Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____*Комп'ютерних наук*_____
(повна назва)
Кафедра_____*Системотехніки*_____
(повна назва)

# КВАЛІФІКАЦІЙНА РОБОТА
# Пояснювальна записка

рівень вищої освіти _____другий (магістерський)_____

Дослідження методів інтелектуального аналізу даних для виявлення цільових груп клієнтів у готельному бізнесі
(тема)

Виконала:
студентка 2 курсу, групи *ІТПм-22-1*
*Семенцова А.М.*
(прізвище, ініціали)

Спеціальність *122 Комп'ютерні науки*
(код і повна назва спеціальності)
Тип програми освітньо-професійна
Освітня програма *Інформаційні технології проектування*
(повна назва освітньої програми)

Керівник *доц. Імангулова З.А.*
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ *Гребеннік І.В.*
(підпис) (прізвище, ініціали)

2024 р.

*Я як студентка ХНУРЕ розумію і підтримую політику закладу із академічної доброчесності. Я не надавав і не одержував недозволену допомогу під час підготовки кваліфікаційної роботи. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.*

*19.01.2023* *Семенцова А.М.*

*Кваліфікаційна робота не містить відомостей заборонених до відкритого опублікування*

*Керівник кваліфікаційної роботи* 

*Кваліфікаційна робота виконана у відповідності до стандартів, що діють в Україні*

*Керівник кваліфікаційної роботи* 

*Попередній захист проведено* *19.01.2024*

*Керівник кваліфікаційної роботи* 

Харківський національний університет радіоелектроніки

Факультет *Комп'ютерних наук*

Кафедра *Системотехніки*

Рівень вищої освіти другий (магістерський)

Спеціальність *122 Комп'ютерні науки*
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма *Інформаційні технології проектування*
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«___» _____20___ р.

# ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові *Семенцовій Анастасії Миколаївні*
(прізвище, ім'я, по батькові)

1. Тема роботи: *Дослідження методів інтелектуального аналізу даних для виявлення цільових груп клієнтів у готельному бізнесі*
затверджена наказом по університету від *20.11* 2023 р. № *1373Ст*
2. Термін подання студентом роботи до екзаменаційної комісії: 23.01.2024 р
3.Вихідні дані до роботи: Експериментальний аналіз кластерів сегментів ринку для готельного бізнесу. Проведення аналізу предметної області, конкурентів в даній області та їх програмні рішення. Визначення переліку необхідних вимог, вхідних та вихідних даних. Перелік використовуваних програмних засобів: Jupyter Notebook.

4. Перелік питань, що потрібно опрацювати в роботі: *Вступ. Аналіз предметної області. Опис сучасного стану розвитку інформаційних систем. Аналіз застосування досліджуваних методів в існуючих системах. Актуальність досліджуваних методів. Постановка задачі. Постановка задачі на дослідження. Опис об'єкта дослідження. Вхідні та вихідні дані. Загальні мета та критерії даного дослідження. Огляд методів інтелектуального аналізу даних. Методи інтелектуального дослідження даних. Методи кластерного аналізу. Методи класифікації. Методи асоціативних правил. Методи регресійного аналізу. Узагальнення методів для вирішення задачі сегментації цільових груп. Математичний опис методів інтелектуального аналізу даних. Математичний опис кластерного аналізу. Експериментальний аналіз кластерів сегменту ринку для готельного бізнесу. Узагальнена інформація про аналіз та його інструменти. Опис середовища розробки. Вхідні данні. Опис вхідних даних. Кластерізація k-means.*

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій: Об'єкт, предмет та мета дослідження. Актуальність теми. Приклад використання ІАД. Постановка задачі. Засоби проведення експерименту. Вхідні данні. Метод ліктя. Реалізація k-means алгоритму. Завдання кольорів для кожного сегменту. Початковий графік даних. Згенеровані п'ять кластерів. Згенеровані три кластери. Висновки.

6. Консультанти розділів роботи

| Найменування розділу | Консультант (посада, прізвище, ім'я, по батькові) | Позначка консультанта про виконання розділу | |
|---|---|---|---|
| | | підпис | дата |
| *Розділи спеціальної частини* | *доц. Імангулова З.А.* | | 19.01.2024 |
| | | | |

**КАЛЕНДАРНИЙ ПЛАН**

| № | Назва етапів роботи | Терміни виконання етапів роботи | Примітка |
|---|---|---|---|
| *1* | *Отримання завдання на виконання роботи* | *16.10.2023* | |
| *2* | *Огляд критеріїв для дослідження методів інтелектуального аналізу даних* | *17-23.11.2023* | |
| *3* | *Огляд літератури* | *01-17.11.2023* | |
| *4* | *Опис математичної моделі методів інтелектуального аналізу даних* | *18-30.11.2023* | |
| *5* | *Проведення експерименту* | *29.12.2023* | |
| *6* | *Збір та аналіз даних експерименту* | *29.12.2023* | |
| *7* | *Оформлення пояснювальної записки* | *01-18.01.2024* | |
| *8* | *Представлення на рецензування* | *19.01.2024* | |

Дата видачі завдання 16.10.2023 р.

Студент _____ *Семенцова А.М.*
(підпис)

Керівник роботи _____ *доц.Імангулова З.А.*
(підпис)

# РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи магістра містить: 98 с., 2 табл., 29 рис., 2 додатки, 19 джерел інформації.

ЕЛЕКТРОННА КОМЕРЦІЯ, ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, ІНФОРМАЦІЙНІ СИСТЕМИ, КЛАСИФІКАЦІЯ, КЛАСТЕРНИЙ АНАЛІЗ, МАСШТАБУВАННЯ, РЕГРЕСІЯ, K-MEANS

Об’єкт досліджень – процес замовлення клієнтами послуг готелю в інформаційної системи управління готельним бізнесом. Під час обробки замовлень система виконує збір та обробку даних клієнтів, наприклад їхні демографічні характеристики, інформацію про покупки, уподобання, звички, частоту використання певних послуг.

Предметом дослідження є методи інтелектуального аналізу даних, які можна застосувати для обробки даних про клієнтів готельного бізнесу.

Мета досліджень – пошук та аналіз інформації щодо існуючих готельних послуг, їх розповсюдженості та популярності, дослідження бажань клієнтів та їх замовлення, що дозволить виділити цільові групи клієнтів що готові використовувати готельні сервіси.

Методи дослідження – методи інтелектуального аналізу даних, такі як методи кластерного аналізу та класифікації та їх поєднання.

Сфера застосування – галузь надання готельних послуг.

# ABSTRACT

Explanatory note to the qualification work of the masters contains: 98 pages, 2 tables, 29 figures, 3 appendices, 19 sources of information.

CLASSIFICATION, CLUSTER ANALYSIS, E-COMMERCE, INFORMATION SYSTEM,INTELLEGE DATA ANALISYS, K-MEANS, REGRESSION, SCALING.

The object of the research is the process of ordering hotel services by clients in the information management system of the hotel business. During order processing, the system collects and processes customer data, such as their demographic characteristics, information about purchases, preferences, habits, and the frequency of using specific services.

The subject of the research is the methods of intelligent data analysis that can be applied to process data about clients in the hotel business.

The goal of the research is to search for and analyze information about existing hotel services, their prevalence and popularity, the study of customer preferences and their orders, which will allow identifying target customer groups willing to use hotel services.

Research methods include intelligent data analysis methods, such as cluster analysis and classification methods, and their combinations.

Application area: the field of providing hotel services.

**ЗМІСТ**

# ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

АІС – автоматизована інформаційна система;

БД – база даних;

ІАД – інтелектуальний аналіз даних;

ADR – Average Daily Rate;

AI – Artificial Intelligence (штучний інтелект);

CRS & CM – Central Reservation System and Channel Managers (комп’ютерна система бронювання та дистрибуції);

HTML – HyperText Markup Language (мова розмітки гіпертексту);

ІС – інформаційна система;

IOT – Internet of Things (інтернет речей);

ML – Machine Learning (машинне навчання);

PMS – Property Management System (система управління майном);

URL – Uniform Resource Locator (єдиний вказівник на ресурс).

# ВСТУП

Готельний бізнес є одним з напрямів сучасного бізнесу, що неперервно розвивається, змінюється та осучаснюється кожний рік. Саме ця галузь є частиною світової туристичної структури яка приносить великий прибуток. На сьогоднішній день існує чимало різноманітних готелів на будь-який смак та кишеню. Також проживання в готелі для багатьох гостей є не тільки тимчасовим місцем мешкання, але і можливістю комфортно відпочити.

У минулому замовлення обслуговування номерів у готелі було предметом розкоші. Проте з часом це стало джерелом ресурсів для власників бізнесу. Середнє замовлення з телефону займає три хвилини і вимагає більше персоналу з більш завантажених готелів. Програмні технології для гостей готелю можуть вирішити ці проблеми, автоматизуючи весь процес бронювання послуг.

Перш ніж переходити до деталей, важливо зазначити, що, хоча прибуток від обслуговування номерів знизився, він все ще становить близько 1000 доларів США за номер, а галузеві звіти показують, що мобільні рішення можуть збільшити кошик замовлень на цілих 20%.

Замовлення сервісів в номер зможе зробити будь-який відпочинок набагато комфортнішим та зручнішим для гостя. Але досі мало які готелі змогли перейти на автоматизовані рішення заказів сервісів та частково позбутися застарілого методу заказу «віч-на-віч» чи за допомогою телефонного дзвінка.

Але зробити застосунок без доволі глибокого аналізу даних неможливо, бо без цінних уявлень, прогнозування тенденцій та прийняття важливих рішень на основі обробки великих обсягів даних прикладна програма може мати доволі великі недоліки в користуванні.

Розвиток методів запису і зберігання даних привів до бурхливого зростання обсягів збираної та аналізованої інформації. Обсяги даних настільки великі, що людина просто невзмозі проаналізувати їх самостійно, хоча необхідність проведення такого аналізу цілком очевидна, адже в цих “сирих даних” є знання, які можуть бути використані для прийняття рішень [1].

Алгоритми традиційної математичної статистики тривалий час, як основні, підтримували концепцію усереднення з вибірки, що зводиться до операцій над фіктивними величинами (типу середньої температури аудиторій в усіх приміщеннях університету, середньої висоти будинку міста, що складається з палаців і халуп тощо). Методи математичної статистики виявилися корисними переважно для перевірки заздалегідь сформульованих гіпотез (verification-driven data mining) і для "грубого" розвідницького аналізу, що становить основу оперативної аналітичної обробки даних (online analytical processing, OLAP).

Традиційна математична статистика ще довгий час претендувала на роль основного інструменту аналізу даних, не відповідала проблемам, що виникали. Тому виникла необхідність у розвитку нових сучасних методологій обробки та аналізу даних. Такою новою методологією і став інтелектуальний аналіз даних ІАД. Причини популярності ІАД такі:

– стрімке накопичення даних (рахунок йде вже на экзабайти);

– загальна комп'ютеризація бізнес-процесів;

– проникнення Інтернет в усі сфери діяльності; • прогрес в області інформаційних технологій: вдосконалення СУБД і сховищ даних; прогрес в області виробничих технологій: стрімке зростання продуктивності комп'ютерів, обсягів накопичувачів, впровадження Grid систем.

Алгоритми, що використовуються в ІАД, вимагають великої кількості обчислень. Раніше це було стримувальним чинником широкого практичного застосування ІАД, проте сьогоднішнє зростання продуктивності сучасних процесорів зняло гостроту цієї проблеми. Тепер за прийнятний час можна провести якісний аналіз сотень тисяч і мільйонів записів. ІАД – міждисциплінарна галузь, що виникла і розвивалася на основі таких наук, як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних тощо [2].

Саме за допомогою інтелектуального аналізу даних можна дізнатися певні шаблони та тренди сучасного готельного бізнесу послуг, прогнози на фоні теперішніх рішень інших готелів. В результаті можлива оптимізація рішень (наприклад, процесів надання послуг, системи обліку, маркетинг та інше) та

найголовніше – виявлення цільової аудиторії послуг готелю. Ці методи є запорукою підвищення загальної ефективності бізнесу та виявлення можливостей для покращення різних аспектів обслуговування клієнтів.

В кваліфікаційній роботі розглядається задача дослідження методів інтелектуального аналізу даних для виявлення цільових груп клієнтів у готельному бізнесі.

Актуальним рішенням задачі сегментації клієнтів буде провести аналіз даних клієнтів, виділити критерії сегментації, виконати кластерний аналіз а також класифікацію та зведення до суттєвих ознак. Таким чином можна отримати більш точні та деталізовані профілі груп клієнтів готельного бізнесу, що в свою чергу допомагає вдосконалити маркетингові стратегії та персоналізовані підходи до клієнтів.

Об’єктом досліджень є процес замовлення клієнтами послуг готелю в інформаційної системи управління готельним бізнесом. Під час обробки замовлень система виконує збір та обробку даних клієнтів, наприклад їхні демографічні характеристики, інформацію про покупки, уподобання, звички, частоту використання певних послуг. Застосування методів аналізу отриманих даних а також інтерпретація отриманих результатів дозволять визначити групи цільової аудиторії, які мають певні спільні риси, потреби або поведінку.

Предметом дослідження є методи інтелектуального аналізу даних, які можна застосувати для обробки даних про клієнтів готельного бізнесу.

Методи дослідження – методи кластерного аналізу та класифікації.

Результати даної роботи доповідалися і обговорювалися на ІІ Міжнародній науковій конференції «Період трансформаційних процесів в світовій науці: задачі та виклики» (Кривий ріг, 2024) [3].

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

## 1.1 Опис сучасного стану розвитку інформаційних систем

Спочатку треба розглянути, що представляє собою готель. Готель - це установа, яка забезпечує тимчасове проживання та різноманітні послуги для подорожуючих, які мають різну мету (бізнес-подорожі, відпочинок тощо). Вони можуть мати різний рівень комфорту, розміщення та сервісу, від економ-готелів до розкішних курортів. Готелі можуть надавати різноманітні послуги, такі як проживання, харчування, конференц-зали, спортивні заклади, розважальні програми, спа-центри та інші додаткові сервіси для задоволення потреб своїх гостей.

У 2001 р. один з провідних провайдерів в електронному резервуванні готелів UTELL за участю як готелів, так і турагентів розробив власну систему категоризації готелів, яку запропонував до використання чотирьом найбільшим GDS (Amadeus, Galileo, Worldspan, Sabre). Система поділяє готелі на три основних категорії за комфортом:

– luxury (Superior Deluxe, Deluxe, Moderate Deluxe згідно класифікації OHG) – найвищий стандарт якості, що відповідає 5 зіркам або 4 зірки плюс;

– superior – підвищена якість послуг для споживачів середнього класу – 4-3 зірки плюс (Superior First Class, First Class, Limited Service First Class за OHG);

– value – економічний клас – від скромних 3-зіркових до 1-зіркових готелів з обмеженим сервісом за помірну ціну (за OHG – Moderate First Class, Superior Tourist Class, Tourist Class, Moderate Tourist Class).

