання моделей ІАД може підвищити продуктивність виробництва шляхом виявлення та усунення неефективностей в процесах, виявлення та передбачення можливих збоїв обладнання.
- **Ризик і ймовірність:** вибір найбільш придатних замовників для цільової розсилки, визначення точки рівноваги для ризикованих сценаріїв, призначення ймовірностей діагнозами або інших результатів;
- **Планування асортименту продукції:** Інтелектуальний аналіз даних може значно поліпшити процес планування асортименту продукції. Використовуючи аналіз тенденцій попиту, історичних продажів, поведінки покупця та ринкових умов, компанії можуть точніше прогнозувати, які продукти будуть затребувані. Це дозволяє ефективно управляти запасами, мінімізувати витрати на зберігання та знижувати ризик надлишків чи дефіциту товарів.

Величезні обсяги даних спричинили стрімке зростання інтересу до розширених методів інтелектуального аналізу даних, оскільки кількість інформації зросла і стала значно різноманітнішою та складнішою. Простий статистичний аналіз вже не є достатнім при роботі з великими наборами даних. Наприклад, при аналізі 30 чи 40 мільйонів деталізованих записів про покупки, недостатньо лише знати, що два мільйони з них були здійснені в одному місці. Важливо також розуміти, чи ці два мільйони покупок належать до певної вікової групи і який у них середній дохід [7].

Такі бізнес-потреби вимагають переходу від простого пошуку та аналізу даних до більш складного інтелектуального аналізу. Для вирішення бізнес-задач потрібен аналіз, який дозволяє розробити модель для детального опису інформації, що в кінцевому рахунку призводить до створення змістовних звітів та інсайдів. Процес аналізу даних зображено на рисунку 2.1.

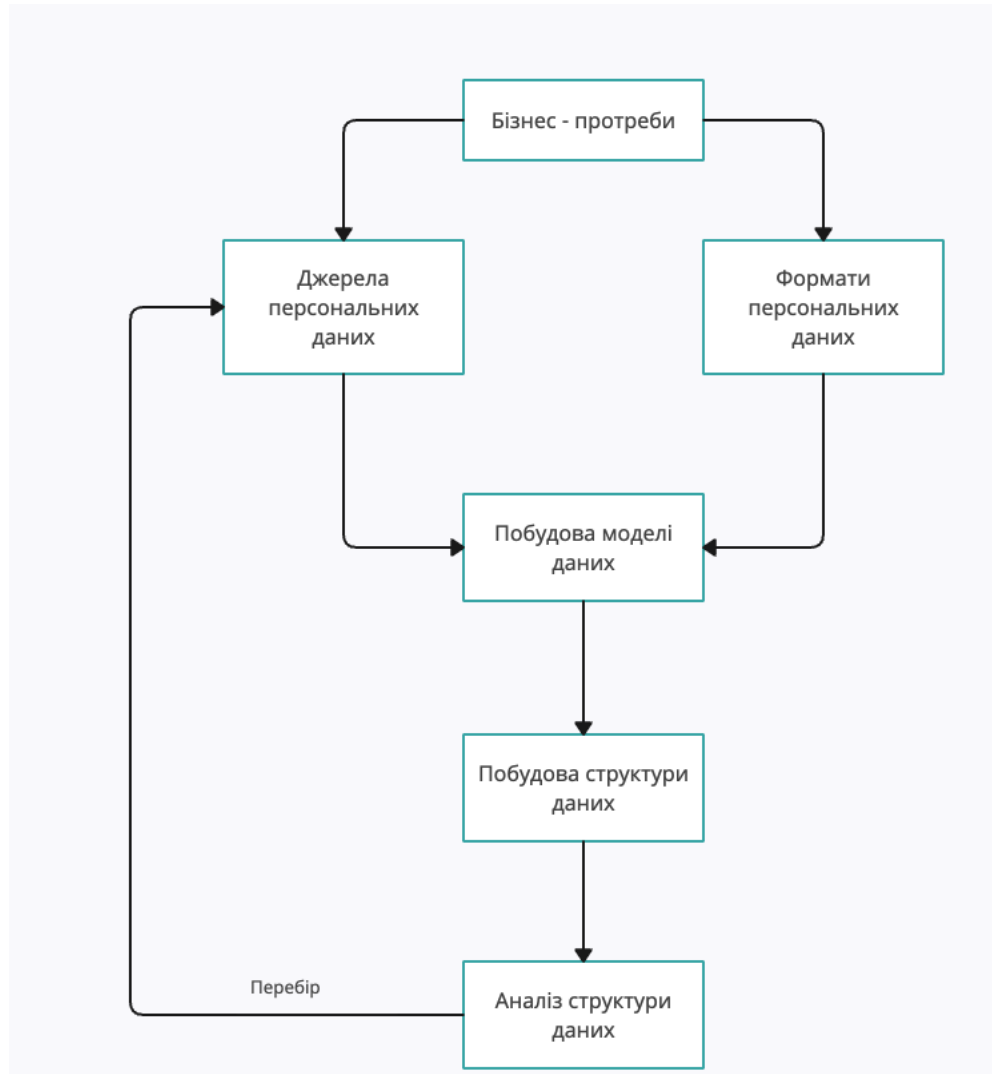


Рисунок 2.1 Блок-схема аналізу даних.

Процес аналізу даних, включаючи виявлення, пошук та розробку моделей, часто відбувається у вигляді ітерацій. Це пов'язано з необхідністю виявлення повторюваних закономірностей та прихованих паттернів, які зазвичай містяться в об'ємах великих даних. Важливим аспектом є також здатність з'єднувати, перетворювати та інтегрувати ці дані з іншими інформаційними ресурсами для досягнення бажаних результатів. В рамках цього процесу, після ідентифікації нових елементів і характеристик даних, може виникати необхідність перегляду підходів

до використання джерел і форматів даних. Це, в свою чергу, може вимагати адаптації та модифікації існуючих методів аналізу, з метою забезпечення глибшого і більш точного зіставлення зібраної інформації з встановленими бізнес-цілями і критеріями оцінки [8].

2.2.1 Кластерний аналіз

Аналізуючи один або декілька характеристик чи категорій, можливо здійснити групування окремих частин даних, тим самим формуючи структуроване узагальнення. У базовому аспекті кластеризація включає використання одного або більше характеристик як фундаменту для групування схожих результатів в межах одного кластера. Цей метод є ефективним у виявленні різноманітності інформації, оскільки він дозволяє виявляти зв'язки з іншими даними, виявляючи схожості та межі співпадіння між різними наборами даних.

Кластерний аналіз, відомий також як кластеризація, є методом статистичного аналізу, який використовується для групування об'єктів або точок даних у відносно однорідні групи, звані кластерами. Цей метод базується на ідеї знаходження підмножин у великому наборі даних, де елементи в межах кожного кластера є більш схожими один на одного, ніж на елементи в інших кластерах. Відмінності та схожості між даними визначаються на основі заданих характеристик або атрибутів, таких як відстань, щільність, або властивості. Кластерний аналіз застосовується у різних областях, таких як маркетинг для сегментації ринку, у біології для класифікації рослин і тварин, у медицині для ідентифікації груп пацієнтів з подібними симптомами або характеристиками. Цей метод дозволяє визначати природні групування в даних і використовувати ці групування для різноманітних аналітичних або практичних застосувань.

Кластеризація відіграє ключову роль у сфері інтелектуального аналізу даних і є одним із головних статистичних методів, що застосовується у широкому спектрі дисциплін. Цей процес полягає у групуванні об'єктів у такий спосіб, що елементи в межах одного кластера демонструють високу ступінь схожості, в той час як кластери між собою відрізняються. Це означає, що об'єкти всередині кожного

кластера розташовані якомога ближче один до одного, враховуючи всі доступні атрибути, тоді як різні кластери відділені один від одного на максимально можливу відстань. У порівнянні з кластеризацією, класифікація зосереджена на розділенні DataSet на взаємовиключні групи таким чином, що елементи в кожній групі максимально подібні один до одного, а групи між собою максимально різні з огляду на певні визначені змінні, які ви прагнете передбачити.

Для ТОВ "Світ каменю" це може мати велике значення, особливо якщо в їхньому асортименті є різноманітні товари, такі як будівельні матеріали, декоративні елементи тощо. Оптимізація асортименту дозволяє легше визначати стратегії маркетингу, контролювати залишки товарів та надавати клієнтам більш зрозумілу та структуровану пропозицію.

Створення моделі кластеризації у інструменті для аналізу даних передбачає розробку джерела даних та структурування процесу інтелектуального аналізу цих даних. Кластерний аналіз відіграє ключову роль у вивченні та аналізі даних, особливо коли існують численні інстанції даних без явної агрегації. В таких випадках алгоритми кластеризації допомагають у виявленні природних групувань. Кластеризація також може виступати в ролі важливого етапу попередньої обробки даних для виділення однорідних груп, що будуть використані у подальшому моделюванні. Існує широкий спектр алгоритмів для проведення кластерного аналізу, серед яких особливо популярні є кластеризація методом k-середніх (K-mean algorithm), просторова кластеризація на основі щільності для додатків із шумами (DBSCAN) та самоорганізаційна карта Кохонена (SOM) [9].

Кластеризація є процесом прийняття рішень. Використання різних методів кластеризації на одному й тому ж наборі даних часто призводить до отримання різних груп. Це вимагає проведення оцінки та верифікації дійсності кластерів, сформованих за різними методами, щоб забезпечити точність та об'єктивність отриманих результатів.

Багато традиційних методів кластеризації ґрунтуються на принципі вимірювання відстаней між об'єктами, що обмежує їх здатність до виявлення лише

скупчень кулястої форми, з труднощами у розпізнаванні кластерів неправильних форм. На відміну від них, розробка DBSCAN базується на концепції щільності. Центральна ідея цього підходу полягає в тому, що для кожної точки в межах кластера існує визначений радіус, всередині якого повинна бути зафіксована мінімально необхідна кількість інших точок, таким чином, щільність у цьому радіусі має перевищувати заданий пороговий рівень. Це дозволяє DBSCAN ефективно виявляти кластери різноманітних форм та розмірів.

Визначення ключових параметрів:

eps (epsilon): Максимальна відстань між двома точками, за якої вони вважаються сусідніми.

MinPts (мінімальна кількість точок): Мінімальна кількість точок, яка потрібна для того, щоб точка вважалася центральною (або ядерною) в кластері.

Класифікація точок:

Ядерні точки: Точка є ядерною, якщо в межах відстані eps від неї знаходиться не менше MinPts точок.

Граничні точки: Точка є граничною, якщо вона не є ядерною, але в межах відстані eps від неї є хоча б одна ядерна точка.

Шум: Точки, які не є ні ядерними, ні граничними.

Формування кластерів:

Ядерні точки, разом із точками, які знаходяться в межах eps від них, формують кластер.

Граничні точки включаються до найближчого кластера, але не можуть бути використані для розширення кластера.

Розширення кластерів:

Алгоритм продовжує обробляти всі ядерні точки, розширюючи кластери, поки не будуть оброблені всі точки, що можуть бути включені в кластери.

Обробка шуму:

Точки, які не входять до будь-якого кластера, класифікуються як шум. Ці точки або залишаються некласифікованими, або можуть бути оброблені за допомогою додаткових методів.

Основні точки (Core points) – всі точки, які щонайменше мають *MinPts* точок в своїх ϵ -околицях;

Крайові точки (Border points) – всі точки, які мають менше *MinPts* точок в своїх ϵ -околицях, але знаходяться досить близько до деяких основних точок;

Точки шуму (Noise point) – будь-яка точка, яка не є основною або крайовою точкою.