Крім цього UTELL пропонує класифікувати готелі також за призначенням відповідно до їх ніші на ринку: Style – готелі з власним іміджом (історичні будинки або оригінальний проект); Resort – курортні готелі для відпочинку; Apartment – апартготелі довгострокового проживання; Airport – транзитні готелі для пасажирів, розташовані в аеропорту [4].

Готельні послуги можуть включати номери різних категорій (одномісні, двомісні, люкс), ресторани, бари, басейни, тренажерні зали, конференц-зали для зборів та інші зручності, що залежать від рівня комфорту самого готелю. Крім того, готелі можуть надавати різні пакети послуг для відпочинку, бізнес-подорожей чи спеціальних заходів.

В готелях регулюються послуги за допомогою певних пунктів Правил № 19. Наприклад, готель може надавати (п. 1.3 Правил № 19):

– основні послуги – обсяг послуг готелю (проживання, харчування тощо), що входить до ціни номера (місця) і який надають споживачу згідно з укладеним договором;

– додаткові послуги – обсяг послуг, що не належать до основних послуг готелю та які споживач замовляє й оплачує додатково за окремим договором.

Готель самостійно визначає перелік і ціну основних (які входять до ціни номера) та додаткових послуг (п. 3.5 Правил № 19). Водночас така інформація повинна бути в кожному номері готелю (п. 2.3 Правил № 19).

Відповідно до Класифікатора ДК 009:2010 готельні послуги відповідають групі 55.10 "Діяльність готелів і подібних засобів тимчасового розміщування". Цей вид діяльності має бути зазначено у Єдиному державному реєстрі юридичних осіб, фізичних осіб – підприємців та громадських формувань [5].

Готельна галузь є важливою складовою туристичної індустрії та займає значне місце в господарському розвитку багатьох країн. Вона сприяє залученню туристів, створює робочі місця та сприяє розвитку інфраструктури у регіонах з підвищеним туристичним потенціалом.

Сучасні інформаційні системи у сфері готельних послуг використовуються для поліпшення обслуговування клієнтів, автоматизації процесів управління готелями та оптимізації бізнес-процесів. Ось деякі основні тенденції:

– системи управління готелем (PMS): це програмні засоби для автоматизації бронювання, обліку клієнтів, керування номерами та послугами готелю. Вони дозволяють контролювати резервування, фіксувати платежі, управляти інвентарем номерів тощо;

– системи бронювання та дистрибуції (CRS & CM): ці системи дозволяють готелям ефективно розподіляти свої номери через онлайн-канали бронювання та підтримувати актуальні дані про доступність та ціни;

– інтернет-технології та мобільні додатки: готелі активно використовують Інтернет для маркетингу, продажу та збільшення зручностей для клієнтів, а також розробляють мобільні додатки для бронювання та зв'язку з гостями;

– аналітика та бізнес-інтелект: готелі використовують системи аналітики для збору та аналізу даних про клієнтів, що допомагає приймати стратегічні рішення та покращувати обслуговування;

– інтеграція технологій "інтернету речей" (IoT): використання IoT дозволяє підключати різні пристрої та системи у готелях для автоматизації контролю за комфортом гостей, енергоефективності тощо.

Ці інформаційні системи сприяють ефективному функціонуванню готельних бізнесів, покращенню обслуговування клієнтів, оптимізації бізнес-процесів та підвищенню конкурентоспроможності у сучасному ринковому середовищі.

Сфера застосування ІАД нічим не обмежена – вона скрізь, де є якісь дані. Але насамперед методи ІАД сьогодні зацікавили комерційні підприємства, що розгортають свої проекти на основі інформаційних сховищ даних (Data Warehousing). ІАД являють собою велику цінність для керівників і аналітиків у їх повсякденній діяльності. Ділові люди усвідомили, що за допомогою методів ІАД вони можуть одержати відчутні переваги у конкурентній боротьбі.

Досвід багатьох підприємств показує, що віддача від використання ІАД може сягати 1000 %. Наприклад, відомі повідомлення про економічний ефект, що в 10 – 70 разів перевищив первісні витрати від 350 до 750 тис. дол. Відома інформація про проект у 20 млн. дол., що окупився усього за 4 місяці. Інший приклад – річна економія 700 тис. дол. за рахунок впровадження ІАД в мережі універсамів Великобританії. Нижче розглянуто сучасні системи, в основу яких покладений ІАД [6].

Кожна нова сучасна інформаційна система не може створюватися та розвиватися без використання інтелектуального аналізу даних хоча б через те, що маркетинговий, економічний та соціальний потенціал без цього падає. Без інформаційного аналітичного дослідження будь-який застосунок може зустрітися з такими ризиками:

– недостатня ефективність прийняття рішень (ускладняються процеси прийняття стратегічних рішень через відсутність об'єктивних даних та уявлень);

– втрата можливостей для оптимізації процесів, якщо упущено використання даних для виявлення проблемних аспектів або областей оптимізації, як результат - втрата ефективності та конкурентоспроможності;

– недостатня персоналізація та адаптація до потреб користувачів;

– недооцінка конкурентного середовища (неврахування потенційних зовнішніх загроз та можливостей) [3].

Отже, кожна сучасна інформаційна система величезною мірою залежить від методів інтелектуального аналізу даних через необхідність автоматизованого прийняття рішень, прогнозування та передбачення майбутніх тенденцій, виявлення ризиків та підозрілих змін, персоналізації потреб користувачів та загальної оптимізації бізнес-процесів.

## 1.2 Аналіз застосування досліджуваних методів в існуючих системах

Під час пошуку різних варіацій вже існуючих автоматизованих систем готельних сервісів з використанням методів інтелектуального аналізу даних були зроблені висновки щодо доцільності та актуальності даного дослідження.

На прикладі застосунку "Data Governance"(рис. 1.1) від розробників Alation ми можемо побачити робочий варіант повного управління даними в одному застосунку. Функції активного керування даними в Alation допомагають забезпечити належне використання даних. Людей навчають використовувати дані відповідно до вимог законодавства та навчають у робочому процесі політикам щодо даних, які вони обробляють [7].

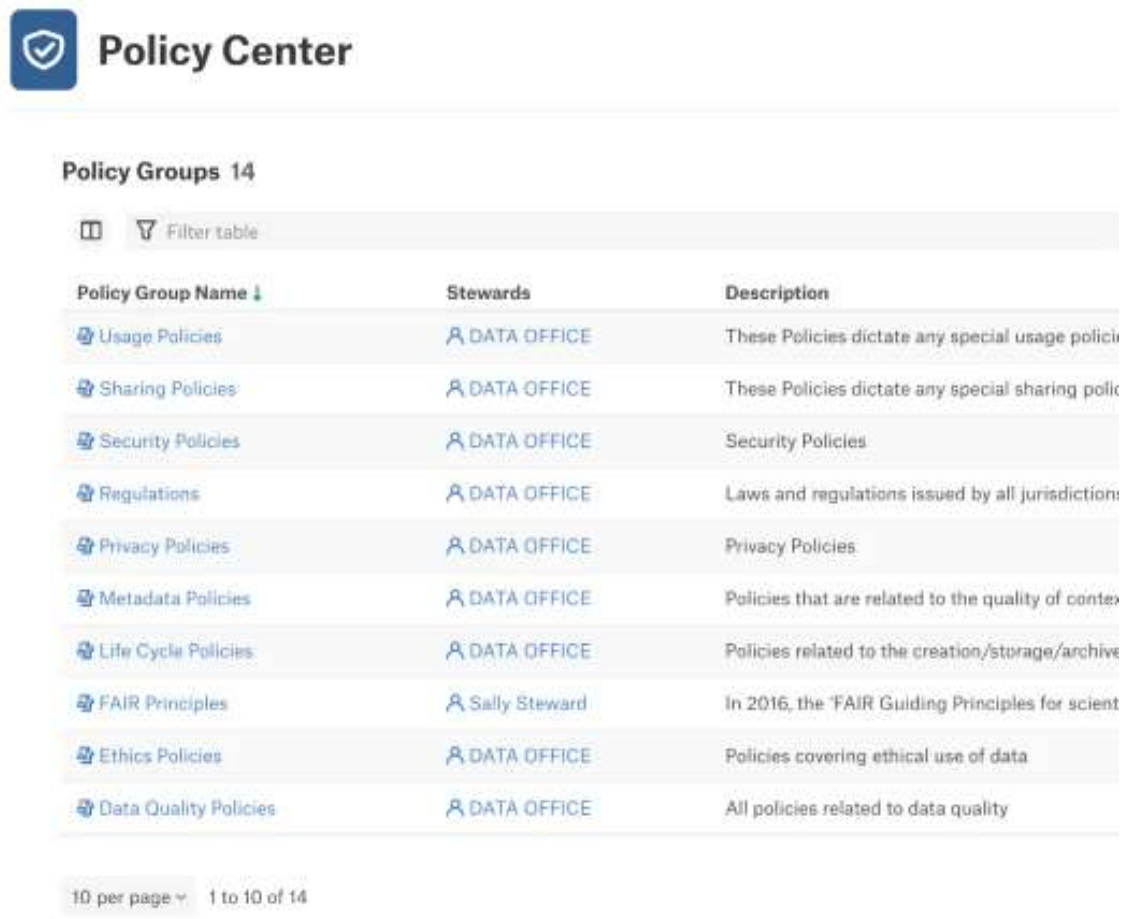


Рисунок 1.1 – Сторінка з центром керування політиками компанії

Даний застосунок пропонує центр керування політиками компанії у онлайн форматі який в реальному часі керує ризиками та перевіряє процеси на відповідність до умов.

Також, данний застосунок не обходиться без використання ML та AI-технологій. Також, даний застосунок пропонує робочий стіл для управління даними за допомогою класифікації даних та призначення політик даних. Data Governance пропонує допомогу у створенні та заповненні власного робочого столу за допомогою бізнес-глосарію Alation, який централізує створення термінів, таких як метрики та інші важливі визначення.

На рисунку 1.2 показана сторінка з робочим столом застосунку Data Governance, на якому ми бачимо доступ до різних дата сетів компанії, різноманітні інсайти та аналітики, сформовані алгоритмами даного застосунку тощо.

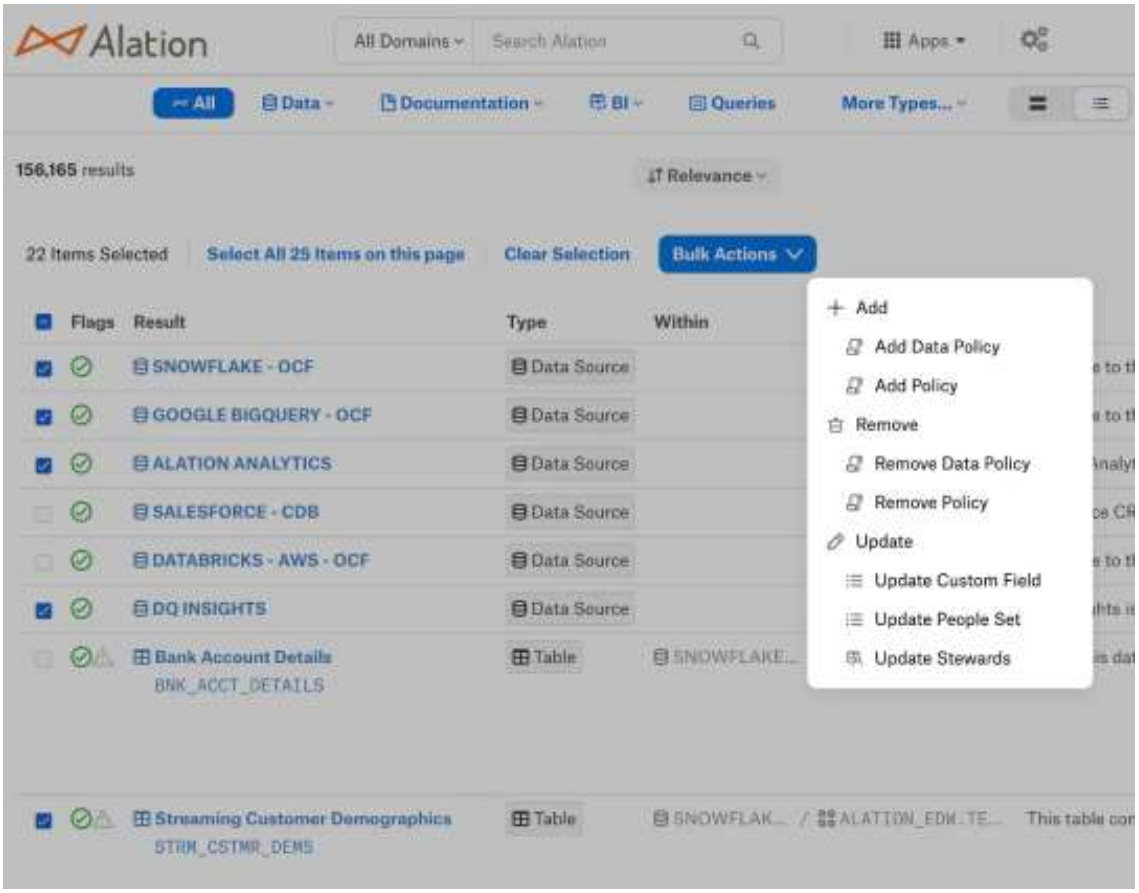


Рисунок 1.2 – Сторінка з робочим столом управління даними

Використовуючи методи інтелектуального аналізу даних дана прикладна програма вміє вимірювати вплив зусиль у сфері управління відносно поставлених цілей за допомогою візуалізації даних на сторінці інструментів управління (наприклад, графіки, діаграми). Інструмент управління надає керівникам можливість спостерігати за метриками прогресу кураторства для визначення того, чи досягаються поставлені цілі (рис. 1.3).

Успішне використання даного застосунку підтвердили Virgin Australia Group. З ростом компанії зростала і її IT-інфраструктура. Історично бізнес-підрозділи вирішували свої зростаючі технологічні потреби самостійно. Це призвело до сильно відокремленої архітектури та комунікаційних прогалин між напрямками бізнесу. Через це, керівники отримували суперечливу інформацію з різних сфер бізнесу, що ускладнювало прийняття рішень.


Рисунок 1.3 – Сторінка з інструментами управління

Команда Data Platforms почала шукати потрібний інструмент для задоволення потреб Virgin Australia, однак багато варіантів, які вони розглядали, вимагали занадто багато технічних знань, тому вони обрали просте та доступне рішення від Alation. До переходу до нового та сучасного застосунку Data Governance компанія займалася управлінням на індивідуальній основі, використовуючи різноманітні інструменти та функції програмного забезпечення для керування даними. Тепер Alation підтримує нову структуру, застосовуючи стандарти, політики та глосарії до даних у точці споживання. У результаті кожен, хто використовує дані у Virgin Australia, працюватиме на основі загального набору визначень, а їхній доступ до даних регулюватиметься ретельно розробленим набором політик [8].