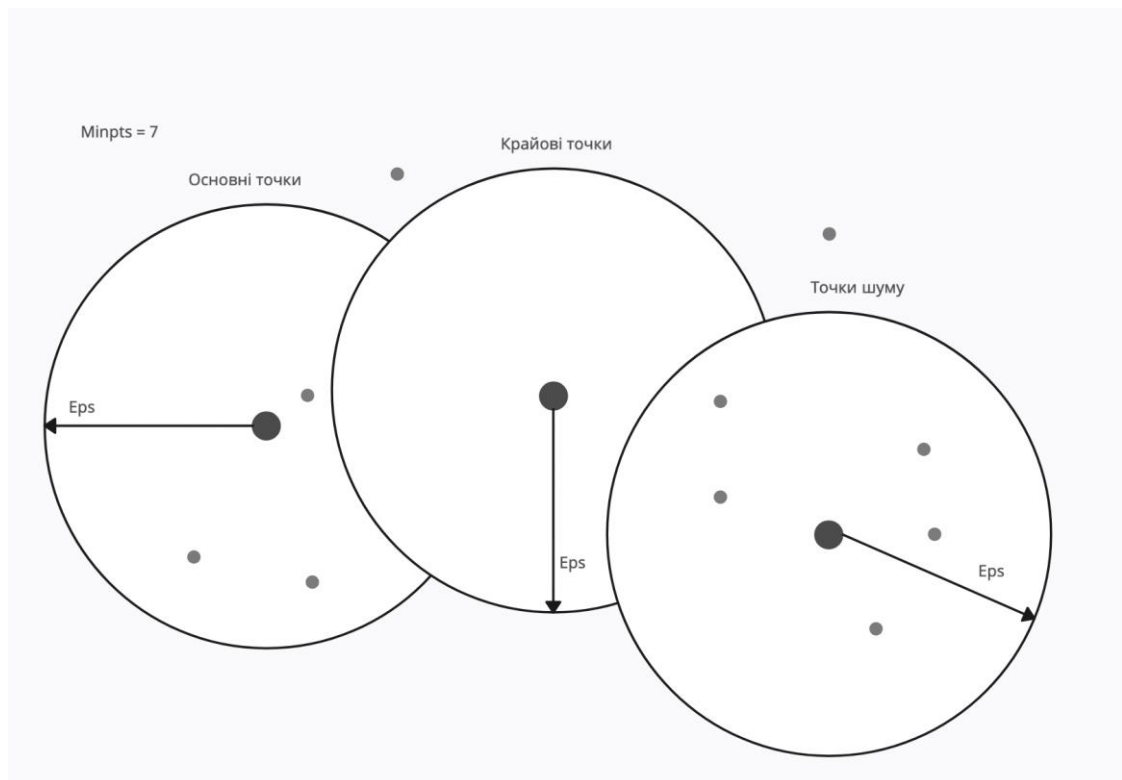


Рисунок 2.2. DBSCAN. Основні, крайові та точки шуму.

Двома глобальними параметрами для алгоритмів DBSCAN є:

1. ϵ (*Eps*) – Максимальний радіус околиці;
2. *MinPts* – Мінімальна кількість точок у ϵ -околиці цієї точки.

Околиця ϵ від точки p , позначеної $N_{eps}(p)$, є

$$Neps(p) = \{q \in D \mid dist(p, q) \leq \varepsilon\}. \quad (2.1)$$

Скажемо, що точка p *безпосередньо досягає щільності з точки q* , відносно (*wrt*) деякої ε та $MinPts$, якщо

$$p \in Neps(q) \text{ і } |Neps(q)| \leq MinPts, \quad (2.2)$$

де $|Neps(q)|$ – кількість елементів $Neps(q)$.

Точка p є *щільно-досяжною для точки q* , якщо існує послідовність точок $p_1, p_2, \dots, p_n, p_1 = q, p_n = q$, така, що p_{i+1} безпосередньо досягає щільності з p_i .

Точка p є *щільно-зв'язною з точкою q* , якщо існує так точка o , що обидві, p і q є щільно-досяжними з o .

Алгоритм DBSCAN (Density-Based Spatial Clustering of Applications with Noise) є надійним методом кластеризації, оскільки він ефективно виявляє кластери різної форми та розмірів, відмінно справляється із шумом у даних і не вимагає попереднього визначення кількості кластерів. Центральними елементами алгоритму є два параметри: ε (максимальна відстань для визначення сусідніх точок) та $MinPts$ (мінімальна кількість точок для формування ядра кластера). Це дозволяє DBSCAN адаптуватися до різних типів даних, роблячи його гнучким у виявленні кластерів, особливо в нелінійних структурах. Така гнучкість, поряд із здатністю ігнорувати випадкові шуми та визначати точки, що не належать до жодного кластеру, робить DBSCAN особливо надійним у широкому спектрі задач аналізу даних [9].

2.3 Алгоритм k-means

Алгоритм k-середніх, широко відомий у світі ітеративних методів кластеризації та часто називаний методом швидкого кластерного аналізу, відрізняється від ієрархічних методів своєю залежністю від попередніх припущень щодо кількості кластерів. Важливою умовою для його застосування є наявність ідеї або гіпотези про оптимальну кількість кластерів. Цей алгоритм розробляє k кластерів, які розташовані на значних відстанях один від одного, з орієнтацією на

сценарії, де очікувана кількість кластерів є заздалегідь відома і відрізняється від інших. Визначення цієї кількості (k) може бути засноване на результатах попередніх досліджень, теоретичних обґрунтуваннях, або навіть на інтуїтивному розумінні [10].

Методика працює на основі мінімізації загальної суми квадратів відстаней між кожним елементом даних та центром призначеного йому кластера.

$$\sum_{i=1}^k \sum_{x \in D_i} \|x - c_i\|^2 \quad (2.3)$$

де x - набір векторів що належать до i -го кластеру, c_i - середнє значення цих векторів.

$$c_i = \frac{\sum_{x \in D_i} x}{N_i}, x \in D_i \quad (2.4)$$

Ключовий принцип полягає в тому, що на кожному етапі алгоритму відбувається перерахунок центру ваги для кожного кластера. Після цього, вектори даних перекласифікуються у нові групи засновано на близькості до цих новообчислених центрів, використовуючи відповідну метрику відстані. Приклад кластеризації k -means зображено на рисунку 2.4.

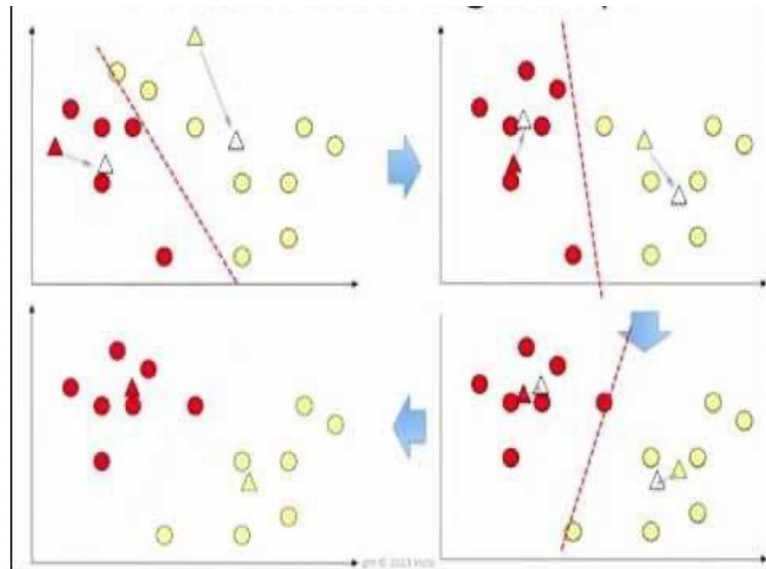


Рисунок 2.3. Приклад кластеризації k-means

Переваги алгоритму k-середніх:

- Легкість у застосуванні.
- Висока швидкість обробки даних.
- Простота, прозорість та зрозумілість алгоритму.

Недоліки алгоритму k-середніх:

- Висока чутливість до аномалій (викидів), які можуть спотворювати середнє. Можливим рішенням є застосування модифікації алгоритму, наприклад, k-медіан.
- Потенційно повільна робота на великих наборах даних. Цю проблему можна вирішити за допомогою вибірки даних.

2.4 Огляд ПП «Orange»

Orange є відкритим програмним забезпеченням, призначеним для інтелектуального аналізу даних та машинного навчання. Розроблений університетом в Люблянці, Словенія, цей інструмент вирізняється своїм інтуїтивно зрозумілим графічним інтерфейсом, який полегшує використання навіть для користувачів без глибоких знань у програмуванні. Orange дає можливість

проводити різноманітні аналітичні операції - від простого візуального аналізу даних до складних процедур машинного навчання [9].

Функціонал Orange охоплює широкий спектр аналітичних інструментів. Користувачі можуть виконувати кластерний аналіз, класифікацію, регресію, а також використовувати різні алгоритми машинного навчання, включаючи нейронні мережі, дерева рішень та випадкові ліси. Важливою особливістю Orange є можливість візуалізації даних, що допомагає користувачам краще розуміти дані та результати їхнього аналізу. Ця функція особливо корисна для виявлення закономірностей та аномалій у великих наборах даних. Розглянемо кілька прикладів:

- **Файл і таблиця даних:** Основні одиниці аналізу даних в Orange називаються віджетами. У цьому робочому процесі віджет «Файл» зчитує дані. Віджет «Файл» передає ці дані у віджет «Таблиця даних», який відображає дані в електронній таблиці. Вихід файлу з'єднаний із входом таблиці даних.
- **Візуалізація підмножин даних:** Деякі віджети візуалізації, як точкова діаграма та кілька віджетів проєкції даних, можуть показувати екземпляри даних у підмножині даних. У цьому робочому процесі точкова діаграма візуалізує дані з файлу вхідних даних, а також позначає точки даних, вибрані в таблиці даних (вибрані рядки).
- **Зведена таблиця:** Зведена таблиця може допомогти нам агрегувати та трансформувати дані. Цей робочий процес використовує проекти Kickstarter і агрегує їх за місяцями. Ми можемо перевіряти частоту опублікованих проєктів на місяць і спостерігати різницю між фінансованими та нефінансованими проєктами.
- **Інтерактивні візуалізації:** Більшість візуалізацій в Orange є інтерактивними. Наприклад, діаграма розсіювання. Двічі клацніть його значок, щоб відкрити його, і клацніть і перетягніть, щоб вибрати кілька точок даних із графіка. Вибрані дані автоматично

поширюватимуться в таблицю даних. Двічі клацніть його, щоб перевірити, які дані вибрано.

Ієрархічна кластеризація це Робочий процес кластеризує елементи даних у наборі даних райдужної оболонки, спочатку перевіряючи відстані між екземплярами даних. Матриця відстані передається до ієрархічної кластеризації, яка відображає дендрограму (рисунок 2.5).