Передивившись деякі існуючі варіанти рішення автоматизованих інформаційних систем для внутрішньо готельних сервісів були обрані для ретельного вивчення такі приклади як мобільний застосунок «Guest Service.App» (рисунок 1.4) та веб-додаток Hotefy (рисунок 1.5). В кожному з цих додатків є свої переваги та недоліки, вони мають свої обмеження та створені для різноманітних вимог щодо цінової політики та об’єму сервісів [9].

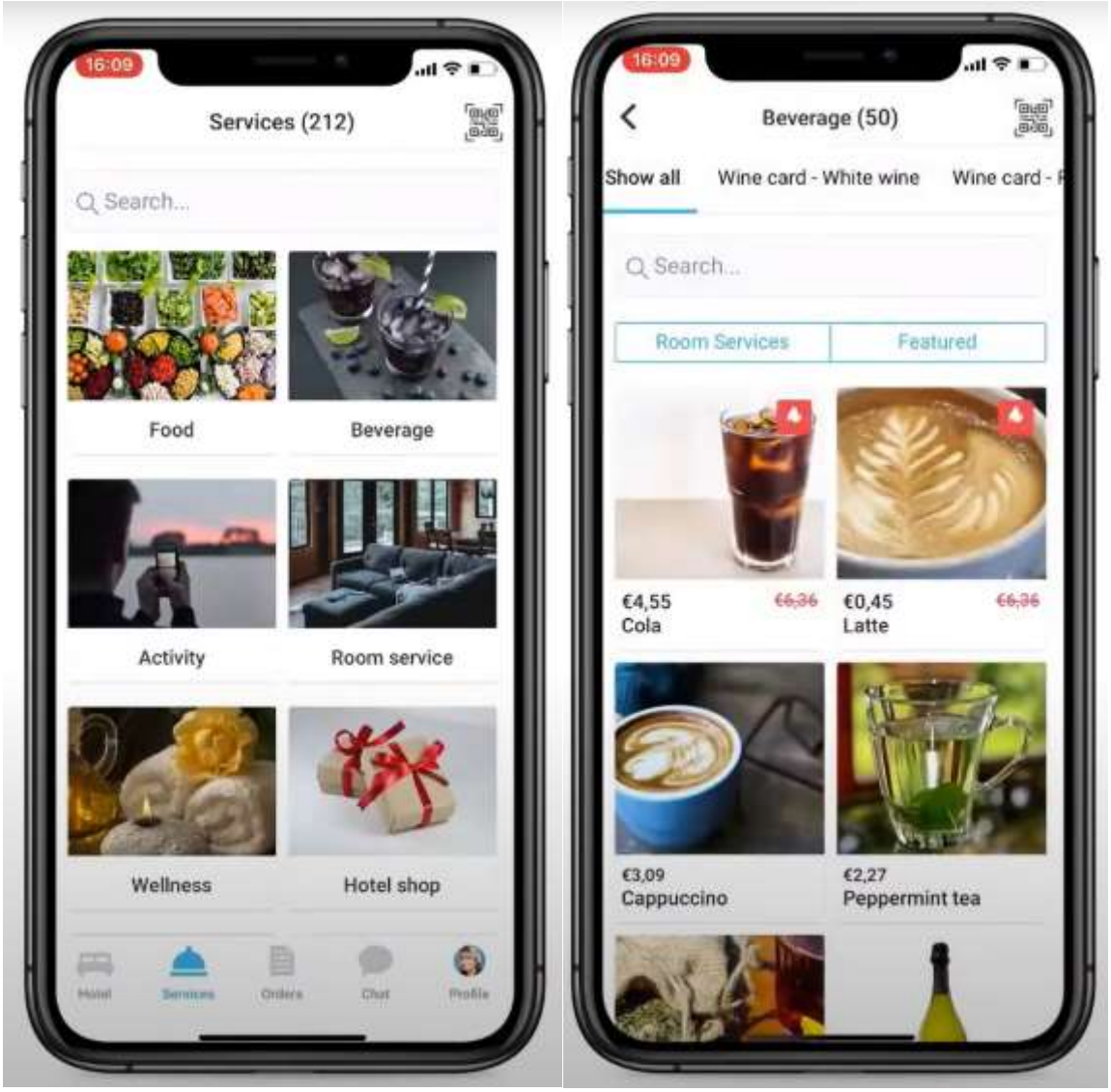


Рисунок 1.4 – Скріншот мобільного додатку Hotel Guest Service

Мобільний додаток Hotel Guest Service націлений в першу чергу на полегшення взаємодії з готелями для людей, що багато подорожують. Вони можуть розміщувати замовлення, планувати відправлення та навіть платити безпосередньо

зі своїх смартфонів до прибуття в якості гостя. Додаток зручно поділяє свої послуги на «їжу», «напої», «активності» тощо. Гості можуть додавати послуги до своїх кошиків і встановлювати час і дату, коли вони хочуть, щоб їхні запити були доставлені.

Цей гостьовий додаток дозволяє отримувати оновлення з підтвердженням замовлень за допомогою сповіщень без необхідності розмовляти з менеджером – розробники стверджують, що готелі можуть покращити показники продажів без залучення персоналу. За кілька натискань на смартфоні гості можуть повечеряти у своїх номерах. Додаток навіть дозволяє готелю пропонувати індивідуальні ціни для своїх постійних клієнтів.

Застосунок Hotel Guest Service має спеціальну систему бонусів для постійних клієнтів у барі зі спеціальними знижками для завзятих кавоманів та любителів легких та прохолодниз літніх коктейлів. Для залучення нових клієнтів має спеціальні персоналізовані знижки що змінюються кожну неділю в залежності від сезону, святковий день в країні де знаходиться готель, спеціальне меню та багато іншого.

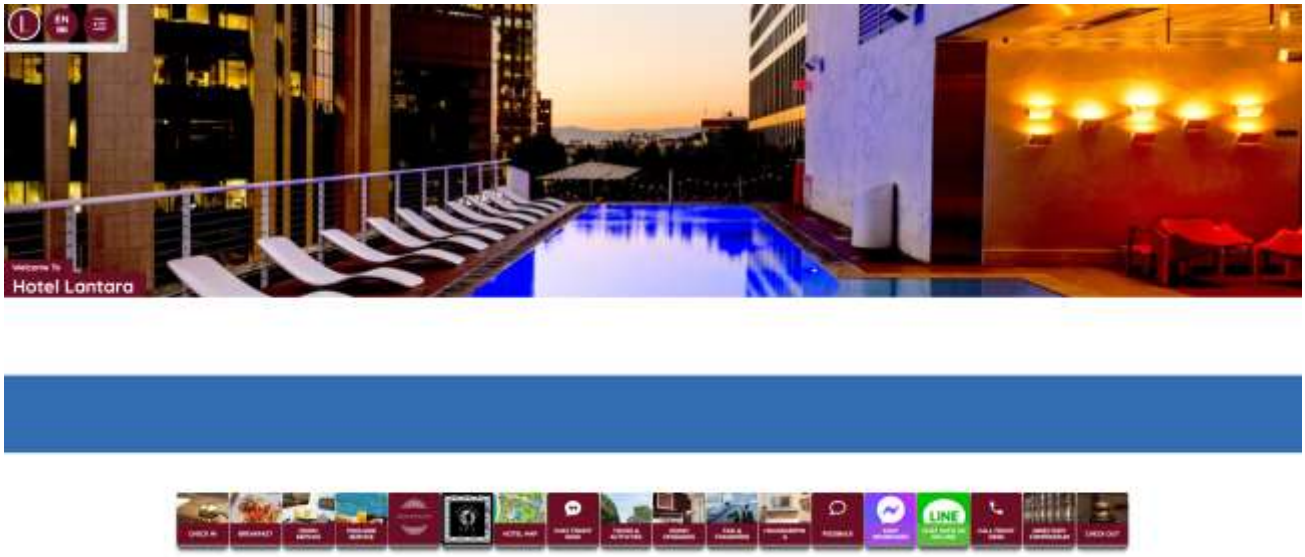


Рисунок 1.5 – Скріншот веб-додатку Hotefy

Hotefy розроблено як “додаток-консьєрж”, який може виконувати запити гостей зі смартфона. Він був створений з урахуванням обмежених бюджетів, з

якими можуть працювати готелі, і дозволяє надавати додаток для обслуговування номерів у найкоротші терміни. Раніше програма допомогла збільшити свої доходи на 30% декільком готелям і курортам .

Розробники програми Hotefy чутливо ставилися до проблем гостей, коли справа доходить до таких речей, як чистота. Запити на прибирання доступні лише одним дотиком. Крім того, готелі можуть значно заощадити на витратах, оскільки їм не потрібно вкладати кошти в персонал для перевірок зон готелю. Загалом, залучення Hotefy значно підвищило рейтинг готелів, що використовують цей додаток.

## 1.3 Актуальність досліджуваних методів

Актуальність інформаційних систем у готельній галузі з використанням методів інтелектуального аналізу даних полягає в їхній здатності досягати більшої ефективності та точності у виявленні клієнтських потреб, оптимізації обслуговування та управління готельним бізнесом.

Аналіз даних дозволяє створювати персоналізовані підходи до клієнтів, враховуючи їхні уподобання, історію проживання та покупок, що робить обслуговування більш ефективним та приємним. Також, дані методи дозволяють прогнозувати попит на певні послуги готелю, а також оптимізувати ціноутворення на основі реальних даних та ринкових тенденцій. Управління запасами та ресурсами, зменшення витрат та оптимізація ефективності, розуміння поведінки клієнтів та трендів – все це є досягненням використання методів інтелектуального аналізу даних.

Такі системи допомагають готелям реагувати на зміни у попиті, бути більш конкурентоспроможними, забезпечуючи високий рівень обслуговування та задоволення потреб клієнтів.

Необхідно провести детальне дослідження методів інтелектуального аналізу даних для виявлення цільової аудиторії послуг готелів.

# 2 ПОСТАНОВКА ЗАДАЧІ

## 2.1 Постановка задачі на дослідження

Дослідити та виявити цільові групи клієнтів готельного бізнесу, що використовують стандартний застосунок готельних послуг. Ми повинні зібрати, проаналізувати та дослідити дані та характеристики клієнтів і їх покупок. Інтерпретувати отримані результати для визначення груп цільової аудиторії, які мають спільні риси, потреби, побажання та поведінку. Ідентифікувати цільову аудиторію користувачів готельних сервісів.

### 2.1.1 Опис об’єкта дослідження

Об'єкт дослідження в цьому контексті – це сам процес застосування методів інтелектуального аналізу даних для виявлення цільових груп клієнтів у готельному бізнесі. Тобто, сам процес аналізу, методи та стратегії, що використовуються для виявлення характеристик, попиту, звичок чи вподобань клієнтів - усе це становить об'єкт дослідження.

Такий аналіз включає в себе збір та обробку даних про клієнтів готельного бізнесу, застосування різних методів аналізу та моделювання, а також інтерпретацію отриманих результатів для визначення цільових аудиторій, які мають певні спільні риси, потреби або поведінку.

Таким чином, об'єктом дослідження є сам процес та методи виявлення цільових груп клієнтів у готельному бізнесі за допомогою інтелектуального аналізу даних.

### 2.1.2 Вхідні та вихідні дані

Потрібні для дослідження дані можуть бути зібрані з різних джерел, таких як бази даних готелів, інтернет-ресурсів, соціальних мереж, опитувань, рейтингів

та інших джерел інформації, і використовуватись для подальшого аналізу та виявлення цільової аудиторії готельного бізнесу.

Для дослідження методів інтелектуального аналізу даних для виявлення цільової аудиторії клієнтів у готельному бізнесі можуть використовуватись різноманітні вхідні дані, які можуть бути зібрані та оброблені, наприклад:

- дані про бронювання та перебування (інформація про історію бронювань клієнтів, типи номерів, тривалість перебування, частота візитів тощо);
- демографічні дані (вік, стать, місце проживання, професія та інші персональні дані);
- інформація про споживання послуг (використання готельних послуг, наприклад, ресторани, спа-центри, фітнес-зали, трансфери тощо);
- дані про відгуки та рейтинги (відгуки та оцінки клієнтів щодо їхнього перебування та послуг готелю);
- дані з соціальних мереж та веб-сайту (інформація з соціальних мереж, веб-сайту готелю, взаємодія з рекламою тощо);
- географічні дані (інформація про місцезнаходження клієнтів, можливі географічні показники);
- інші зовнішні дані (економічні чинники, події або сезонні впливи, які можуть впливати на попит на готельні послуги).

Вихідні дані:

- сегментація аудиторії;
- профілі цільових аудиторій;
- рекомендації для обслуговування та маркетингу;
- прогнозування та рекомендації для розвитку бізнесу.

Ці вихідні дані можуть бути використані для вдосконалення стратегій маркетингу, покращення обслуговування, планування розвитку готельного бізнесу та підвищення його конкурентоспроможності.

### 2.1.3 Загальні мета та критерії даного дослідження

Загальна мета дослідження методів інтелектуального аналізу даних для виявлення цільової аудиторії клієнтів готельного бізнесу полягає у вдосконаленні стратегій обслуговування та маркетингу шляхом глибокого розуміння та сегментації аудиторії. для визначення основних сегментів клієнтів на основі їхніх характеристик, поведінки та уподобань, розробки стратегій та інструментів для створення персоналізованих підходів до кожного сегменту клієнтів.

Також після використання інтелектуальних методів аналізу даних стає можливим покращення маркетингових стратегій застосунку за допомогою використання інсайдів щодо аудиторії для розробки потужних та просунутих маркетингових кампаній та утримання вже існуючих клієнтів, а також повна оптимізація послуг та пропозицій через виявлення бажань та потреб аудиторії для вдосконалення існуючих послуг та розробки нових пропозицій. Важливі критерії успішності даного дослідження що показані в таблиці 2.1.

Таблиця 2.1 – Критерії успішності дослідження

| Назва критерію | Опис |
|---|---|
| Точність сегментації | Здатність точно і чітко розділити аудиторію на різні сегменти за певними параметрами |
| Персоналізація і реактивність | Можливість створювати персоналізовані стратегії з урахуванням індивідуальних потреб різних сегментів |
| Ефективність маркетингу та залучення клієнтів | Підвищення конверсії та збільшення потоку клієнтів відповідно до розроблених стратегій |
| Збільшення задоволення клієнтів та відновлення зв'язку | Підвищення рівня задоволення клієнтів через персоналізований підхід та покращення комунікації |

Ці критерії визначаються для оцінки ефективності та успішності впровадження методів інтелектуального аналізу даних для виявлення цільової аудиторії у готельному бізнесі.