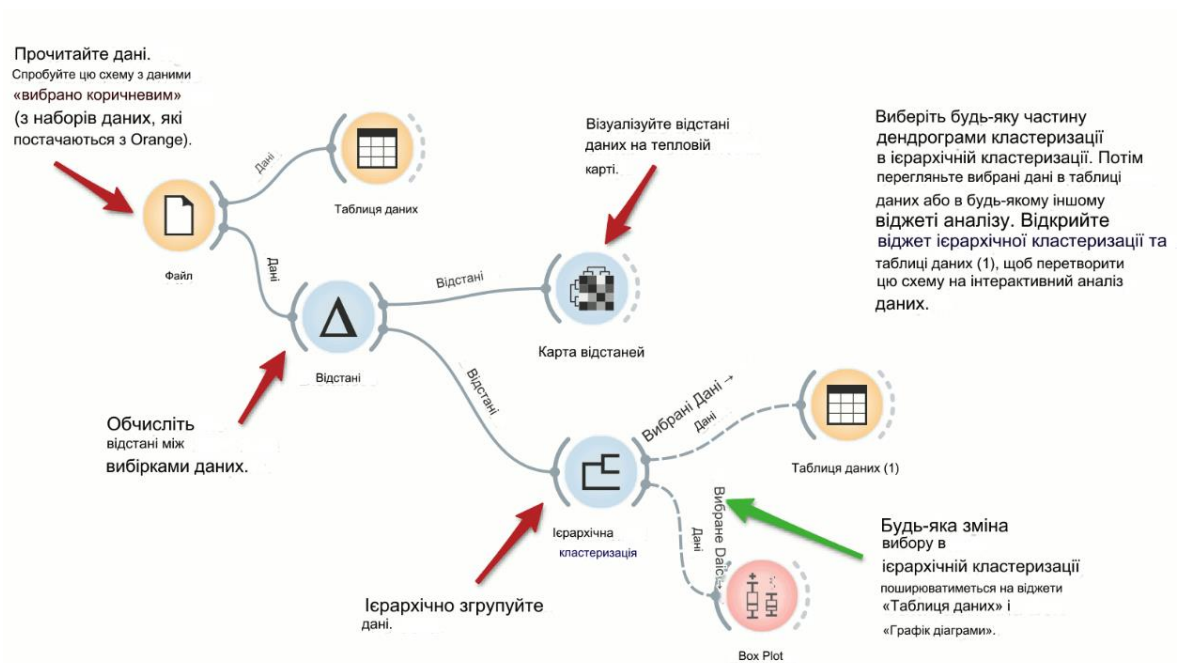


Рисунок 2.4 Ієрархічна кластеризація в «Orange»

Однією з ключових переваг Orange є його легкість у використанні. Інструмент пропонує "перетягування" віджетів, що дозволяє користувачам легко створювати та змінювати аналітичні потоки даних. Ця візуальна програмна платформа спрощує складні процеси аналізу даних, роблячи їх доступними навіть для не-спеціалістів. Також, Orange підтримує розширення за допомогою додатків, що дає користувачам можливість доповнювати його стандартний функціонал згідно із специфічними потребами їхніх проектів.

Orange широко використовується в академічних та дослідницьких цілях, а також знаходить застосування в комерційному секторі, особливо у сферах, де потрібен швидкий та ефективний аналіз даних. Існує активна спільнота користувачів і розробників, що неперервно працює над удосконаленням інструменту, розширюючи його можливості та надаючи підтримку новим користувачам. Ця спільнота допомагає у вирішенні проблем, обміні досвідом та розробці нових віджетів та функціоналу, забезпечуючи постійний розвиток та актуалізацію інструменту.

Однією з основних переваг Orange є його здатність до адаптації та інновацій. Постійні оновлення та розширення функціоналу забезпечують, що інструмент залишається в ногу з останніми тенденціями в галузі аналізу даних та машинного навчання. Це не тільки збільшує його привабливість для дослідників і аналітиків, але й дозволяє охопити більш широкий спектр задач, від простого аналізу даних до складних мультидисциплінарних проектів. Такий підхід забезпечує гнучкість та розширює можливості застосування Orange у різноманітних сферах, включаючи науку, освіту, бізнес та урядові структури.

В майбутньому можна очікувати, що Orange продовжить свій розвиток як потужний інструмент для аналізу даних, зберігаючи свою популярність серед користувачів різних рівнів кваліфікації. Зростання обсягів даних та їх складності вимагають більш інноваційних та інтуїтивно зрозумілих інструментів, і Orange має потенціал задовольнити ці потреби. Його відкритий характер і активна спільнота користувачів створюють оптимальні умови для подальших інновацій та розширення сфери застосування, що робить Orange одним з ключових інструментів у світі аналізу даних на багато років вперед.

2.5 Загальні методи

Різнманітні підходи, які застосовуються в області аналітики даних, зазвичай поділяються за типом аналізу та процесом відновлення даних. Однак, варто зазначити, що не існує уніфікованої термінології серед різних організацій і рішень,

що може призводити до непорозумінь та здаватися зайво складним. Проаналізуємо декілька основних методик та способи їх застосування у сфері аналітики даних.

Метод асоціацій, ймовірно, один з найвідоміших і найпростіших підходів у сфері аналітики даних. Він базується на виявленні зв'язків між двома або більше елементами, часто одного типу. Наприклад, аналізуючи покупки, можна виявити, що полуниця часто купується разом з вершками.

Створення інструментів аналітики даних на основі асоціацій є досить простим завданням. Взяти хоча б InfoSphere Warehouse, де існує функція майстра, яка дозволяє налаштовувати інформаційні потоки для аналізу асоціацій, досліджуючи вхідні дані, основу для прийняття рішень та результати.

Класифікація даних може бути використана для отримання уявлення про типи покупців, товарів або об'єктів на основі різних характеристик та поведінкових патернів. Ось як це може працювати:

Класифікація покупців:

- Демографічні дані: Класифікація може бути заснована на віці, статі, освіті, доходах та інших демографічних ознаках.
- Поведінкові патерни: Аналіз поведінки покупців, як-от частота покупок, вибір товарів, історія покупок, для ідентифікації груп з подібними звичками.
- Психографічні характеристики: Класифікація за цінностями, інтересами, ставленнями, що допомагає розуміти мотивацію покупця.

Класифікація товарів:

- Категоризація за типом: Розділення товарів на категорії.
- За рівнем популярності: Класифікація товарів за ступенем популярності серед покупців.
- Цінові сегменти: Розділення товарів на діапазони цін для визначення цінової категорії.

Класифікація об'єктів:

- За географічним розташуванням: Класифікація магазинів або філій компанії залежно від їх місцезнаходження.
- Функціональність: Класифікація об'єктів за їх функціональним призначенням.
- За розміром або потужністю: Розділення об'єктів за їх фізичними характеристиками або потужністю.

Також, результати класифікації можуть слугувати основою для подальших аналітичних методів. Наприклад, дерева рішень часто використовуються для визначення параметрів класифікації. Подальше застосування кластеризації може допомогти у виявленні груп або кластерів, об'єднаних спільними атрибутами різних класифікаційних категорій.

2.6 Огляд методів інтелектуального аналізу даних

Інструменти інтелектуального аналізу даних є важливою частиною сучасного процесу обробки та аналізу великих обсягів інформації. Вони дозволяють організаціям перетворювати великі і складні набори даних на зрозумілі та корисні інсайти. Ці інструменти включають різноманітні технології, такі як машинне навчання, статистичний аналіз, шаблони пошуку та текстовий аналіз. Використання таких систем дозволяє автоматизувати збір, обробку та візуалізацію даних, роблячи інформацію доступною для керівників та аналітиків для прийняття рішень.

Інструменти аналізу даних можна класифікувати за кількома категоріями залежно від їх функціональності. До найбільш поширених відносяться:

- Дерева рішень та ансамблі: Використовуються для класифікації та прогнозування, допомагають виявляти складні структури в даних.
- Кластерний аналіз: Дозволяє виявити групи схожих об'єктів в наборі даних, корисний для сегментації ринку або виявлення патернів.

- Аналіз асоціацій: Виявляє зв'язки та закономірності між різними елементами в наборах даних.
- Текстовий аналіз: Використовується для обробки та аналізу текстових даних, наприклад, для сентимент-аналізу або тематичного моделювання.

Використання інструментів інтелектуального аналізу даних має широкий спектр застосувань у різних галузях. У бізнесі вони можуть допомогти виявити тенденції продажів, покращити цільовий маркетинг, аналізувати поведінку споживачів, а також сприяти ефективному управлінню ресурсами. У фінансовій сфері вони використовуються для виявлення шахрайства та кредитного скорингу. У науці про здоров'я інструменти аналізу даних можуть бути використані для досліджень ефективності лікування, прогнозування епідемій та аналізу медичних записів. Таким чином, інструменти інтелектуального аналізу даних відіграють ключову роль у перетворенні сирих даних на цінну інформацію, що може бути використана для рішення різноманітних завдань.

2.6.1 Прогнозування

Прогнозування в аналізі даних – це процес використання історичних даних для передбачення майбутніх подій, тенденцій або поведінки. Це ключовий компонент у багатьох галузях, від фінансів до маркетингу та оперативного планування, де інсайти, отримані з даних, можуть допомогти в прийнятті обґрунтованих рішень. За допомогою статистичних методів, машинного навчання та алгоритмів аналізу часових рядів, аналітики можуть ідентифікувати закономірності та зв'язки в даних, що дає можливість зробити передбачення про майбутнє на основі минулих та сучасних даних.

Прогнозування включає в себе різні методи, від простих статистичних технік до складних моделей машинного навчання. Лінійна регресія, авторегресійні інтегровані моделі ковзання середніх (ARIMA), та нейронні мережі є декількома прикладами. Вибір методу залежить від специфіки даних та цілей аналізу. Наприклад, для короткострокових прогнозів можуть використовуватися більш

прості моделі, тоді як для складних, нелінійних взаємозв'язків у даних застосовуються більш складні техніки, такі як машинне навчання.

Блок-схема яка зображена на рисунку 2.5 аналітичної обробки даних представляє процес аналітичної обробки даних, який використовується для перетворення сировини даних у корисні прогнози та інсайти, які можуть сприяти прийняттю обґрунтованих рішень. Блок-схеми є надати візуальне зображення послідовності кроків, які необхідно виконати для перетворення сировини даних із бази даних у інформацію, яка може бути використана для прогнозування майбутніх тенденцій чи моделей [9].