# 3 ОГЛЯД МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

## 3.1 Методи інтелектуального дослідження даних

До методів і алгоритмів ІАД належать такі: штучні нейроні мережі, дерева рішень, символьні правила, методи найближчого сусіда і к-найближчого сусіда, метод опорних векторів, байєсові мережі, лінійна регресія, кореляційно-регресійний аналіз; ієрархічні методи кластерного аналізу, неієрархічні методи кластерного аналізу, зокрема і алгоритми к-середніх і к-медіани; методи пошуку асоціативних правил, зокрема алгоритм Apriori; метод обмеженого перебору, еволюційне програмування і генетичні алгоритми, різноманітні методи візуалізації даних і безліч інших методів. Варто зазначити, що більшість методів ІАД була розроблена у межах теорії штучного інтелекту. Єдиної думки щодо того, які задачі необхідно зараховувати до ІАД, немає. Більшість авторитетних джерел перераховує такі: класифікація, кластеризація, прогнозування, асоціація, візуалізація, аналіз виявлення відхилень, оцінювання, аналіз зв'язків, підведення підсумків. Розглянемо деякі з них [2]:

– кластерний аналіз – метод, що групує клієнтів за схожими характеристиками. Він дозволяє ідентифікувати сегменти аудиторії, які мають спільні риси чи поведінку. Наприклад, використовуючи дані про покупки чи демографічні дані, можна створити групи клієнтів, які мають подібні смаки або споживчі звички;

– класифікація – метод, що використовується для призначення клієнтів до певних категорій на підставі їх характеристик. Наприклад, можна створити модель класифікації, яка визначає, чи належить клієнт до певної цільової групи на основі його покупок, відвідувань сайту або інших даних;

– асоціативні правила – метод, що допомагає виявити зв'язки між продуктами або послугами, які часто придбаються разом. Наприклад, якщо багато клієнтів купують певний продукт разом з іншим, це може вказувати на специфічні потреби або вподобання;

– регресійний аналіз – метод, що дозволяє прогнозувати поведінку клієнтів на основі історичних даних. Він допомагає зрозуміти, які фактори впливають на певні відповіді чи дії клієнтів.

Ці методи можуть використовуватися окремо або в поєднанні один з одним для отримання більш точних та комплексних результатів при визначенні цільових груп клієнтів.

### 3.1.1 Методи кластерного аналізу

Кластерний аналіз – це інтелектуальний метод аналізу даних, спрямований на групування схожих об'єктів у кластери або групи. Основна ідея полягає у виявленні природних зв'язків або патернів у наборі даних, щоб об'єкти всередині кожного кластеру були якомога схожі між собою, а об'єкти різних кластерів – якомога більш відмінними.

Спочатку визначаються мета та цільові показники кластеризації. Це може бути сегментація аудиторії, виявлення патернів або групування подібних об'єктів. Наступним кроком обираються змінні, за якими будуть вимірювати схожість об'єктів.

Для вимірювання схожості можуть використовуватися різні метрики, такі як відстань Евкліда, косинусна схожість, коефіцієнт кореляції та інші. Потім обирається серед існуючих методів кластерізації той, що більше підійде для дослідження, а саме: метод k-середніх, ієрархічна кластеризація, DBSCAN та інші. Кожен метод має свої переваги та обмеження, і вибір залежить від специфіки даних та мети аналізу.

В кінці ми отримуємо дані, які можна кластеризувати, і результати візуалізуються для аналізу. Потім кластери можна інтерпретувати, вивчати та використовувати для подальших досліджень чи прийняття рішень.

Важливо оцінити якість кластеризації, особливо якщо результати будуть використовуватися для прийняття рішень. Це може включати внутрішні методи та зовнішні методи.

Кластерний аналіз використовується для виявлення структури в даних і отримання цінних інсайтів.

### 3.1.2 Методи класифікації

Класифікація – це інтелектуальний метод аналізу даних, який полягає у розподілі об'єктів на певні класи або категорії на основі їхніх характеристик. В результаті розв’язання задачі класифікації виявляються ознаки, які характеризують групи об'єктів досліджуваного набору даних – класи; за цими ознаками новий об'єкт можна зарахувати до того чи іншого класу. Для розв’язання задачі класифікації можуть використовуватися методи: найближчого сусіда (Nearest Neighbor); к-ближнього сусіда (к-Nearest Neighbor); байєсових мереж (Bayesian Networks); індукції дерев рішень; нейронних мереж (neural networks) [2].

Конкретні кроки класифікації показані в таблиці 3.1

Таблиця 3.1 – Кроки методу класифікації

| № | Кроки класифікації | Опис кроків |
|---|---|---|
| 1 | Підготовка даних | Збір, очищення та підготовка даних для аналізу. Це включає в себе вибір характеристик, які будуть використовуватися для класифікації, а також розподіл даних на навчальний та тестовий набори |
| 2 | Вибір моделі класифікації | Обирається підхід або алгоритм для побудови моделі. Це може бути, наприклад, метод дерева рішень, метод k-найближчих сусідів (k-NN), наївний баєсівський класифікатор, логістична регресія, або інші |

Продовження таблиці 3.1

| № | Кроки класифікації | Опис кроків |
|---|---|---|
| 3 | Тренування моделі | Використання навчального набору даних для навчання моделі. Модель виробляється та оптимізується для класифікації на основі вхідних даних та їх класів |
| 4 | Оцінка моделі | Використання тестового набору для оцінки точності та ефективності моделі. Це включає оцінку метрик, таких як точність (accuracy), чутливість (recall), специфічність (specificity) і F1-міра |
| 5 | Використання моделі для класифікації нових даних | Після успішного тренування та оцінки моделі вона може бути використана для класифікації нових об'єктів або прийняття рішень на основі нових даних |

Класифікація застосовується в багатьох галузях, таких як медицина (наприклад, діагностика за симптомами), фінанси (виявлення шахраїв), рекомендаційні системи (фільтрація контенту для користувачів) та інші для різних цілей, таких як передбачення, класифікація чи прийняття рішень.

### 3.1.3 Методи асоціативних правил

Асоціація (Associations). У процесі розв’язання задачі пошуку асоціативних правил відшукуються закономірності між зв'язаними подіями в наборі даних. Відмінність асоціації від двох попередніх задач ІАД: пошук закономірностей здійснюється не на основі властивостей об'єкта, що аналізується, а між кількома подіями, які відбуваються одночасно. Найвідоміший алгоритм розв’язку задачі пошуку асоціативних правил – алгоритм Apriori [2].

Процес виявлення асоціативних правил умовно розбито на такі кроки:

- пошук наборів елементів, що часто зустрічаються, також відомих як item sets (визначення наборів елементів, які з'являються разом в даних частіше, ніж певний поріг підтримки);
- побудова правил асоціацій (пари чи більші набори, які задовольняють задані пороги підтримки та цікавості, перетворюються в асоціативні правила. Наприклад, якщо в даних певні продукти часто купують разом, то може бути сформоване правило типу "Купуючи А, часто купують В";
- визначення і використання порогів (встановлення порогів підтримки та цікавості для відбору найбільш цікавих асоціативних правил);
- оцінка та використання правил (оцінка знайдених асоціативних правил та їх використання для прийняття рішень або отримання нової інформації).

Ключові поняття які потрібно розуміти під час використання асоціативних правил: підтримка (Support, ймовірність конкретного набору елементів, які можуть з’явитися разом у наборі даних), цікавість (Confidence, міра частоти елементу В виявитися в наборі даних, коли є елемент А) та підтримка асоціацій (Association Support, вимір підтримки для всієї асоціації або правила, яке містить два чи більше елементів).

Алгоритми побудови асоціативних правил вивчають можливі зв’язки чи комбінації об’єктів, подій, умов або висновків, оцінюють для них підтримку та формують правило прийняття рішень на цій основі [12].

### 3.1.4 Методи регресійного аналізу

Регресійний аналіз – це статистичний метод, який досліджує зв'язок між однією чи декількома незалежними (пояснюючими) змінними і залежною (пояснюваною) змінною. Основна мета регресійного аналізу – прогнозування значення залежної змінної на основі незалежних змінних. На рисунку 3.1 наведена узагальнена схема множинної лінійної регресії.

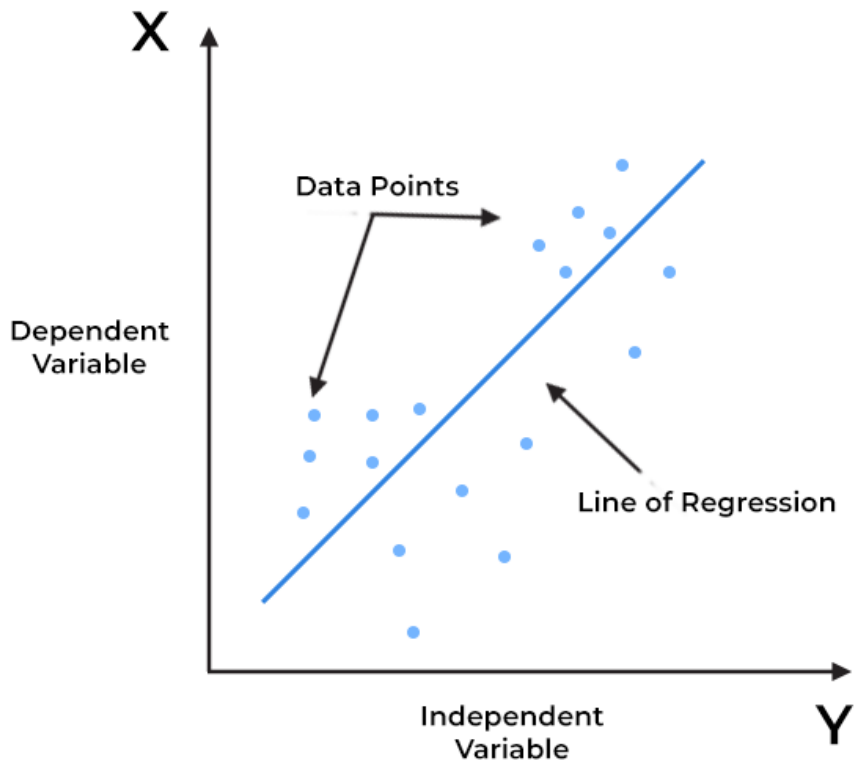


Рисунок 3.1 – Узагальнена множинна лінійна регресія

Регресія – один із найпоширеніших типів моделей машинного навчання, оцінює зв’язки між змінними. Тоді як моделі класифікації визначають, до якої категорії належить спостереження, регресійні моделі оцінюють числове значення [13].

Маємо певні типи регресійного аналізу, наприклад:

– проста лінійна регресія – модель, де залежна змінна залежить від однієї незалежної змінної, і зв'язок між ними описується лінійною функцією;

– множинна лінійна регресія – модель, де залежна змінна залежить від багатьох незалежних змінних;

– нелінійна регресія – модель, де залежність між змінними описується нелінійною функцією;

– логістична регресія – використовується для прогнозування ймовірності виникнення категоріальної залежної змінної.

Ключові кроки регресійного аналізу:

а) підготовка даних:

1) збір та очищення даних: включає в себе збір необхідних даних та видалення аномальних або відсутніх значень;

2) відбір змінних: визначення незалежних та залежних змінних для аналізу;

б) вибір типу регресії:

1) лінійна, множинна, нелінійна, логістична тощо: вибір конкретного типу регресії залежить від характеристик даних та цілей дослідження;

в) побудова моделі регресії:

1) підгонка моделі: використання алгоритмів для побудови математичної моделі, що описує зв'язок між змінними;

2) оцінка коефіцієнтів: оцінка параметрів моделі, таких як нахил та зсув лінії/площини;

г) оцінка моделі:

1) перевірка адекватності: оцінка того, наскільки добре модель відображає реальні дані;

2) статистичні тести: використання статистичних тестів для перевірки статистичної значущості коефіцієнтів та моделі в цілому;

д) прогнозування та інтерпретація:

1) прогнозування: використання моделі для прогнозування значень залежних змінних для нових даних;

2) інтерпретація: аналіз результатів, визначення впливу незалежних змінних на залежну змінну;

е) валідація та підтвердження:

1) перевірка: використання додаткових даних для перевірки точності моделі;

2) коригування: якщо потрібно, вносити корективи у модель для поліпшення результатів.

Регресійний аналіз застосовується у багатьох галузях, від економіки до медицини, для прогнозування, розуміння залежносте та виявлення факторів, які впливають на певні явища чи події.

### 3.1.5 Узагальнення методів для вирішення задачі сегментації цільових груп

Після ретельного вивчення інтелектуальних методів аналізу даних було обрано декілька методів для використання в даному дослідженні, а саме методи кластерного аналізу, класифікації та регресійного аналізу даних. Дані методи достатньо повно покривають потреби дослідження, а обрані методи ІАД дозволяють виконати завдання виявлення цільових груп користувачів у готельному бізнесі.

Окрема увага приділяється кластерному аналізу, який дозволяє згрупувати користувачів у сегменти на основі подібності у їх характеристиках (наприклад, за допомогою алгоритмів, таких як k-середніх або ієрархічна кластерізація). Методи класифікації та регресійного аналізу можуть бути використані для передбачення та класифікації вже нових даних на основі характеристик аудиторії.

## 3.2 Математичний опис методів інтелектуального аналізу даних

### 3.2.1 Математичний опис кластерного аналізу

Метод кластерного аналізу - це статистичний метод, що застосовується для групування схожих об'єктів разом у кластери на основі їх характеристик чи подібностей. Основна мета полягає в тому, щоб об'єкти всередині кожного кластера були між собою схожими, а об'єкти в різних кластерах були якнайменше схожими.