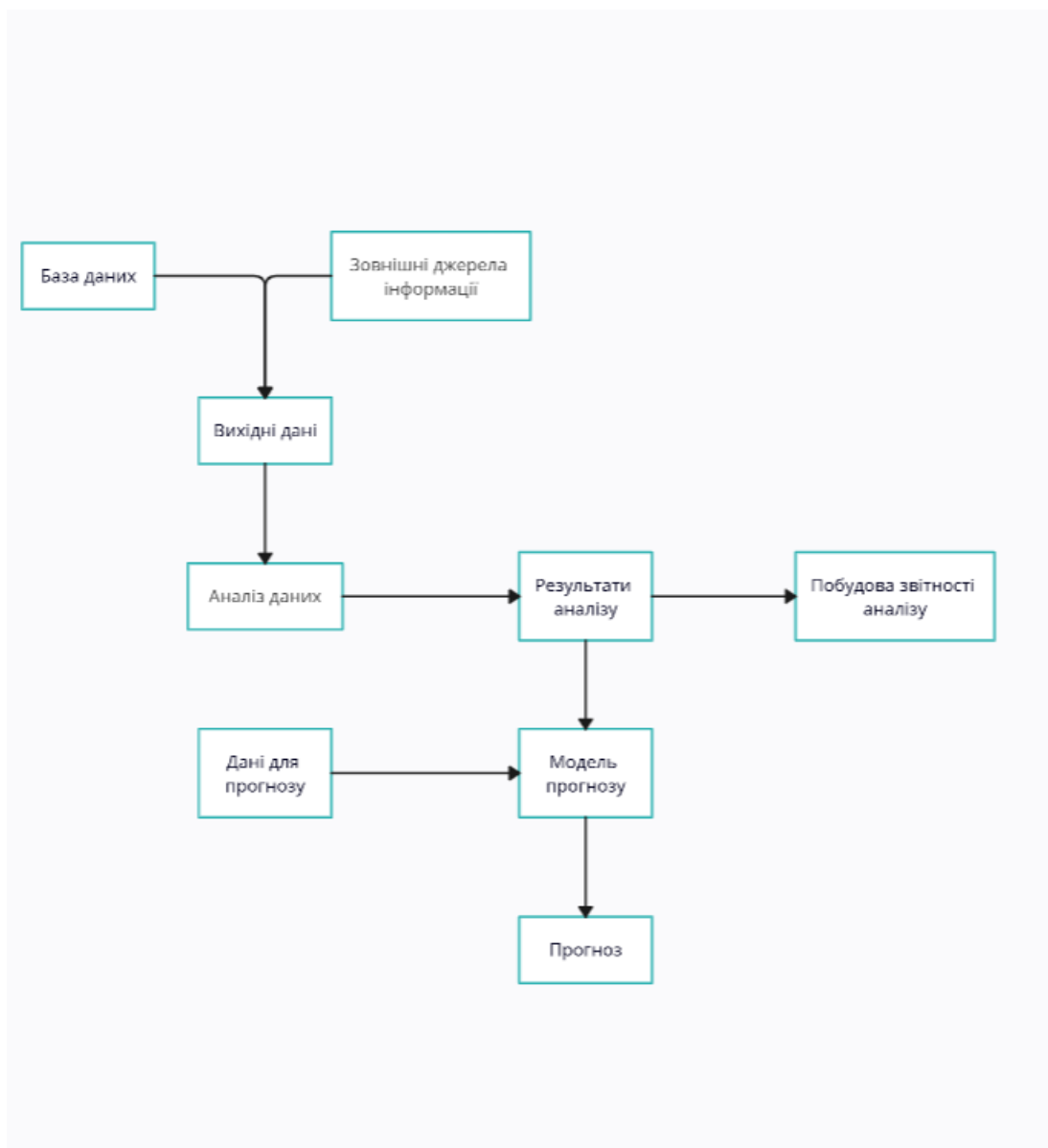


Рисунок 2.5. Блок-схема процесу прогнозування.

Прогнозування має широке застосування в різних галузях. У фінансах це може бути використано для прогнозування цін на акції або валютних курсів. У маркетингу - для передбачення тенденцій споживчого попиту або ефективності рекламних кампаній. У сфері здоров'я прогнозування допомагає в аналізі тенденцій захворюваності або результатів лікування. У логістиці та виробництві, прогнозування використовується для оптимізації запасів та планування виробничих процесів. Ефективне прогнозування дозволяє компаніям підвищити ефективність, знизити витрати та краще реагувати на майбутні виклики та можливості [10].

2.6.2 Аналіз часових рядів

Аналіз часових рядів - це метод статистичного аналізу, що використовується для спостереження за наборами даних, які змінюються протягом часу. Цей метод особливо корисний для ідентифікації та аналізу трендів, циклічних змін та сезонності у даних. Основна мета аналізу часових рядів полягає в моделюванні даних для прогнозування або визначення майбутніх значень заснованих на минулих спостереженнях. Це важливий інструмент в багатьох областях, таких як економіка, фінанси, наукові дослідження, соціологія, погодні прогнози, і багато інших, де змінні є функціями часу. Аналіз часових рядів дозволяє виявляти кореляції та причинно-наслідкові зв'язки в даних, що є ключовим для розуміння динаміки систем та прийняття обґрунтованих рішень [10].

Будь-який часовий ряд можна представити як суму детермінованого та випадкового компонентів:

$$y_t = d_t + r_t, \quad t = 1, T. \quad (2.5)$$

В свою чергу, детермінований компонент складається з трьох частин: трендового, сезонного, циклічного компонентів.

$$d_t = tr_t + s_t + c_t, \quad t = 1, T. \quad (2.6)$$

Таким чином, будь-який часовий ряд можна розглядати як суму:

$$y_t = tr_t + s_t + c_t + r_t, \quad t = 1, T. \quad (2.7)$$

Такий вигляд часового ряду отримав назву адитивної моделі. Якщо ж замість реальних значень компонентів використовувати їх логарифми, то отримаємо мультиплікативну модель:

$$\ln y_t = \ln tr_t + \ln s_t + \ln c_t + \ln r_t, \quad (2.8)$$

Або

$$y_t = tr_t \cdot s_t \cdot c_t \cdot r_t, \quad t = 1, T. \quad (2.9)$$

Детермінований елемент часових рядів змінюється згідно з визначеними правилами, які можна виявити через аналіз і дослідження. Часто час є основним фактором, що впливає на цей елемент.

У процесі аналізу часових рядів першим кроком є ідентифікація трендового компонента, що зазвичай легко виявляється при аналізі графіка. Науковці можуть використовувати аналітично визначені криві для опису цих змін.

Сезонний компонент відображає коливання, які відбуваються навколо основного тренду. Наприклад, у економічних часових рядах це може бути пов'язано зі змінами у виробництві та споживанні, які відбуваються сезонно. Аналіз сезонних коливань полягає у порівнянні даних за відповідні періоди.

Циклічний компонент розміщується між трендовим і сезонним компонентами. Він представляє собою плавні зміни, які відбуваються протягом довгого часу, відрізняючись від періодичних сезонних коливань, але і не належить до них.

Випадковий компонент є тією частиною часового ряду, яка залишається після видалення трендових, сезонних та циклічних компонентів. Він може містити наслідки непередбачених природних подій або випадкові людські дії. Наявність цього компонента робить точне прогнозування часових рядів складним і завжди супроводжується певною помилкою, адже будь-який реальний процес включає елемент випадковості [10].

2.7 Методи прогнозування витрат

Одним з найпростіших методів прогнозування ринкових тенденцій є екстраполяція, яка передбачає продовження існуючих трендів, що сформувались у

минулому, у майбутнє. Економічні показники, що розвиваються за певними тенденціями, до певної міри можуть передбачити їх подальші значення. Також важливо враховувати інерційність багатьох ринкових процесів, що особливо помітно при короткостроковому прогнозуванні. Однак, при довгостроковому прогнозуванні важливо враховувати потенційні зміни умов ринкового середовища.

- Традиційні методи.
- Активнісно-орієнтовані методи (ABC - Activity-Based Costing).
- Стратегічні методи.

Традиційні методи обсягу витрат базується на стандартному поділі витрат на прямі та непрямі. Прямі витрати легко відстежуються та призначаються конкретному продукту чи послугі, наприклад, матеріали та робоча сила, безпосередньо використані у виробництві. Непрямі витрати, такі як адміністративні витрати та витрати на оренду, розподіляються за допомогою різних баз розподілу, як-то виробнича площа або години робочої сили. Ці методи вважаються досить простими у використанні, але вони можуть не враховувати складність сучасних виробничих процесів [10].

Активнісно-орієнтовані методи калькулювання витрат зосереджується на витратах, пов'язаних з різними видами діяльності, необхідними для виробництва продукту чи надання послуги. Метою цього методу є точніше відображення використання ресурсів у виробничому процесі. Витрати класифікуються за активностями, а потім відносяться на продукти з урахуванням реального використання ресурсів. Цей метод дає краще розуміння того, як і чому виникають витрати, і дозволяє більш точно оцінити собівартість продукції та послуг [11].

Стратегічні методи обсягу витрат фокусуються на довгостроковій перспективі та включають аналіз витрат, що впливає на стратегічне планування і рішення. Ці методи використовуються для визначення вартості життєвого циклу продукту, аналізу витрат на ціннісний ланцюг, а також для ціноутворення з урахуванням факторів зовнішнього середовища та конкуренції. Стратегічні методи

дозволяють оцінити, як зміни у виробничих процесах, інновації, або зміни у ринкових умовах можуть вплинути на витрати та доходи компанії. Вони забезпечують ширший контекст для прийняття рішень, що виходять за рамки простого обліку витрат.

Прогнози обсягу витрат за допомогою експертів можуть бути отримані в одній з трьох форм:

- точкового прогнозу;
- інтервального прогнозу;
- прогнозу розподілу ймовірностей.

Точковий прогноз обсягу витрат представляє собою визначення конкретної суми витрат. Цей тип прогнозування є найбільш базовим, оскільки він включає мінімальну кількість даних. Зазвичай припускається, що точковий прогноз має ризик невірності, проте в методології не передбачені засоби для оцінки ймовірності або розміру помилки прогнозу. Тому в практичній діяльності більш часто використовуються альтернативні методи прогнозування, такі як інтервальний та ймовірнісний прогнози.

Інтервальний прогноз обсягу витрат включає визначення діапазону, в межах якого очікується реальне значення показника з певною ступенем достовірності. Для прикладу можна взяти будь яку компанію.

Таблиця 2.1 Приклад імовірності обсягу продаж компанії

Обсяг продаж компанії млн. грн.	Імовірність
23 – 24,5 млн. грн.	0,25
13 - 15 млн. грн.	0,50
28 - 29 млн. грн.	0,75

Під час створення прогнозу завжди існує шанс, що реальний обсяг продажів не впишеться у передбачений діапазон, проте аналітики вважають, що ця ймовірність настільки низька, що її можна не брати до уваги у процесі планування.

Ефективний прогноз та аналіз даних у сфері маркетингу сприяє плануванню оптимальних витрат та своєчасному управлінню запасами, що має ключове значення для збереження конкурентоспроможності компанії на ринку в майбутньому.

Висновок до розділу 2

У цьому розділі було розглянуті існуючі інструменти інтелектуального аналізу даних. Також було визначення потреби в створенні інформаційної системи для планування асортименту.

Використання інструментів Data Mining в Orange в цій системі сприятиме покращенню процесів прийняття рішень планування асортименту. Це також включає ІАД, що дозволяє компанії систематизувати зібрані дані та представляти їх у зручних форматах для ухвалення управлінських рішень.

Впровадження інтелектуального аналізу даних, заснованої на вітрині даних MS SQL Server 2010, забезпечить ТОВ "Світ каменю" більш ефективні інструменти для обробки та аналізу даних. Це, в свою чергу, створить основу для формування рекомендаційних стратегій планування асортименту.

РОЗДІЛ 3. ЗАСТОСУВАННЯ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ПЛАНУВАННЯ АСОРТИМЕНТУ

3.1 Розробка вітрини даних

Розглянемо задачі планування асортименту. Виявлено, що на підприємстві неправильно побудована стратегія планування асортименту, вона не ефективно відображає потреби ринку та можливості виробництва, що призводить до втрати доходів та нераціонального використання ресурсів. Для вирішення цієї проблеми було вирішено провести кластерний аналіз існуючих даних з метою ідентифікації потенційних груп продуктів, які мають схожі характеристики.

Для вирішення цієї проблеми потрібно:

- Згенерувати вітрину даних, на основі якої можна буде проводити інтелектуальний аналіз даних.
- Застосувати кластерний аналіз для групування продуктів на основі їхніх характеристик та поведінки споживачів.
- Проаналізувати структуру кластерів для розуміння їхніх особливостей та вплив на планування асортименту.
- Зробити висновки та дати рекомендації для оптимізації асортименту продукції на основі результатів кластерного аналізу.

Вітрина даних (ВД) значно спрощує процес створення запитів для обробки та аналізу даних, оскільки воно забезпечує централізоване, оптимізоване та інтегроване зберігання даних з різних джерел. Це дозволяє аналітикам зосередитись на видобутку даних, а не на вирішенні технічних проблем, пов'язаних з доступом до даних. Вітрина даних спрощує доступ до історичних та актуальних даних, уніфікуючи формати і структури, що знижує складність запитів і збільшує швидкість отримання результатів, що створює сприятливе середовище для глибокого аналізу, дозволяючи аналітикам легко здійснювати складні запити, об'єднання даних та їх візуалізацію, що є ключовим для інформованого прийняття рішень [14].