Математичні підходи до кластерного аналізу часто базуються на відстанях або подібностях між об'єктами даних, а також на певних параметрах (наприклад, кількості кластерів у K-середніх). Вибір конкретного методу залежить від природи

даних та цілей аналізу. Ось математична основа для певних методів кластерного аналізу:

- k-середніх (k-means) – найбільш поширений метод є K-середніх. Цей метод розділяє набір даних на K кластерів, де кожен кластер представляється своїм центром (центроїдом) (рис. 3.2). Центроїди розташовуються так, щоб мінімізувати внутрішній різницю кожного кластера та максимізувати відстань між кластерами. Математично, це відбувається шляхом мінімізації суми квадратів відстаней між кожним об'єктом та центроїдом свого кластера;

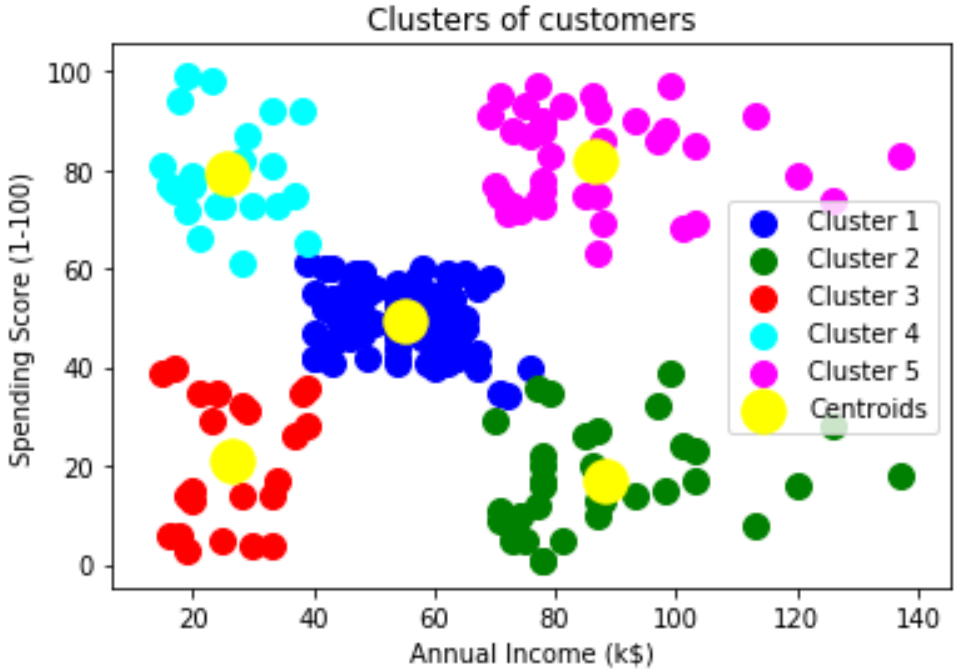


Рисунок 3.2 – Візуальний приклад розділення даних на кластери з позначенням центроїдів

– ієрархічна кластеризація – метод, що розділяє об'єкти на кластери в ієрархічній структурі. Він може бути агломеративним (злиття кластерів) або дівізійним (розділення кластерів). Для агломеративного підходу, кожен об'єкт спочатку вважається окремим кластером, а потім кластери зливаються разом на основі визначених відстаней (зазвичай евклідової відстані або кореляції) до тих пір, поки не утворять один кластер (рис.3.3);

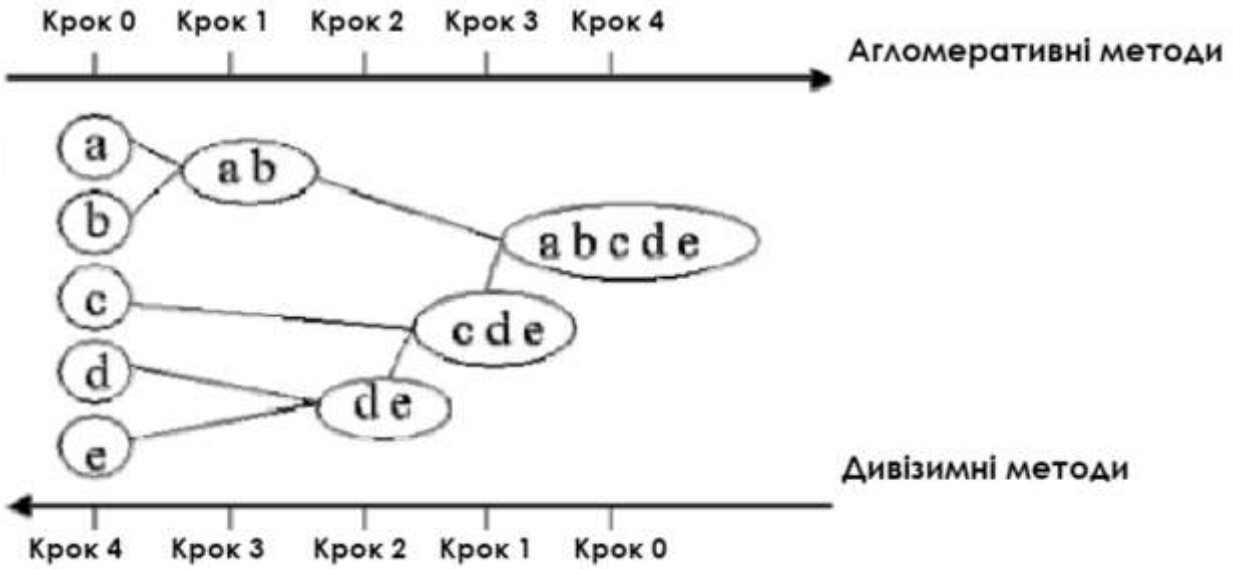


Рисунок 3.3 – Візуальне трактування роботи агломеративних та дивізимних алгоритмів

– DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – метод, що розділяє точки даних на кластери, використовуючи густину даних. Він визначає кластери як групи точок, які мають достатню густину, з'єднані між собою, і виділяє відокремлені точки як "шум" або "випадкові". На рисунку 3.4 відображено як даний метод справляється з шумом та випадковими точками.

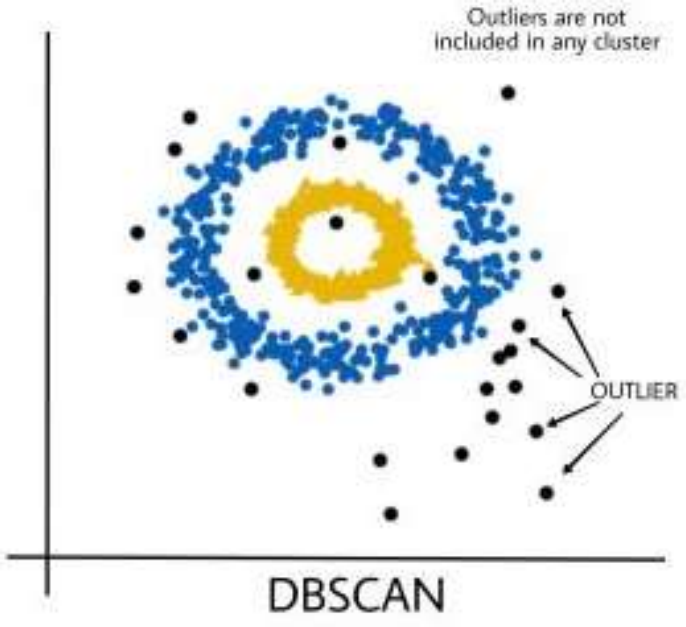


Рисунок 3.4 – Вплив викидів на DBSCAN

Опишемо кожен з методів окремо.

У випадку ієрархічної кластерізації маємо дві класифікації даних алгоритмів:

– агломеративні (характеризуються послідовним об'єднанням вихідних елементів та відповідним зменшенням числа кластерів (тобто, побудовою кластерів знизу вгору));

– дивізимні (характеризуються зростанням числа кластерів, починаючи з одного, в результаті чого утворюється послідовність розщеплюючих груп (тобто, побудовою кластерів зверху донизу))

Узагальнене уявлення агломеративного методу:

крок 1: уся множина об’єктів предметної області I уявляється як множина кластерів вигляду:

$$c_1 = \{i_1\}, c_2 = \{i_2\}, \dots, c_p = \{i_p\}, \dots, c_q = \{i_q\}, \dots, c_m = \{i_m\} \qquad (3.1)$$

крок 2: обираються два найбільш близькі один до одного кластери (наприклад $c_p$ та $c_q$) та поєднуються в один кластер, складається нова множина з m - 1 кластерів

$$c_1 = \{i_1\}, c_2 = \{i_2\}, \dots, c_p = \{i_p, i_q\}, \dots, c_m = \{i_m\} \qquad (3.2)$$

крок 3: повторення кроків 1 та 2 до тих пір, поки не буде сформований кластер, що складається з m об’єктів та співпадаючий з першопочатковою множиною об’єктів

$$d_{rs} = \alpha_p d_{ps} + \alpha_q d_{qs} + \beta d_{pq} + \gamma |d_{ps} - d_{qs}|, \qquad (3.3)$$

де $d_{rs}$ – відстань від нового кластеру $c_r$ як результат поєднання кластерів $c_p$ та $c_q$ до кластеру $c_s$;

$d_{ps}$ – відстань від кластеру $c_p$ до кластеру $c_s$;

$d_{qs}$ – відстань від кластеру $c_q$ до кластеру $c_s$; $d_{pq}$ – відстань поміж кластерами $c_p$ та $c_q$.

Головна особливість дивізимних методів це на кожному кроці дивізимних алгоритмів передбачається рекурсивне розподілення одного вихідного кластера на два дочірніх. Таке рекурсивне розподілення продовжується до тих пір, поки усі кластери не будуть складатися з одного об'єкта , або поки всі члени одного кластеру не будуть мати нульову відмінність однин від одного.

Основні кроки дивізимного алгоритму, що був запропонований у 1965 р. Смітом Макнаотоном:

крок 1: усі елементи множини I поміщаються в один кластер $C_1$;

крок 2: обирається елемент множини I, у якого середнє значення відстані від інших елементів в цьому кластері найбільше

$$D_{C_1} = {}^{1}/_{N_{C_1}} \times \sum_{p=1}^{m} \sum_{q=1}^{m} d\left(i_p, i_q\right) \forall i_p, i_q \in C_1, p \neq q; \qquad (3.4)$$

крок 3: обраний на кроці 2 елемент видаляється з кластеру $C_1$ та формує елементі другого кластеру $C_1$;

крок 4: обирається елемент кластеру $C_1$ , для якого різниця поміж середньою відстанню до елементів, що знаходяться у кластері $C_2$ , та середньою відстанню до елементів, що залишилися в кластері $C_1$ , є найбільшою;

крок 5: обраний на кроці 4 елемент переноситься у кластер $C_2$;

крок 6: повторювати виконані кроки 4 та 5 до тих пір, поки відповідні різниці середніх не стануть від'ємними(тобто, поки існують елементи, що прихильні до елементів кластеру $C_2$ ближче, ніж до елементів кластеру $C_1$). Потім завершити метод.

Один з варіантів розвитку дивізимного методу запропонований у 1990 році Кауфманом та Роузеувом метод вибору для розщеплення кластеру з найбільшим діаметром, який вичислюється за формулою:

$$D_C = \max d(i_p, i_q) \forall i_p, i_q \in C \quad (3.5)$$

Найбільш вигідно та доречно використовувати метод k-means, якщо данні за своєю природньою натурою розподіляються на компактні, приблизно сферичні групи. Даний алгоритм використовується для групування набору даних у певну кількість (k) кластерів на основі їх характеристик або ознак.

K-means ефективний для великих обсягів даних та звичайно застосовується в багатьох областях, таких як маркетингові дослідження, обробка зображень, біоінформатика та інше. Важливо пам'ятати, що вибір початкових центроїдів може вплинути на кінцевий результат, і деякі конфігурації даних можуть призвести до різних результів кластеризації.

Концептуальний опис алгоритму k-means:

крок 0. Вибирається до k довільних вихідних центрів - точок в просторі всіх об'єктів досліджуваних даних. Не критично, які це будуть центри – процедура їх вибору вплине, головним чином, лише на час розрахунку;

крок 1. Усі об'єкти розбиваються на k груп, найближчих до кожного з центрів. Близькість визначається відстанню, яка обчислюється одним з варіантів розрахунку відстані, наприклад, береться евклідова відстань:

$$d_2(x_i, x_j) = \sqrt{\sum_{t=1}^{m} (x_{it} - x_{jt})^2}, \quad (3.6)$$

де i, j – ідентифікатори об’єктів даних, які описують набором значень атрибутів $x_i, x_j$;

t = 1,m – ідентифікатор окремих атрибутів, що утворюють набори $x_i, x_j$.

крок 2. Обчислюються нові центри кластерів, наприклад, як середні значення змінних об'єктів, віднесених до сформованих груп;

крок 3. Повторення Кроків 1 і 2 до тих пір, поки центри кластерів не перестануть змінюватися.

Базова множина вихідних даних:

$$M = \{m_j\}_{j=1}^{d},$$

де d – кількість точок(векторів) даних.

Метрика відстані розраховується за формулою:

$$d_A^2(m_j, c^{(i)}) = \|m_j - c^{(i)}\|_A^2 = (m_j - c^{(i)})^t A(m_j - c^{(i)}). \quad (3.7)$$

Вектор центрів кластерів:

$$C = \{c^{(i)}\}_{i=1}^{c},$$

де

$$c^{(i)} = \frac{\sum_{j=1}^{d} u_{ij} m_j}{\sum_{j=1}^{d} u_{ij}}, 1 \le i \le c. \quad (3.8)$$

Матриця розбиття:

$$U = \{u_{ij}\},$$

де

$$u_{ij}^{(1)} = \begin{cases} 1 \text{ при } d(m_j, c_i^{(1)} = \min\limits_{1 \le k \le c} d(m_j, c_k^{(1)})) \\ 0 \text{ в інших випадках.} \end{cases} \quad (3.9)$$

Цільова функція:

$$J(M, U, C) = \sum_{i=1}^{c} \sum_{j=1}^{d} u_{ij} d_A^2(m_j, c^{(i)}) \quad (3.10)$$

Набір лімітів

$$u_{ij} \in \{0,1\};\ \sum_{i=1}^{c} u_{ij} = 1; 0 < \sum_{j=1}^{d} u_{ij} < d, \quad (3.11)$$

визначає, що кожен вектор даних може належати тільки одному кластеру та не належати іншим. В кожному кластері міститься не менше однієї точки, але й не менше загальної кількості точок.

крок 1: проініціалізувати початкове розбиття(наприклад, випадковим чином), обрати точність δ (використовується як умова завершення алгоритму), проініціалізувати номер ітерації l = 0;

крок 2: визначити центри кластерів за формулою:

$$c_1^{(i)} = \frac{\sum_{j=1}^{d} u_{ij}^{(1-1)} m_j}{\sum_{j=1}^{d} u_{ij}^{(1-1)}}, 1 \le i \le c. \quad (3.12)$$

крок 3: оновити матрицю розбиття з тим, щоб мінімізувати квадрати помилок, використовуючи формулу (3.8);

крок 4: перевірити умову $\left\| U^{(1)} - U^{(1-1)} \right\| < \delta$. Якщо умова виконується, завершити процедуру, якщо ні – перейти до другого кроку з номером ітерації 1=1+1.

# 4 ЕКСПЕРИМЕНТАЛЬНИЙ АНАЛІЗ КЛАСТЕРІВ СЕГМЕНТУ РИНКУ ДЛЯ ГОТЕЛЬНОГО БІЗНЕСУ

## 4.1 Узагальнена інформація про аналіз та його інструменти

Проведемо аналіз кластерів сегменту ринку для готельного бізнесу за допомогою консольного застосунку, написаним мовою програмування Jupyter Notebook, схожий на мову програмування Python, та візуалізацію графіків за допомогою застосунку Jupyter Notebook Output.

Jupyter Notebook – це інтерактивне середовище для програмування, яке дозволяє комбінувати код, текстовий контент, графіку та інші елементи в одному документі. "Output" в Jupyter Notebook вказує на результат виконання комірки коду або інших введених команд (рис. 4.1).