ВД також підтримує високий рівень даних та їх якість, забезпечуючи надійність та точність інформації, що є критично важливим для аналітичних висновків. Завдяки їх відокремленню від операційних систем, ВД забезпечує стабільність та консистенцію даних, що знижує ризики, пов'язані з обробкою та аналізом. Окрім цього, використання ВД дозволяє імплементувати розширені аналітичні функції, такі як машинне навчання, що розширює можливості аналітики та підвищує цінність витягнутих даних. Таким чином, ВД не тільки спрощує обробку даних, а й підвищує ефективність аналітичних процесів, дозволяючи компаніям робити більш обґрунтовані та стратегічні рішення [14].

Вітрина даних оптимізоване для ефективного вилучення даних. Так як агрегація даних відбувається на сервері сховища даних, а не в Microsoft Office Excel, це зменшує обсяг даних, які потрібно передавати в Excel при створенні або оновленні звітів. Ця стратегія дозволяє обробляти значно більші обсяги даних, ніж при традиційному підході, де Excel мусить завантажувати кожен звіт окремо для подальшого обчислення загальних значень.

Побудова моделі кластеризації в Orange передбачає створення вітрини даних та структури інтелектуального аналізу даних.

Використовуючи AllFusion ERWin Data Modeler була побудована структура вітрини даних, що є основою для розмірної моделі вітрини даних, зображеної на рисунку 3.1.

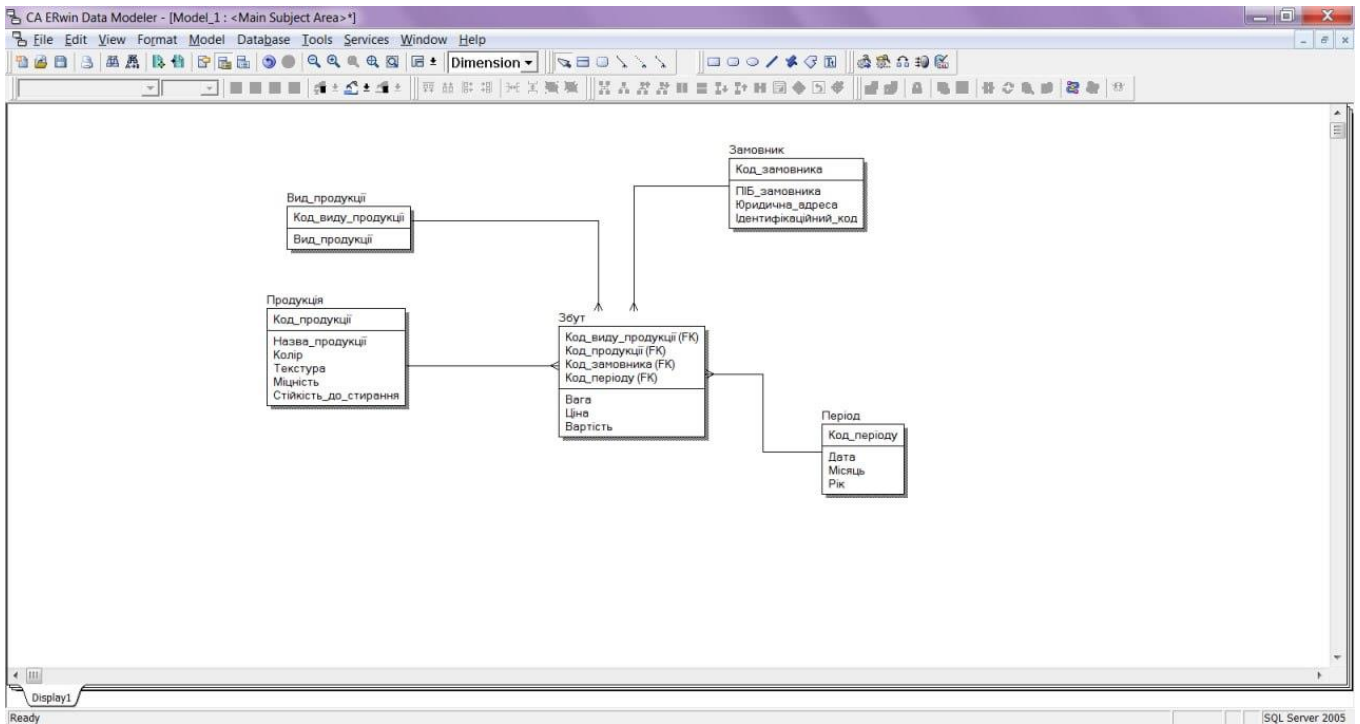


Рисунок 3.1 Вітрина даних в AllFusion ERWin Data Modeler.

Фізична модель вітрини даних з використанням ERWin Data Modeler згенерована в СУБД MS SQL Server, вітрина даних наведена на рисунку 3.2.

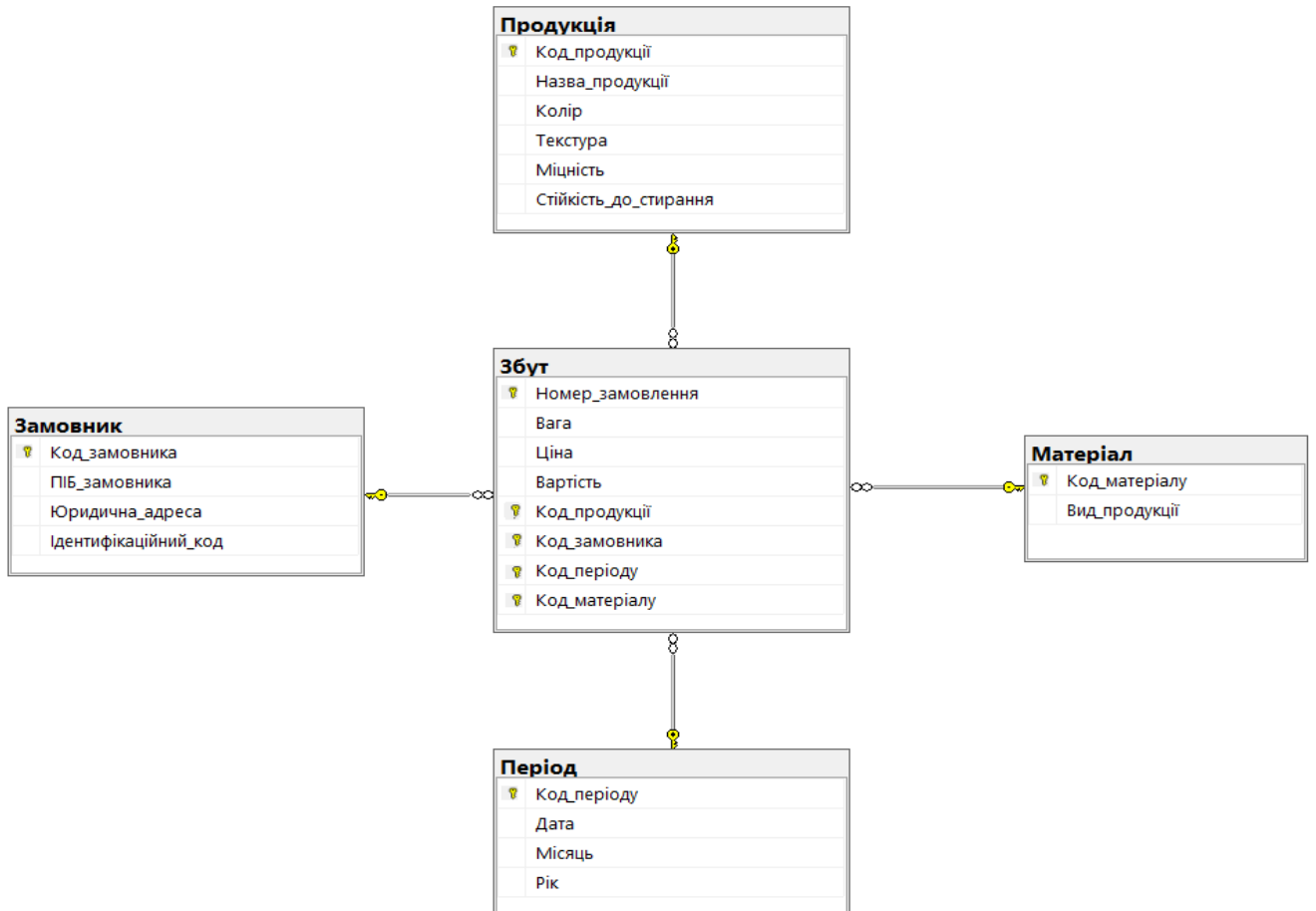


Рисунок 3.2 Схема вітрини даних в MS SQL Server.

На основі даних, розміщених у вітрині даних, побудовано представлення, наведене на рисунку 3.3, що є джерелом даних для аналізу інформації і побудови аналітичних звітів.

Вид_проду...	Назва_про...	Колір	Текстура	Міцність	Стійкість_д...	Місяць	Рік	ПБ_замов...	Юридична...	Номер_за...	Вага	Вартість
Штучний кв...	Капустянськ...	Червоний	Зерниста	Висока	Висока	Січень	2023	Шевчук Ол...	Дрогобич, ...	1	219,00	8322,00
Онiкс	Кіліманджа...	Світло-кор...	Прозора з ...	Низька	Низька	Січень	2023	Мельник Д...	Ковель, вул...	2	88,00	3872,00
Онiкс	Китайський...	Коричневий	Зерниста	Висока	Висока	Січень	2023	Коваль Сер...	Стрий, вул ...	3	126,00	3654,00
Штучний кв...	Мультиколі...	Різнокольо...	Гладка з жи...	Середня	Середня	Січень	2023	Лисенко Д...	Нововолин...	4	142,00	6106,00
Онiкс	Фінляндськ...	Темно-синій	Зерниста	Висока	Висока	Січень	2023	Бондаренк...	Луцьк, вул ...	5	73,00	2044,00
Мармур	Індійський ...	Рожево-жо...	Зерниста	Висока	Висока	Січень	2023	Лисенко Вл...	Дрогобич, ...	6	83,00	3735,00
Мармур	Онiкс Оранж	Оранжевий	Прозора з ...	Низька	Низька	Січень	2023	Коваль Ма...	Сарни, вул ...	7	92,00	3496,00
Інші види о...	Кварцовий...	Різнобарвн...	Однорідна ...	Висока	Висока	Січень	2023	Лисенко О...	Нововолин...	8	139,00	6811,00
Граніт	Штучний кв...	Зелений з ...	Однорідна ...	Висока	Висока	Січень	2023	Коваленко ...	Ковель, вул...	9	183,00	4941,00
Граніт	Онiкс Грідж...	Сірий	Прозора з ...	Низька	Низька	Січень	2023	Кравченко ...	Львів, вул ...	10	129,00	4515,00
Штучний кв...	Онiкс Драг...	Сірий	Зерниста	Висока	Висока	Лютий	2023	Коваленко ...	Луцьк, вул ...	11	144,00	4032,00
Травертин	Кварцевий ...	Білий з бли...	Однорідна ...	Висока	Висока	Лютий	2023	Шевченко ...	Луцьк, вул ...	12	92,00	3496,00
Онiкс	Крема	Світло-кре...	Прозора з ...	Низька	Низька	Лютий	2023	Мельник Ів...	Дрогобич, ...	13	127,00	5461,00
Онiкс	Індійський ...	Рожево-жо...	Зерниста	Висока	Висока	Лютий	2023	Кравченко ...	Львів, вул ...	14	151,00	6946,00
Граніт	Кварцовий...	Різнобарвн...	Однорідна ...	Висока	Висока	Лютий	2023	Мельник М...	Червоногр...	15	92,00	3496,00
Штучний кв...	Онiкс Драг...	Червоний ...	Прозора з ...	Низька	Низька	Лютий	2023	Коваленко ...	Дубно, вул ...	16	117,00	2340,00
Штучний кв...	Кварцовий...	Кремовий ...	Однорідна ...	Висока	Висока	Лютий	2023	Коваленко ...	Луцьк, вул ...	17	192,00	7872,00
Онiкс	Коричневи...	Коричневий	Гладка з жи...	Середня	Середня	Лютий	2023	Кравченко ...	Львів, вул ...	18	183,00	4941,00
Граніт	Індійський ...	Золотистий...	Зерниста	Висока	Висока	Лютий	2023	Мельник Ів...	Ковель, вул...	19	183,00	4941,00
Травертин	Кварцовий...	Багатобарв...	Однорідна ...	Висока	Висока	Лютий	2023	Кравченко ...	Червоногр...	20	192,00	7872,00
Граніт	Дідковицьк...	Світло-сіри...	Зерниста	Висока	Висока	Березень	2023	Мельник Ів...	Ковель, вул...	21	92,00	3496,00
Онiкс	Кварцовий...	Різнобарвн...	Однорідна ...	Висока	Висока	Березень	2023	Шевченко ...	Сарни, вул ...	22	219,00	8322,00
Штучний кв...	Китайський...	Коричневий	Зерниста	Висока	Висока	Березень	2023	Мельник Д...	Ковель, вул...	23	205,00	7790,00
Травертин	Делікато	Блідо-роже...	Прозора з ...	Низька	Низька	Березень	2023	Лисенко Вл...	Дрогобич, ...	24	139,00	6811,00
Штучний кв...	Капустянськ...	Червоний	Зерниста	Висока	Висока	Березень	2023	Коваль Іва...	Червоногр...	25	147,00	4263,00
Травертин	Кварцовий...	Різнобарвн...	Однорідна ...	Висока	Висока	Березень	2023	Шевченко ...	Дубно, вул ...	26	142,00	6106,00
Штучний кв...	Лабрадори...	Синій з зол...	Зерниста	Висока	Висока	Березень	2023	Бондаренк...	Луцьк, вул ...	27	142,00	6106,00
Інші види о...	Граніт Влас...	Чорний з м...	Зерниста	Висока	Висока	Березень	2023	Кравченко ...	Дрогобич, ...	28	83,00	3735,00
Мармур	Вене Ціано	Світло-зеле	Прозора з ...	Низька	Низька	Березень	2023	Лисенко Ів...	Сарни, вул ...	29	126,00	3654,00

Рисунок 3.3 Представлення даних у вітрині даних.

3.2 Кластерний аналіз даних

Обсяг збуту продукції демонструє, яка продукція користується попитом та які товари мають найбільші обсяги продажів. На основі, наведених на рисунку 3.4 даних, лідером продаж є «Онiкс алабастро», всього було продано 15 одиниць продукції, у відсотковому співвідношенні це 8,15% від загального обсягу збуту. За ним розташований «мармур» та «мармур червоний», разом було продано 24 одиниці продукції, в відсотковому співвідношенні це 13,04%.

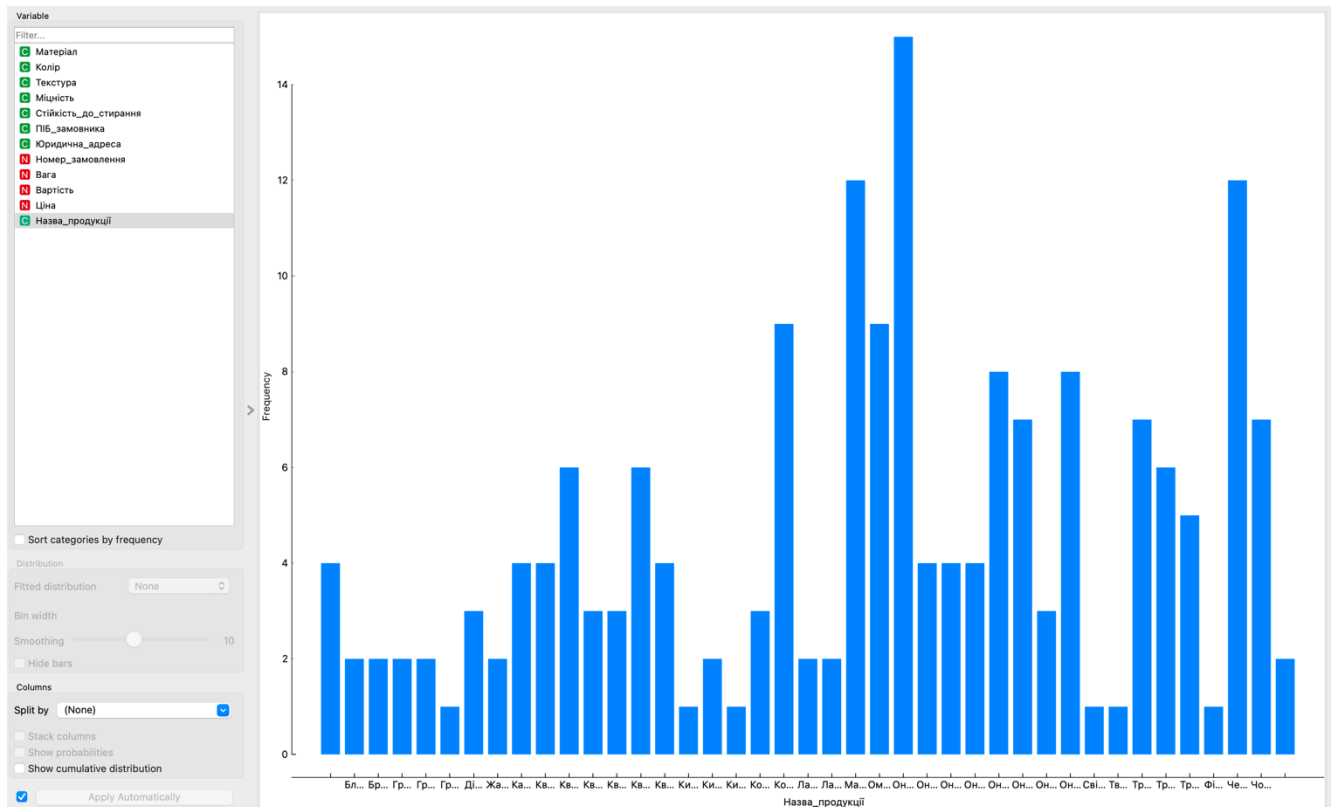


Рисунок 3.4 Обсяг збуту продукції.

Обсяг збуту продукції (Рисунок 3.5) демонструє, який матеріал був найбільш використовуваним для виготовлення продукції, що мала найбільший обсяг збуту. Матеріал «Онікс», відповідає найбільшому обсягу збуту, всього було продано 55 одиниць продукції, що у відсотковому співвідношенні складає 29,89%, наступним буде мармур, всього було продано 46 одиниць продукції це 25% від обсягу збуту.

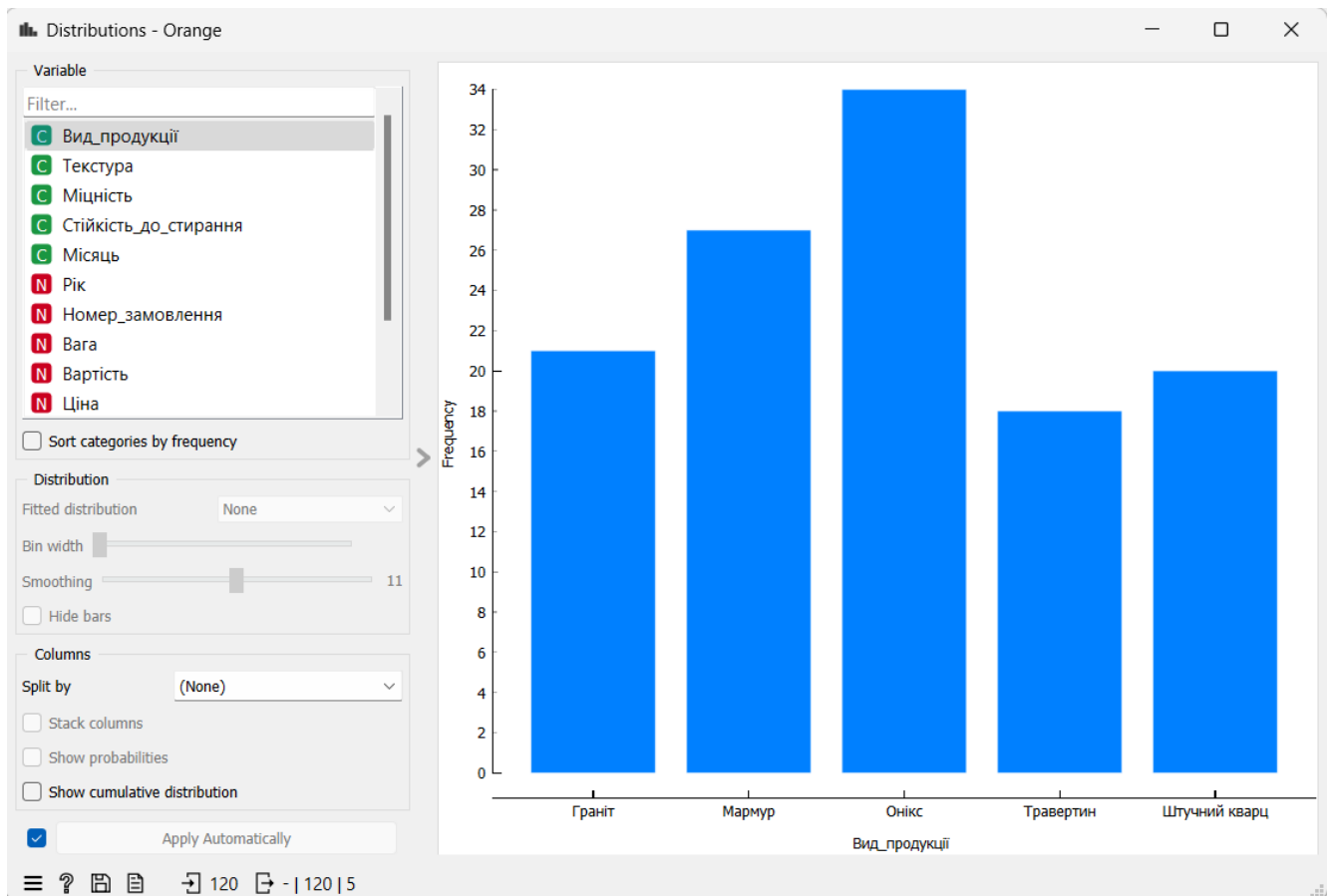


Рисунок 3.5 Обсяг збуту в залежності від матеріалу виготовлення.

Наведений на рисунку 3.6 робочий процес в ПП Orange призначений для організації та виконання структурованого кластерного аналізу даних для виявлення взаємопов'язаних груп або кластерів у наборі даних, пов'язаних з продукцією підприємства. Через візуалізацію розподілу даних із застосуванням алгоритму k-середніх, схема допомагає ідентифікувати приховані патерни та структури в даних, що слугує основою для прийняття обґрунтованих рішень щодо планування асортименту продукції. Це сприяє підвищенню ефективності виробничого планування та вдосконаленню стратегій.