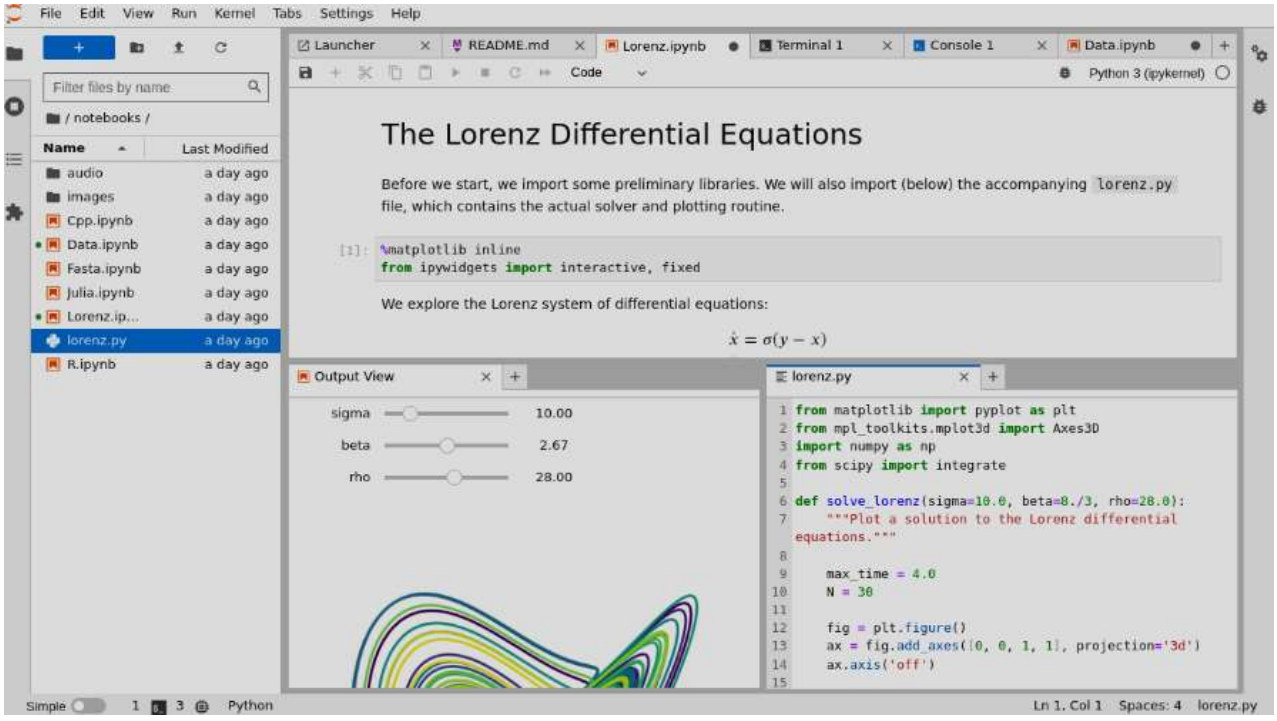


Рисунок 4.1 – Скріншот з програми Jupyter Notebook

Блокнот – це документ, яким можна ділитися, який поєднує в собі комп’ютерний код, описи простою мовою, дані, різноманітні візуалізації, як-от 3D-

моделі, діаграми, графіки та малюнки, а також інтерактивні елементи керування. Ноутбук разом із редактором (наприклад, JupyterLab) забезпечує швидке інтерактивне середовище для прототипування та пояснення коду, дослідження та візуалізації даних, а також обміну ідеями з іншими [17].

В залежності від того, що ви вводите в комірці коду, "Output" може приймати різні форми, наприклад, такі основні з них, як:

– текстовий вивід – за допомогою команди для виведення текстової інформації результат можна буде відображено прямо в Jupyter Notebook;

– графіки та зображення – з використанням бібліотеки для візуалізації даних (наприклад, Matplotlib, Seaborn), графіки та зображення також можуть відображатися безпосередньо в Output;

– табличний вивід – результати, представлені у вигляді таблиць (наприклад, використовуючи бібліотеку Pandas), також можуть бути виведені в Output.

Важливо відзначити, що результат виконання кожної комірки виводиться безпосередньо під нею в тому ж самому документі, що дозволяє легко комбінувати код, викладки та результати в єдиному файлі для подальшого аналізу та представлення.

Даний аналіз базується на основі оригінального дослідження Antonio, Almeida та Nunes [18].

## 4.2 Опис середовища розробки

IPython виконує дві основні ролі:

– термінал IPython, відомий як REPL;

– ядро IPython - IPykernel, яке забезпечує обчислення та зв’язок із зовнішніми інтерфейсами, такими як ноутбук.

Коли ви вводите ipython, ви отримуєте оригінальний інтерфейс IPython, який працює в терміналі.

Усі інші інтерфейси – Блокнот, консоль Qt, консоль ipython у терміналі та сторонні інтерфейси – використовують ядро IPython. IPykernel – це окремий процес, який відповідає за виконання коду користувача та такі речі, як обчислення можливих завершень. На рисунку 4.2 наведено огляд зв'язків проекту Jupiter.

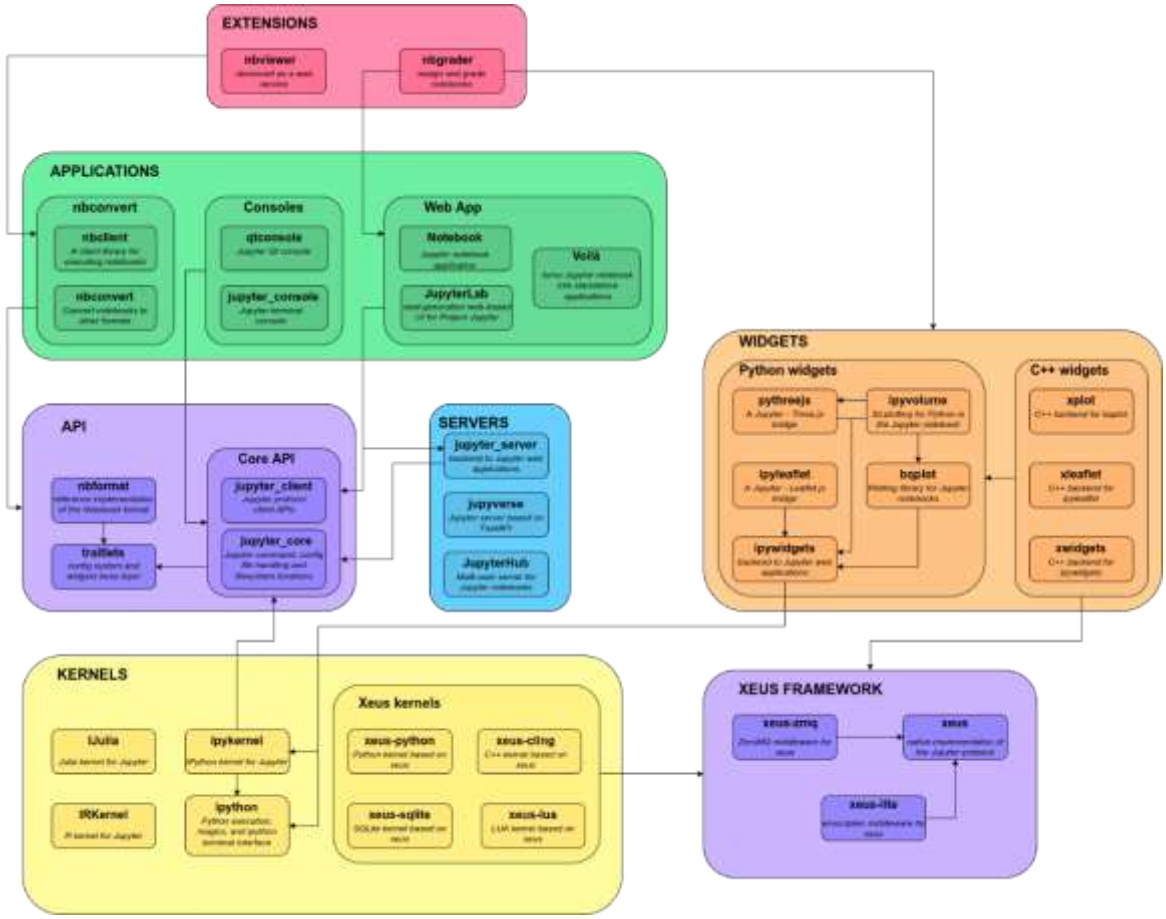


Рисунок 4.2 – Візуальний огляд зв'язків проекту Jupiter на високому рівні

Інтерфейси, такі як блокнот або консоль Qt, спілкуються з ядром IPython за допомогою повідомлень JSON, які надсилаються через сокети ZeroMQ; протокол, який використовується між інтерфейсами та ядром IPython, описано в Messaging в Jupyter. Основний механізм виконання для ядра використовується спільно з терміналом IPython (рис.4.3).

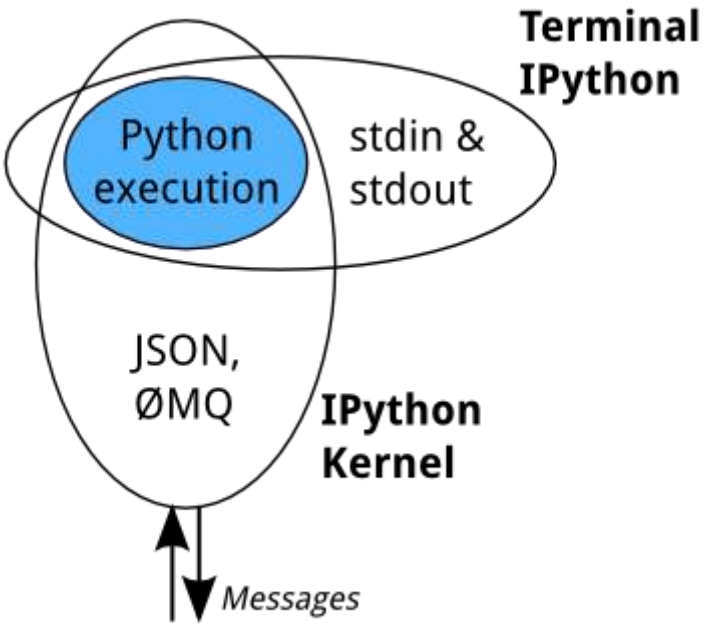


Рисунок 4.3 – Результат взаємодії IPykernel та Treminal Ipython

Процес ядра може бути підключений до кількох зовнішніх інтерфейсів одночасно. У цьому випадку різні інтерфейси матимуть доступ до однакових змінних.

Ця конструкція мала на меті полегшити розробку різних інтерфейсів на основі одного ядра, але вона також дозволила підтримувати нові мови в тих самих інтерфейсах, розробляючи ядра цими мовами, і ми вдосконалюємо IPython, щоб зробити це більш практичним. Сьогодні існує три способи розробки ядра для іншої мови (рис.4.4):

– ядра-оболонки повторно використовують механізм зв’язку з IPykernel і реалізують лише базову частину виконання;

– рідні ядра реалізують виконання та зв’язок цільовою мовою;

– ядра, засновані на xeus, нативній реалізації протоколу, реалізують мовну частину ядер. На відміну від підходу до оболонки, xeus не залежить від середовища виконання Python.

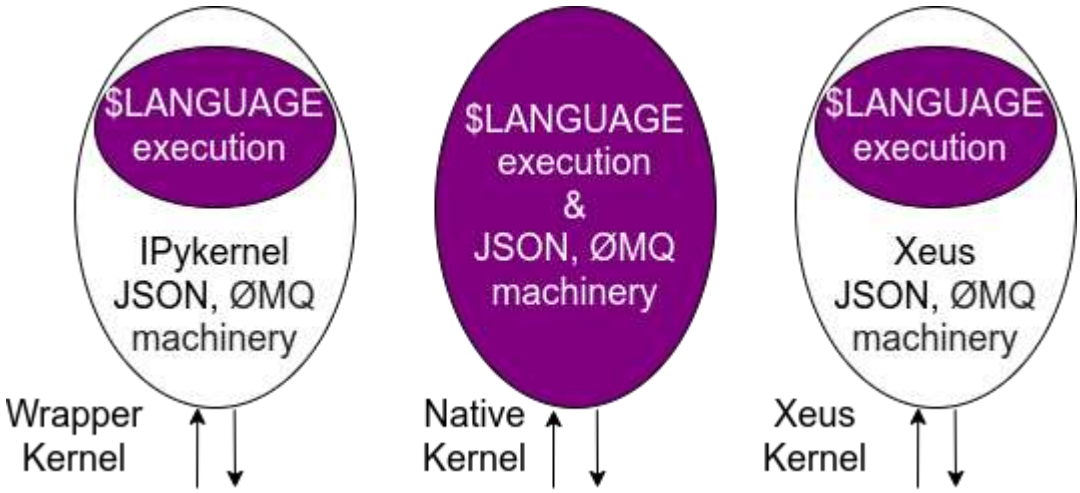


Рисунок 4.4 – Взаємодія трьох способів розробки ядра

Ядра-оболонки легше писати швидко для мов, які мають хороші оболонки Python, як-от octave_kernel, або для мов, де непрактично реалізувати механізм зв'язку, як-от bash_kernel. Рідні ядра, ймовірно, будуть краще підтримуватися спільнотою, яка їх використовує, наприклад IJulia або IHaskell. Ядра Xeus легко писати, коли інтерпретатор мови надає C++ або C API.

Блокноти Jupyter – це структуровані дані, які представляють ваш код, метадані, вміст і результати. Під час збереження на диск блокнот використовує розширення .ipynb і використовує структуру JSON.

Jupyter Notebook і його гнучкий інтерфейс розширює блокнот за межі коду до візуалізації, мультимедіа, співпраці тощо. На додаток до виконання вашого коду, він зберігає код і вихідні дані разом із примітками до розмітки в редагованому документі, який називається блокнот. Коли ви зберігаєте його, він надсилається з вашого браузера на сервер Jupyter, який зберігає його на диску як файл JSON із розширенням .ipynb.

Сервер Jupyter є комунікаційним центром. Браузер, файл блокнота на диску та ядро не можуть спілкуватися один з одним напряму. Вони спілкуються через сервер Jupyter. Сервер Jupyter, а не ядро, відповідає за збереження та завантаження блокнотів, тому можливо редагувати блокноти, навіть якщо у вас немає ядра для цієї мови – юзер просто не зможете запустити код. Ядро нічого не знає про

документ блокнота: йому просто надсилаються клітинки коду для виконання, коли користувач запускає їх (рис.4.5).