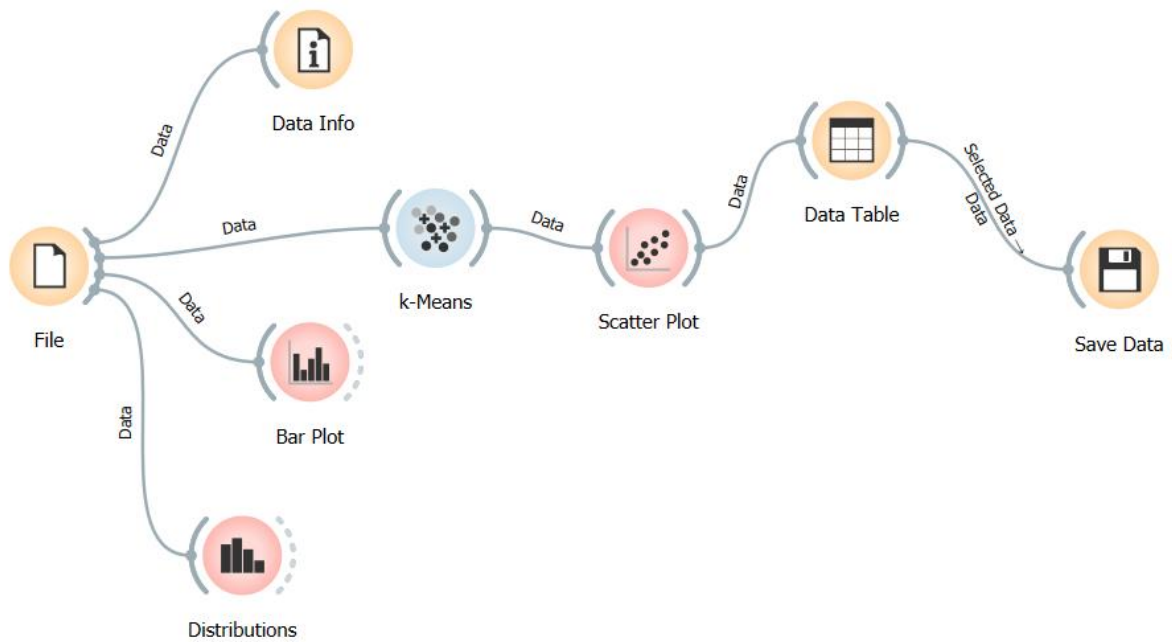


Рисунок 3.6 Схеми робочого процесу в Orange

Розглянемо схему моделі в Orange, що відображає послідовність кроків обробки даних за допомогою різних віджетів:

File: Початковий віджет, який імпортує дані з файлу.

Data Info: Віджет, який надає інформацію про завантажені дані, таку як кількість зразків та ознак, відсутні значення тощо.

k-Means: Віджет для кластеризації даних за допомогою алгоритму k-середніх, який групує дані на основі схожості.

Distributions: Віджет для відображення розподілів значень в ознаках даних, зазвичай у вигляді гістограм.

Scatter Plot: Віджет, який створює розсіяну діаграму, відображаючи взаємозв'язки між двома ознаками даних.

Bar Plot: Віджет для створення стовпчикових діаграм, які можуть відображати різні статистичні дані, такі як кількість, середні значення тощо.

Data Table: Віджет, що відображає дані у табличному форматі, дозволяючи переглядати конкретні значення датасету.

Після побудови схеми робочого процесу в Orange можемо реалізувати графічне представлення результатів кластерного аналізу (Рисунок 3.7).

Вісь x «Стійкість_до_стирання», вісь y "Текстура", кольори кластерів це «Матеріал». у, наведені на рисунку 3.8.

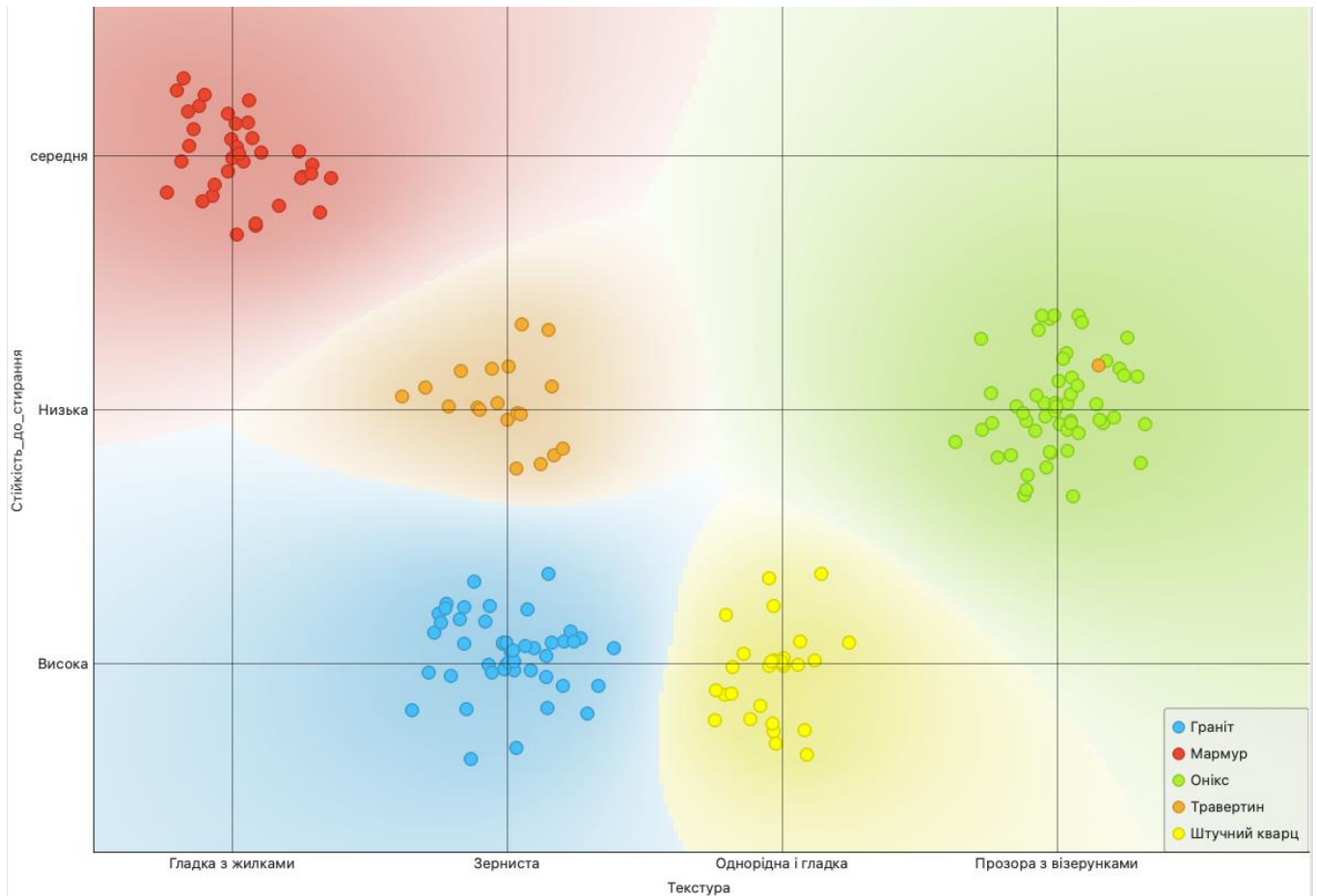


Рисунок 3.7 Графічне представлення результатів кластерного аналізу

	Selected	Назва_продукції	Cluster	Silhouette	Матеріал	Колір	Текстура	Міцність	йкість_до_стирання	Вага	Вартість
1	No	Капустянський граніт Rosso Santiago (...)	C3	0.588692	Граніт	Світло-сіри...	Зерниста	Висока	Висока	219.00	8322.00
2	No	Онікс Медовий	C2	0.600467	Онікс	Медово-жо...	Прозора з в...	Низька	Низька	88.00	3872.00
3	No	Онікс Оранж	C2	0.590257	Онікс	Оранжевий	Прозора з в...	Низька	Низька	126.00	3654.00
4	No	Кварцевий камінь White diamond	C3	0.581707	Штучний кв...	Білий з блис...	Однорідна і ...	Висока	Висока	142.00	6106.00
5	No	Онікс Верде Смеральдо	C2	0.584052	Онікс	Зелений	Прозора з в...	Низька	Низька	73.00	2044.00
6	No	Червоний мармур	C1	0.618166	Мрамур	Червоний	Гладка з жи...	середня	середня	83.00	3735.00
7	No	Червоний мармур	C1	0.621679	Мрамур	Червоний	Гладка з жи...	середня	середня	92.00	3496.00
8	No	Онікс Медовий	C2	0.594852	Онікс	Медово-жо...	Прозора з в...	Низька	Низька	139.00	6811.00
9	No	Корнинський граніт Leopard (GG1)	C3	0.617738	Граніт	Сіро-жовти...	Зерниста	Висока	Висока	183.00	4941.00
10	No	Омелянівський граніт Rosso Toledo (G...	C3	0.616166	Граніт	Темно-черв...	Зерниста	Висока	Висока	129.00	4515.00
11	No	Кварцевий камінь White diamond	C3	0.59529	Штучний кв...	Білий з блис...	Однорідна і ...	Висока	Висока	144.00	4032.00
12	No	Травертин зелений	C2	0.557831	Травертин	Зелений	Зерниста	Низька	Низька	92.00	3496.00
13	No	Онікс Алабастро	C2	0.606734	Онікс	Білий з сіри...	Прозора з в...	Низька	Низька	127.00	5461.00
14	No	Онікс Алабастро	C2	0.597452	Онікс	Білий з сіри...	Прозора з в...	Низька	Низька	151.00	6946.00
15	No	Жадківський граніт Rosa Raveno	C3	0.595903	Граніт	Темно-черв...	Зерниста	Висока	Висока	92.00	3496.00
16	No	Кварцовий камінь Fusion Wow	C3	0.579225	Штучний кв...	Багатобарв...	Однорідна і ...	Висока	Висока	117.00	2340.00
17	No	Кварцовий камінь Perla Venata	C3	0.580695	Штучний кв...	Кремовий з ...	Однорідна і ...	Висока	Висока	192.00	7872.00
18	No	Онікс Медовий	C2	0.575835	Онікс	Медово-жо...	Прозора з в...	Низька	Низька	183.00	4941.00
19	No	Індійський граніт Kashmir Gold	C3	0.608327	Граніт	Золотистий ...	Зерниста	Висока	Висока	183.00	4941.00
20	No	Травертин білий	C2	0.540215	Травертин	Білий	Зерниста	Низька	Низька	192.00	7872.00
21	No	Дідковицький граніт Star of Ukraine (G...	C3	0.589954	Граніт	Світло-сіри...	Зерниста	Висока	Висока	92.00	3496.00
22	No	Онікс Оранж	C2	0.570852	Онікс	Оранжевий	Прозора з в...	Низька	Низька	219.00	8322.00
23	No	Кварцовий камінь Silestone	C3	0.589511	Штучний кв...	Різобарвний	Однорідна і ...	Висока	Висока	205.00	7790.00
24	No	Травертин червоний	C2	0.552056	Травертин	Червоний	Зерниста	Низька	Низька	139.00	6811.00
25	No	Штучний кварц Tartaruga	C3	0.594354	Штучний кв...	Бежевий з т...	Однорідна і ...	Висока	Висока	147.00	4263.00
26	No	Твартин зелений	C2	0.555668	Травертин	Зелений з т...	Зерниста	Низька	Низька	142.00	6106.00
27	No	Штучний кварц Tartaruga	C3	0.58056	Штучний кв...	Бежевий з т...	Однорідна і ...	Висока	Висока	142.00	6106.00
28	No	Онікс Оранж	C2	0.597349	Онікс	Оранжевий	Прозора з в...	Низька	Низька	83.00	3735.00
29	No	Мрамур	C1	0.605554	Мрамур	Білий	Гладка з жи...	середня	середня	126.00	3654.00
30	No	Мрамур	C1	0.618786	Мрамур	Білий	Гладка з жи...	середня	середня	125.00	4625.00
31	No	Мрамур	C1	0.560303	Мрамур	Зелений	Гладка з жи...	середня	середня	212.00	5300.00
32	No	Коричневий мармур	C1	0.598636	Мрамур	Коричневий	Гладка з жи...	середня	середня	163.00	6357.00
33	No	Китайський граніт Yellow	C3	0.586583	Граніт	Жовтий	Зерниста	Висока	Висока	142.00	6390.00
34	No	Травертин білий	C2	0.538465	Травертин	Білий	Зерниста	Низька	Низька	83.00	1909.00
35	No	Онікс Оранж	C2	0.588153	Онікс	Античний к...	Прозора з в...	Низька	Низька	174.00	6960.00
36	No	Травертин білий	C2	0.51833	Травертин	Білий	Зерниста	Низька	Низька	208.00	4576.00
37	No	Блю Онікс	C2	0.59415	Онікс	Синій	Прозора з в...	Низька	Низька	83.00	3818.00
38	No	Травертин червоний	C2	0.557705	Травертин	Червоний	Зерниста	Низька	Низька	84.00	3948.00
39	No	Онікс Алабастро	C2	0.606438	Онікс	Білий з сіри...	Прозора з в...	Низька	Низька	129.00	5547.00
40	No	Кварцовий камінь Dekton®	C3	0.601776	Штучний кв...	Різобарвний	Однорідна і ...	Висока	Висока	156.00	4680.00
41	No	Чорний мармур	C1	0.574971	Мрамур	Чорний	Гладка з жи...	середня	середня	218.00	8720.00
42	No	Онікс Алабастро	C2	0.580397	Онікс	Коричневий	Прозора з в...	Низька	Низька	56.00	2744.00
43	No	Онікс Драгон	C2	0.60121	Онікс	Червоний з ...	Прозора з в...	Низька	Низька	103.00	3914.00
44	No	Кварцовий камінь Dekton®	C3	0.563353	Штучний кв...	Різобарвний	Однорідна і ...	Висока	Висока	51.00	1224.00
45	No	Мрамур	C1	0.603548	Мрамур	Білий	Гладка з жи...	середня	середня	115.00	3220.00
46	No	Кварцовий камінь Dekton®	C3	0.558776	Штучний кв...	Різобарвний	Однорідна і ...	Висока	Висока	79.00	3871.00
47	No	Дідковицький граніт Star of Ukraine (G...	C3	0.608329	Граніт	Світло-сіри...	Зерниста	Висока	Висока	167.00	5845.00
48	No	Лабрадорит Irlina Blue (L2)	C3	0.567684	Граніт	Синій з зол...	Зерниста	Висока	Висока	52.00	1144.00
49	No	Онікс Алабастро	C2	0.593862	Онікс	Білий з сіри...	Прозора з в...	Низька	Низька	171.00	6156.00
50	No	Онікс Miele	C2	0.575391	Онікс	Медовий	Прозора з в...	Низька	Низька	207.00	8280.00
51	No	Чорний мармур	C1	0.605241	Мрамур	Чорний	Гладка з жи...	середня	середня	163.00	6520.00

Рисунок 3.8 Деталізація кластерів у табличному вигляді

Проаналізуємо структуру кластерів, що наведена на рисунку 3.6 і містить загальне розподілення кластерів на площині. Кожна група кольорів, відповідає поведінці покупців по каналах збуту в кластері. Кожному кольору відповідає продукція (Блакитний – «граніт», червоний – «мармур», зелений – «онікс» тощо). Проаналізувавши характеристики кластеру в табличному вигляді можна зазначити, що в кластери C2 входить продукція, яка має найбільший обсяги реалізації, при цьому характеристики цієї продукції є текстура зерниста та прозора з візерунком, міцність низька та висока, стійкість до стирання низька та висока.

Дану групу продукції за цими характеристиками, можна рекомендувати для планування асортименту в залежності від попиту споживачів. Кластер C2, який включає в себе продукти з високими обсягами продажу, вказує на сильні позиції цих товарів на ринку.

Продукція з високою стійкістю до стирання та високою міцністю, як «Омелянський граніт Rosso Toledo» та «Індійський граніт Kashmir Gold», буде ідеальною для місць з високим навантаженням, таких як громадські простори або комерційні підлоги. Продукція з низькою міцністю та прозорою текстурою, як «Онiкс алабастро» та «Онiкс драгон», може бути більш підходящою для декоративних цілей або для внутрішнього оздоблення приміщень, де в пріоритеті естетика, а не функціональне навантаження.

Враховуючи характеристики, можна робити висновок про потребу акцентувати увагу на цих видах продукції при розробці стратегії планування асортименту.

Таблиця 3.1 Список продукції для планування асортименту

Назва продукції	Матеріал	Стійкість до стирання	Міцність	Текстура	Колір
Онiкс алабастро	Онiкс	Низька	Низька	Прозора з візерунком	Білий з сірими візерунками
Онiкс драгон	Онiкс	Низька	Низька	Прозора з візерунком	Червоний з чорними візерунками
Омелянський граніт Rosso Toledo (GR8)	Граніт	Висока	Висока	Зерниста	Темно червоний
Травертин червоний	Травертин	Низька	Низька	Зерниста	Червоний
Індійський граніт Kashmir Gold	Граніт	Висока	Висока	Зерниста	Золотистий з чорними та сірими плямами
Капустянський граніт Rosso Santiago	Граніт	Висока	Висока	Зерниста	Світло-сірий з синіми плямами

Висновок до розділу 3

Існуюча система планування асортименту не ефективно відображає потреби ринку та можливості виробництва, що призводить до втрати доходів та нерационального використання ресурсів. Тому було прийняте рішення провести кластерний аналіз існуючих даних для ідентифікації потенційних груп продуктів з схожими характеристиками.

Інформаційно-аналітична система служить фундаментом для проведення інтелектуального аналізу даних, дозволяючи ефективно вводити, зберігати та обробляти великі обсяги даних. З використанням інструменту «Orange», аналітики мають можливість підключати джерела даних, обробляти їх та застосовувати різноманітні алгоритми для виявлення закономірностей та кластерів. Це особливо важливо, коли існуюча система планування асортименту не відображає ефективно потреби ринку та можливості виробництва, ведучи до фінансових втрат та неефективного використання ресурсів.

Вітрина даних (ВД) значно спрощує процес створення запитів для обробки та аналізу даних, забезпечуючи централізоване, оптимізоване та інтегроване зберігання даних з різних джерел. ВД підтримує високий рівень даних та їх якість, що є критично важливим для аналітичних висновків.

Кластерний аналіз дозволяє виявити взаємопов'язані групи або кластери у наборі даних, пов'язаних з продукцією підприємства, та розробити рекомендації для оптимізації асортименту продукції.

Отримані результати кластерного аналізу вказують на сильні позиції продукції за певними та можуть бути використані для розробки стратегії планування асортименту, що відповідає потребам і перевагам споживачів.

ВИСНОВКИ

В даній магістерській роботі проведено дослідження методів інтелектуального аналізу даних які були застосовані на підприємстві ТОВ «Світ каменю». Було проведено кластерний аналіз, що дало можливість зробити рекомендації по плануванні асортименту, за допомогою статистичного методу Data Mining з використанням алгоритму K-means, реалізовано з використанням інструменту «Orange».

В ході дослідження було проведено:

- Аналіз діяльності підприємства;
- Аналіз існуючих систем;
- Використовуючи AllFusion ERWin Data Modeler була побудована структура вітрини даних;
- Згенеровану вітрину даних в MS SQL Server;
- Проведений кластерний аналіз даних, на основі якого сформовано рекомендаційну стратегію планування асортименту.

Кластерний аналіз допоміг ідентифікувати ключові групи продукції, що користуються попитом, та сформувані рекомендації по плануванню асортименту, забезпечуючи таким чином більшу його відповідність потребам ринку та можливостям виробництва.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Методичні рекомендації до виконання кваліфікаційної роботи на здобуття освітнього ступеня «Магістра» спеціальності 122 «Комп'ютерні науки» освітньо-професійної програми «Комп'ютерні науки» денної та заочної форм навчання / уклад. : Л. Г. Загоровська, О. М. М'якшило, М. П. Костіков. – К. : НУХТ, 2024. – 28 с.
2. Проектування UML діаграм [Електрон. ресурс]. – режим доступу: <https://evergreens.com.ua/ua/articles/uml-diagrams.html>
3. Діаграма прецедентів [Електрон. ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Діаграма_прецедентів
4. IBM SPSS [Електрон. ресурс]. – Режим доступу: <https://www.ibm.com/spss>
5. Інтелектуальний аналіз даних [Електрон. ресурс]. – Режим доступу: <https://science.lpnu.ua/sites/default/files/journal-paper/2017/nov/6688/1049-58.pdf>
6. Інтелектуальний аналіз даних. [Електронний ресурс]: лабораторний практикум для студентів освітнього ступеня "Магістр" спеціальності 122 "Комп'ютерні науки та інформаційні технології" денної та заочної форм навчання / уклад.: О.М. М'якшило, О.В. Харкянен, – К.: НУХТ, 2016. – 25 с.
7. Харкянен О.В. Інтелектуальний аналіз даних [Електронний ресурс]: конспект лекцій для здобувачів освітнього ступеня «магістр» спеціальності 22 “Комп'ютерні науки” освітньо-професійної програми «Комп'ютерні науки» денної та заочної форм навчання / О.В. Харкянен – К.: НУХТ, 2020. – 111 с.
8. Чубукова И. А. Data mining - Електрон. дані (1 файл) – Режим доступу: <https://core.ac.uk/download/pdf/323532557.pdf> (дата звернення 20.02.2018) – Назва з екрану.
9. Cluster Analysis [Електрон. ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Кластерний_аналіз
10. Belmouhcine A., Benkhalifa M. Implicit Links based Web Page Representation for Web Page Classification // Proc. of the 5th Intern. conf. on Web Intelligence, Mining and Semantics. Larnaca, Cyprus, 2015.

11. Data mining - Електрон. дані (1 файл) – Режим доступу: https://ru.wikipedia.org/wiki/Data_mining – Назва з екрану
12. Інтелектуальний аналіз даних. Класифікація і регресія - Електрон. дані (1 файл) – Режим доступу: <http://ukrbukva.net/print:page,1,44734-%2082%20intellektual-nyiy-analiz-dannyh-klassifikaciya-i-regressiya.html> - Назва з екрану.
13. Christopher Pal, Mark Hall, Eibe Frank, Ian Witten. Data Mining: Practical Machine Learning Tools and Techniques, 4rd ed. / Morgan Kaufmann, 2016.
14. Бойко В.В., Савінков В.М. Проектування баз даних інформаційних систем. - М.: Фінанси і статистика, 1989. - 351 с.
15. Гаркуша І.М. Конспект лекцій з дисципліни “Проектування інформаційних систем” для студентів галузі знань 12 “Інформаційні технології” спеціальності 126 “Інформаційні системи та технології”. – Д.: НТУ «ДП», 2020. – 75 с.