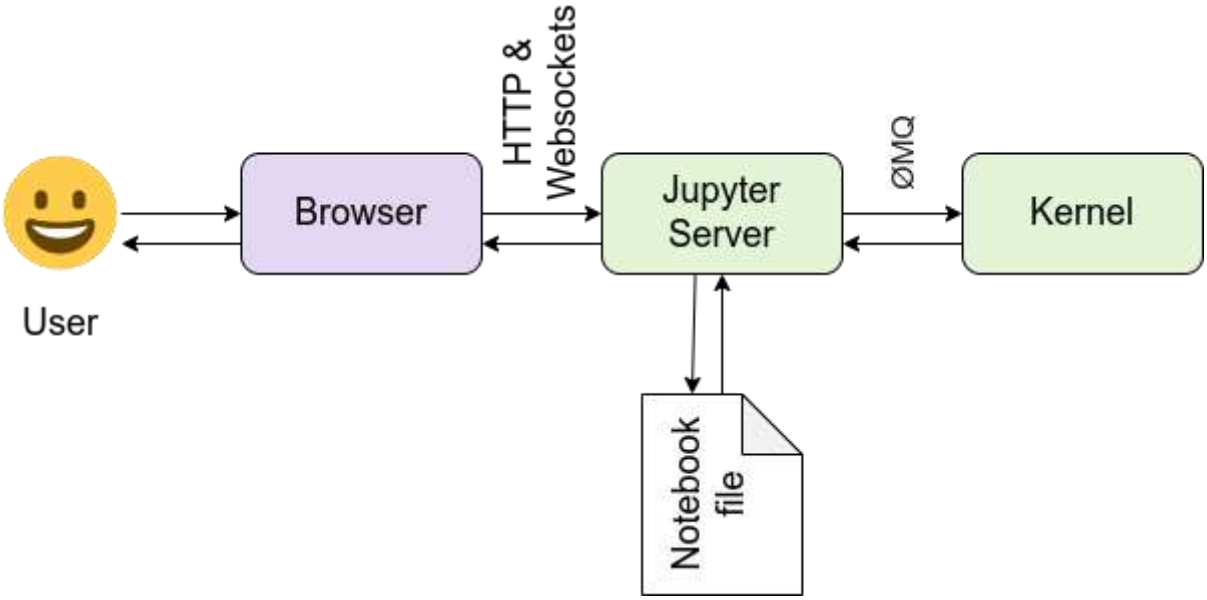


Рисунок 4.5 – Взаємодія користувача, браузера, сервера, файлів програми а також ядра

Інструмент Nbconvert у Jupyter перетворює файли блокнотів в інші формати, такі як HTML, LaTeX або reStructuredText. Це перетворення проходить через ряд кроків (рис.4.6):

– препроцесори змінюють блокнот у пам'яті. наприклад ExecutePreprocessor запускає код у блокноті та оновлює вихідні дані;

– експортер перетворює блокнот в інший формат файлу. Більшість експортерів використовують для цього шаблони;

– постпроцесори працюють з файлом, створеним під час експорту.

Веб-сайт nbviewer використовує nbconvert із експортером HTML. Коли користувач надає йому URL-адресу, він отримує блокнот із цієї URL-адреси, перетворює його на HTML і надає цей HTML користувачу [19].

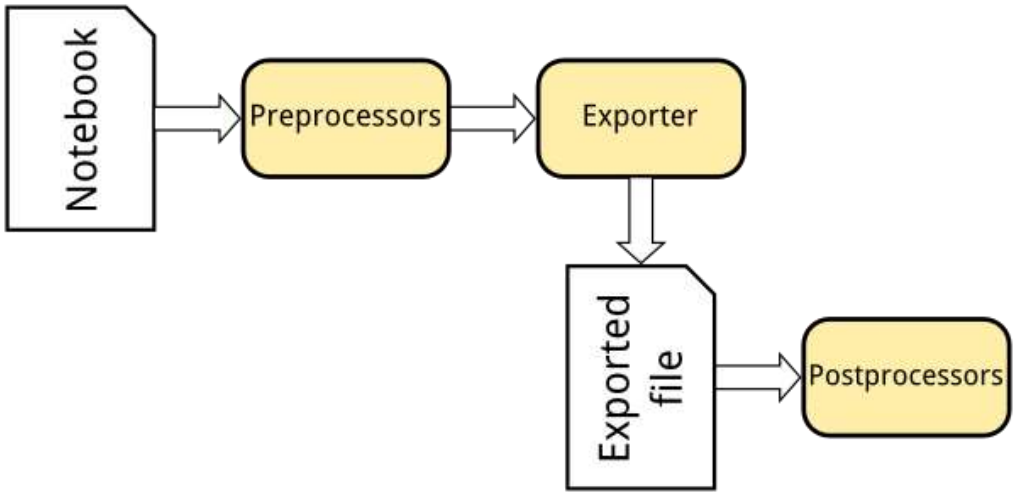


Рисунок 4.6 – Етапи перетворення файлів блокноту в інші формати за допомогою Nbconvert

## 4.3 Вхідні данні

Враховуючи передбачений час (період часу від моменту бронювання клієнтом до фактичного перебування в готелі, називаємо lead time), разом із середньоденною вартістю номеру на одного клієнта (ADR), алгоритм кластеризації k-means використовується для візуальної ідентифікації найбільш прибуткових сегментів ринку для готелю. Клієнт з високим ADR і низьким передбаченим часом є ідеальним, оскільки це означає, що:

– клієнт платить високу денну ставку, що призводить до більшої маржі прибутку для готелю;

– низький передбачений час означає, що клієнт сплачує за своє бронювання швидше – це збільшує потік готівки для зазначеного готелю.

## 4.4 Опис вхідних даних

Маємо більше ста тисяч рядків інформації про передбачений час прибування клієнту в готелі, рік, місяць, тиждень (як номер), день неділі заселення клієнтів,

кількість дорослих, дітей, малюків, замовлення їжі в номер і так далі. Ми вже визначили які параметри(стовпці) інформації нам потрібні (рис.4.7).

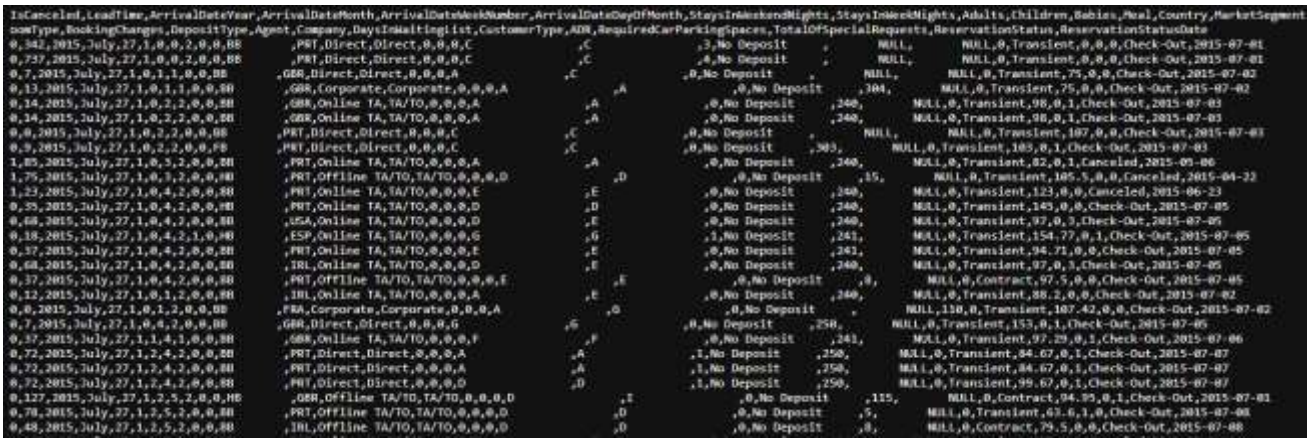


Рисунок 4.7 – Приклад консольного виводу бази даних готелю

Через велику кількість даних ми обираємо та завантажуємо 100 випадкових прикладів.

```
df = pd.read_csv('H1full.csv')
df = df.sample(n = 100)
```

Рисунок 4.8 – Код програми обрання 100 випадкових прикладів даних

Інтервал (або безперервні випадкові змінні) – це час виконання та ADR, визначені нижче:

```
leadtime = df['LeadTime']
adr = df['ADR']
```

Рисунок 4.9 – Безперервні випадкові змінні

Змінні з компонентом категорій визначаються за допомогою «cat.codes», у цьому випадку сегмент ринку.

```
marketsegmentcat=df.MarketSegment.astype("category").cat.codes
marketsegmentcat=pd.Series(marketsegmentcat)
```

Рисунок 4.10 – Змінні з компонентом категорій

Метою цього є присвоєння категорійних кодів кожному сегменту ринку. Наприклад, на рисунку 4.11 наведено фрагмент деяких записів сегментів ринку в наборі даних.

```
10871        Online TA
7752         Online TA
35566    Offline TA/TO
1353         Online TA
17532        Online TA
             ...
1312         Online TA
10364           Groups
16113           Direct
23633        Online TA
23406           Direct
```

Рисунок 4.11 – Фрагмент записів сегментів ринку

Після застосування «cat.codes» отримаємо відповідні категорії (рис.4.12).

Позначки сегментів ринку такі:

– 0 – Corporate;
– 1 – Direct;
– 2 – Groups;
– 3 – Offline TA/TO;
– 4 – Online TA.

```
10871    4
7752     4
35566    3
1353     4
17532    4
        ..
1312     4
10364    2
16113    1
23633    4
23406    1
```

Рисунок 4. 12 – Категорії після застосування «cat.codes»

Час виконання та функції ADR масштабуються за допомогою «sklearn» (рис.4.7).

```
from sklearn.preprocessing import scale
X = scale(x1)
```

Рисунок 4. 13 – Функція масштабування часу виконання та функції ADR

```
array([[ 1.07577693, -1.01441847],
       [-0.75329711,  2.25432473],
       [-0.60321924, -0.80994917],
       [-0.20926483,  0.26328418],
       [ 0.53174465, -0.40967609],
       [-0.82833604,  0.40156369],
       [-0.89399511, -1.01810593],
       [ 0.59740372,  1.40823851],
       [-0.89399511, -1.16560407],
```

Рисунок 4. 14 – Приклад зразка X

4.5 Кластерізація k-means.

Коли справа доходить до вибору кількості кластерів, одним із можливих рішень є використання так званого методу ліктя (рис. 4.15).

Це техніка, за допомогою якої обчислюється внутрішньокластерна дисперсія для кожного кластера – чим нижча дисперсія, тим щільніший кластер.

Цей метод спроектований для визначення точки, на якій зміна в кількості кластерів призводить до значного зниження змістовності моделі.

Основний принцип методу ліктя використовує ідею, що при збільшенні кількості кластерів сума квадратів відстаней в кожному кластері (внутрішня варіація) буде зменшуватися, але при певному моменті зменшення цієї суми стає менш значущим. Графічно, це може виглядати як "ліктьова точка" на графіку, яка вказує на оптимальну кількість кластерів.

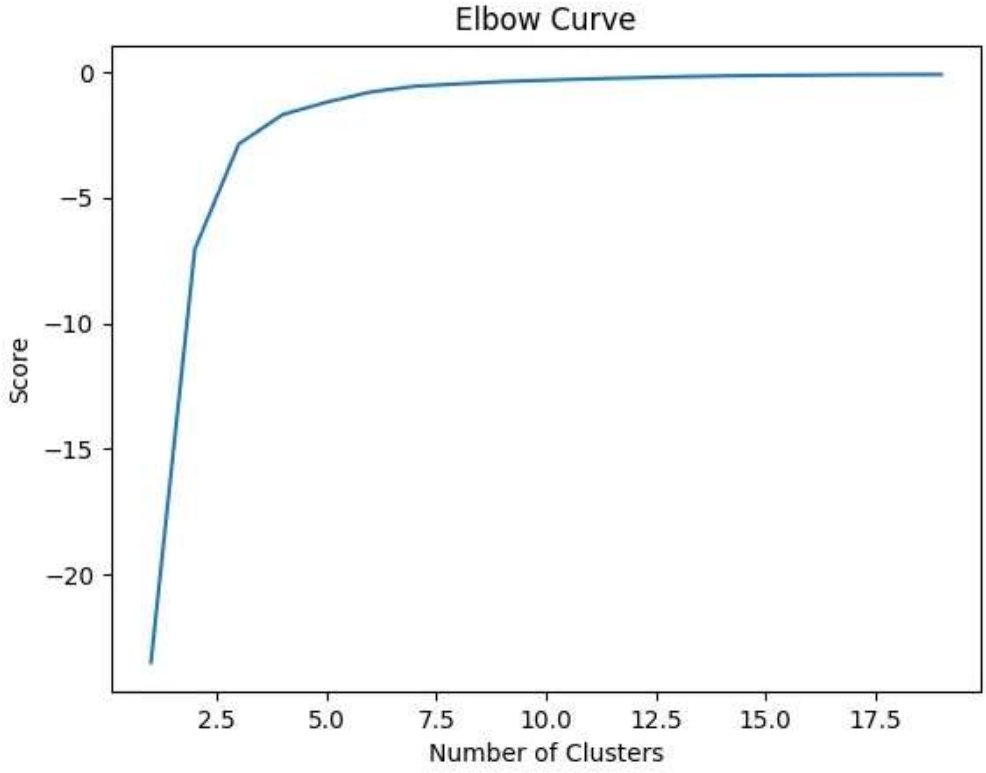


Рисунок 4. 15 – Приклад методу ліктя

Оптимальна кількість кластерів визначається як точка на графіку, де зменшення внутрішньої варіації стає менш значущим.

У зв’язку з цим, оскільки оцінка починає вирівнюватися, це означає, що зменшення дисперсії стає все меншим і меншим, оскільки ми збільшуємо кількість кластерів, що дозволяє нам визначити ідеальне значення для k.

Однак ця техніка не обов’язково підходить для невеликих кластерів. Крім того, ми вже знаємо кількість кластерів (k=5), які ми хочемо визначити, так само вже відома кількість сегментів ринку, які ми хочемо проаналізувати.

```
>>> km = KMeans(n_clusters = 5, n_jobs = None, random_state = None)
>>> km.fit(X)

KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=5, n_init=10, n_jobs=None,
precompute_distances='auto',
       random_state=None, tol=0.0001, verbose=0)
```

Рисунок 4.16 – Реалізація K-means алгоритму

Крім того, хоча методи кластеризації k-середніх можуть також використовувати PCA (або зменшення основної розмірності, Principal Dimensionality Reduction), щоб зменшити кількість функцій, це не підходить у цьому випадку, оскільки єдині дві функції, які використовуються (крім ринкового сегмента), це ADR і час виконання.

Код реалізації методу k-means наведено на рисунку 4.16.

```
# Market Segment Labels: 0 (Complementary) = firebrick, 1 (Corporate)
= dodgerblue, 2 (Direct) = forestgreen, 3 (Groups) = goldenrod, 4
(Offline TA/TO) = rebeccapurple

color_theme = np.array(['firebrick', 'dodgerblue', 'forestgreen',
'goldenrod', 'rebeccapurple'])
```

Рисунок 4.17 – Завдання кольорів для кожного сегменту

Наразі маємо такий графік без згенерованих кластерів за допомогою алгоритму k-середніх (рис. 4.18).

На даному графіку ми бачимо хаотично розкидані різнокольорові точки (лейбли), які на разі не створюють щільну та цільну картину кластерів. Наступним етапом буде генерування кластерів за допомогою k-means алгоритму(рис. 4.13).

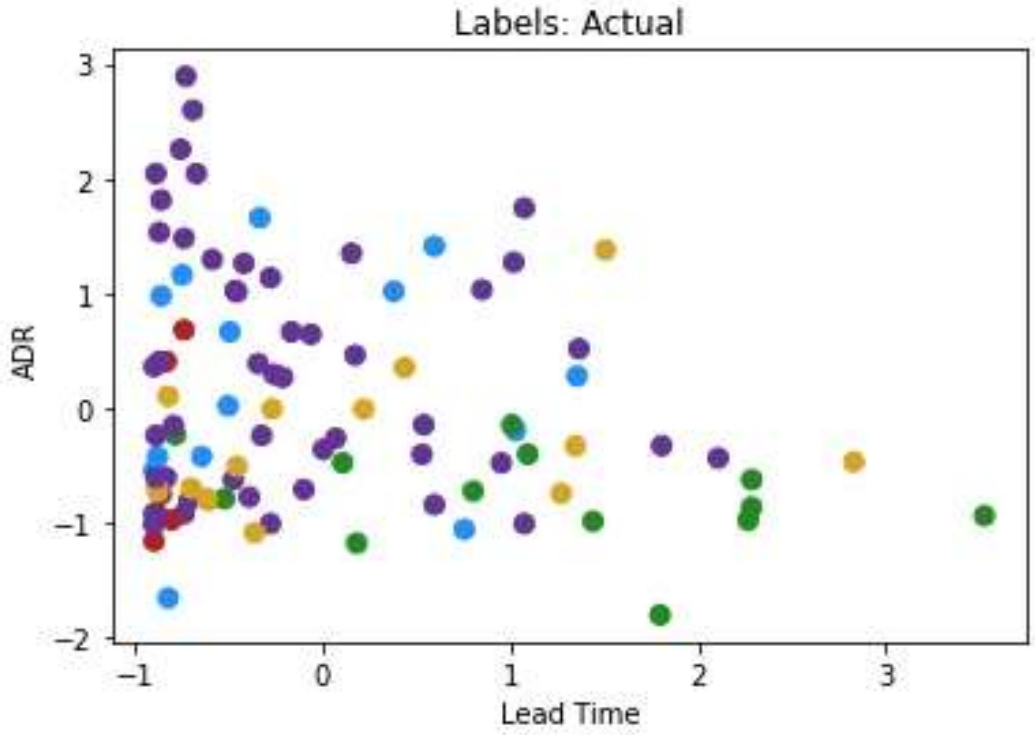


Рисунок 4.18 – Початковий графік даних

Як зазначалося, клієнти з найменшим часом виконання та найвищим ADR вважаються найбільш прибутковими. Наразі, маємо справу з кластером червоного кольору, який описує зазвичай клієнтів, що знімають номер екстрено чи для бізнес подорожей (конференції, відрядження тощо).

Однак, проблемою є те, що багато категорій сегментів ринку були неправильно позначені. Це поширена проблема під час роботи з кластеризацією k-середніх і не обов’язково означає, що модель слід переробити. Натомість це лише натякає на те, що нам потрібно по-іншому думати про наші дані.

Наприклад, ми вже знаємо, які клієнти належать до якого сегмента ринку. У цьому відношенні створення алгоритму кластеризації k-середніх для прогнозування цього ще раз не приносить великої користі.

Натомість сенс використання цього алгоритму полягає в тому, щоб отримати швидке уявлення про те, які типи клієнтів є найприбутковішими.

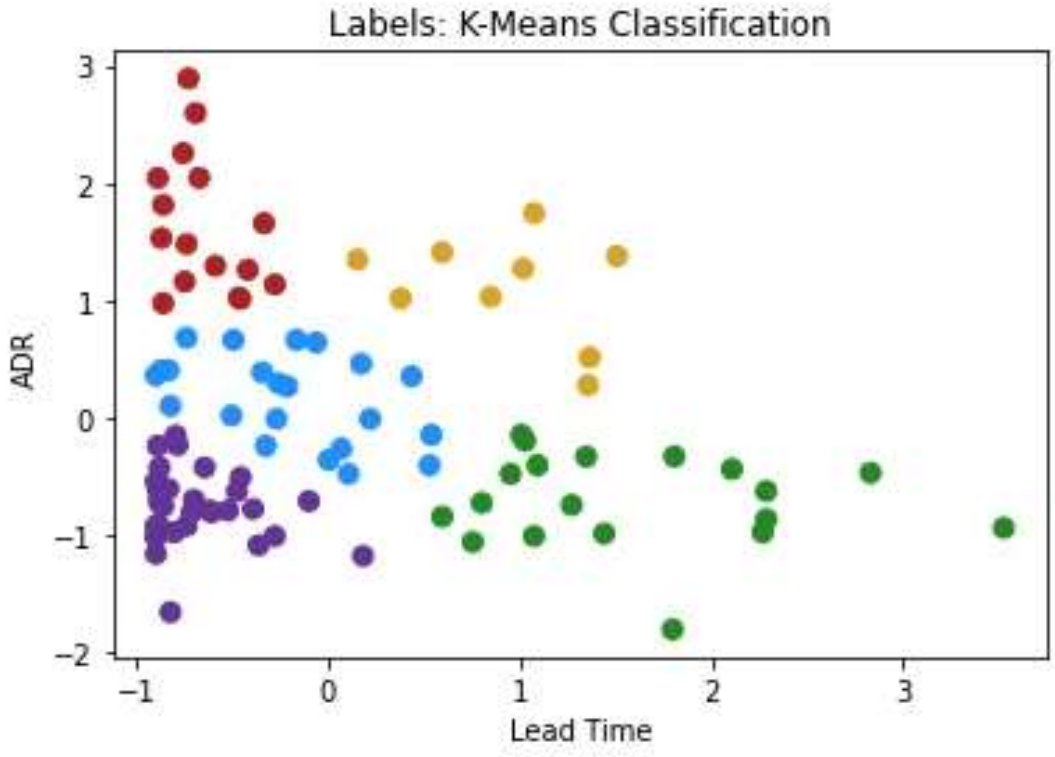


Рисунок 4.19 – Згенеровані п’ять кластерів

Крім того, було розглянуто лише час виконання та ADR як дві функції. Можуть існувати інші особливості, які були не враховані, які б краще вказували, до якого сегменту ринку може належати клієнт, і немає жодних візуальних доказів того, що ми бачили досі, що певні сегменти ринку є більш прибутковими, ніж інші.

У зв’язку з цим ще спростимо аналіз до використання трьох кластерів (рис. 4.20).

Можна помітити, що синя категорія має найвищий ADR і найменший час виконання (найвигідніша), тоді як зелена категорія показує найнижчий ADR і найвищий час виконання (найменш прибутковий).

З цієї точки зору, алгоритм кластеризації k-середніх пропонує ефективний спосіб швидкої класифікації найприбутковіших клієнтів готелю, а подальший аналіз можна провести для аналізу певних атрибутів, які є спільними для клієнтів у кожній групі.

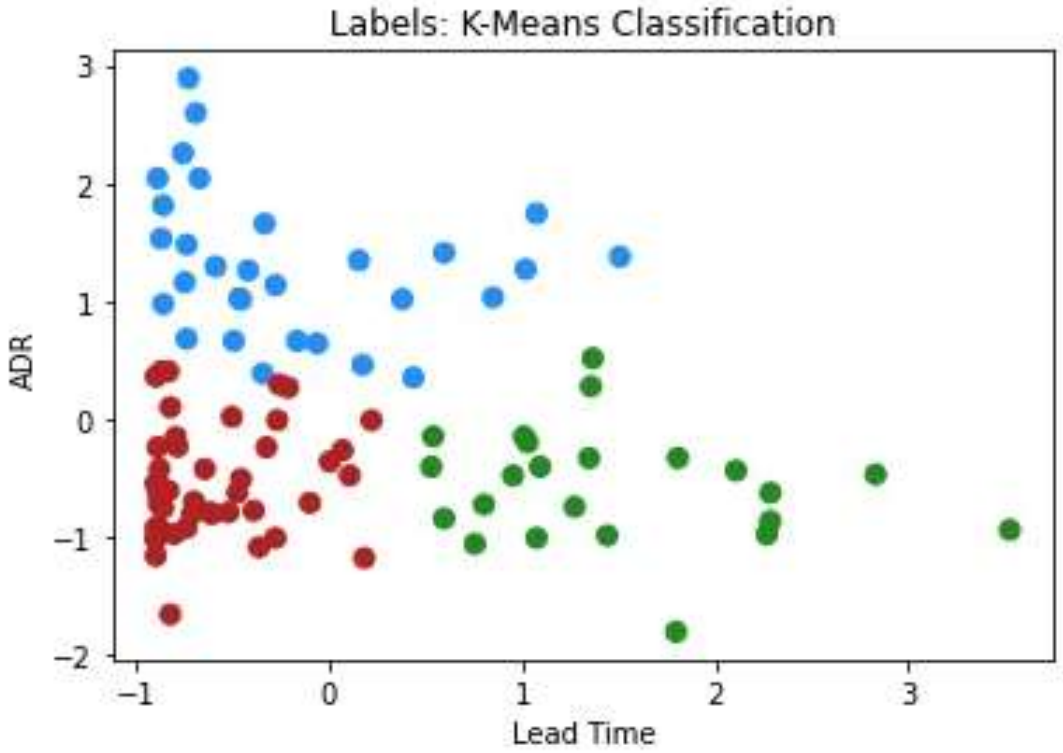


Рисунок 4.20 – Згенеровані три кластери

Коли мова заходить про неконтрольоване навчання – важливо пам’ятати, що це здебільшого дослідницький метод аналізу – метою не обов’язково є передбачення, а скоріше виявити інформацію про дані, які, можливо, не розглядалися раніше. Наприклад, чому певні клієнти мають нижчий час виконання замовлення, ніж інші? Чи клієнти з певних країн більше відповідають цьому профілю? А як щодо різних типів клієнтів?

На всі ці питання алгоритм кластеризації k-середніх може не відповісти прямо, але зведення даних в окремі кластери забезпечує міцну базу для того, щоб поставити такі питання.

# ВИСНОВКИ

В ході виконання кваліфікаційної роботи були досліджені методи інтелектуального аналізу даних, їх використання в сучасних проектах а також було досліджено інструменти візуалізації даних алгоритмів.

Перед початком роботи над використанням методів інтелектуального аналізу даних було детально розібрано всі аспекти предметної області які стосуються роботи готелів. Проведено опис сучасного стану розвитку інформаційних систем, проаналізовано також застосування досліджуваних методів у вже існуючих інформаційних системах. Особливу увагу було приділено процесам взаємодії клієнтів з готелем, а саме обліку замовлень клієнтів, а також процесам збору додаткових даних та інформації про клієнтів.

Під час постановки задачі визначили опис об’єкта дослідження, вхідні та вихідні дані та поставили мету та критерії успіху дослідження.

В ході дослідженнях та експериментів з методами кластерного аналізу та класифікації, визначили що за допомогою методів інтелектуального аналізу даних можна підняти рівень продукту, його економічну, соціальну та маркетингову ефективність.

Виявлення цільових груп клієнтів дозволяє удосконалити персоналізацію обслуговування. Аналізуючи дані про замовлення, інтелектуальна система може прогнозувати індивідуальні потреби гостей та рекомендувати персоналізовані послуги або акції. Це не тільки підвищує задоволення клієнтів, але і створює можливості для додаткового доходу для готелів через пропозицію додаткових послуг, які відповідають конкретним потребам окремих сегментів клієнтів.

Незважаючи на достатню кількість методів ІАД, пріоритет поступово зміщується у бік логічних алгоритмів пошуку в даних причинно-наслідкових правил. За їх допомогою розв’язуються задачі прогнозування, класифікації, розпізнавання образів, сегментації БД, здобування з даних “схованих” знань, інтерпретації даних, установлення асоціацій в БД тощо. Результати таких алгоритмів ефективні й легко інтерпретуються.

# ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ


1. Чубукова И.А. Data Mining: учебн. пособ. – М.: Интернет-университет информационных технологий БИНОМ: Лаборатория знаний, 2006. – 382 с.
2. Колодчак О.М. Інтелектуальний аналіз даних – ЛПНУ, 2013 – ст. 49-58
3. Імангулова З.А., Семенцова А.М. Застосування методів інтелектуального аналізу даних для виявлення цільових груп клієнтів у готельному бізнесі // II Міжнародна наукова конференція «Період трансформаційних процесів в світовій науці: задачі та виклики». Кривий ріг, Україна, 2024. С. 319 – 321.
4. Ладиженська Р. С. Технологія обслуговування в готелях і туристичних комплексах – ХНАМГ, 2010 – 54 с.
5. Готелі: основні нормативні вимоги URL: https://bz.ligazakon.ua/ua/magazine_article/BZ0129 (дата звернення: -13.11.2023)
6. Alation Data Governance. Simplify Governance for Success – Alation [Електронний ресурс] – Режим доступу: https://www.alation.com/product/data-governance/ (Дата звернення: 29.12.2023);
7. Customer Case Study: Virgin Australia. Airline Embarks on a Journey to Become Data-Driven Success – Alation [Електронний ресурс] – Режим доступу: https://www.alation.com/customers/virgin-australia/ (Дата звернення: 29.12.2023);
8. 7 Best Room Service Apps for Hotels 2022 URL: https://get.hotefy.com/7-best-room-service-app-for-hotels/ (дата звернення: -02.05.2022)
9. Types of hotel services URL: https://wiki.otelms.com/en/post/types-of-hotel-services/ (дата звернення: -11.11.2023)
10. Хауссем І. Firebase Cookbook / Houssem Yahiaoui. – Packt Publishing, 2017. –54 с.
11. Економіка готельно-ресторанного бізнесу. Навчальний посібник / Басюк Т.П., Керанчук Т.Л. – К.:НУХТ, 2018. 276 с.
12. Томашевський О.М. та ін.. Інформаційні технології та моделювання бізнес процесів. Навчальний посібник.- К.: Видавництво «Центр учбової літератури», 2012. – 148 с.

13. Regression. What is regression? – DataRoot [Електронний ресурс] – Режим доступу: https://www.datarobot.com/wiki/regression/#:~:text=What%20is%20Regression%3F,models%20estimate%20a%20numeric%20value (Дата звернення: 25.12.2023)

14. Словник готельно-ресторанних термінів/ Вишневська О. О. - Х.: ХНУ імені В. Н. Каразіна, 2012. – 20 с.

15. Knowledge Discovery Through Data Mining: What Is Knowledge Discovery? – Tandem Computers Inc., 1996 – 253 с.

16. Методи інтелектуального аналізу та оброблення даних / В.О. Гороховатський, І.С. Творошенко . – ХНУРЕ, 2021. – 38 с.

17. Project Jupyter Documentation – Docs.Jupyter [Електронний ресурс] – Режим доступу: https://docs.jupyter.org/en/latest/ (Дата звернення: 13.01.2024).

18. Handbook of Research on Holistic Optimization Techniques in the Hospitality, Tourism, and Travel Industry / Pandian Vasant, Kalaivanthan M. – IGI-GlobalEditors, 2016 – с. 140-166.

19. Project Jupyter Documentation – Docs.Jupyter [Електронний ресурс] – Режим доступу: https://docs.jupyter.org/en/latest/projects/architecture/content-architecture.html (Дата звернення: 19.01.2024).