

УДК 621.923.42

## СИСТЕМА КЛАСТЕРИЗАЦІЇ КЛІЄНТСЬКОГО НАБОРУ ДАНИХ ДЛЯ ПОКРАЩЕННЯ ЯКОСТІ ОБСЛУГОВУВАННЯ

**Валентій В. С.**, здобувач вищої освіти, гр. МПІН-191

Науковий керівник: **Дорош М. С.**, д.т.н., доцент  
*Національний університет "Чернігівська політехніка"*

Розвиток методів запису і зберігання даних привело до бурхливого зростання обсягів інформації, що збирається і аналізується. Обсяги даних настільки значні, що людині просто не під силу проаналізувати їх самостійно, хоча необхідність проведення такого аналізу цілком очевидна, адже в цих "сирих" даних укладені знання, які можуть бути використані при прийнятті рішень. Для того щоб провести автоматичний аналіз даних, використовується Data Mining<sup>1</sup>.

Сьогодні майже в будь-якому великому бізнесі є величезна купа даних, які збираються і зберігаються протягом багатьох років. Data Mining - це інструмент аналізу даних бізнесу та знаходження в них певних залежностей. Кластеризація - об'єднання в групи схожих об'єктів - є однією з фундаментальних завдань в галузі аналізу даних і Data Mining. Економічна і політична ситуація швидко змінюється, тому бізнес потребує впевненості не тільки в сьогоdnішньому, але і в завтрашньому дні. Інтелектуальний аналіз даних дає можливість управляти майбутнім бізнесу шляхом точного прогнозування, заснованого на аналізі минулого і сьогоdnення.

**Мета роботи:** створити систему кластеризації наборів клієнтських даних, яка сприятиме покращенню якості обслуговування, збільшенню прибутку, підвищенню лояльності клієнтів, прогнозуванню продажів.

Структура розроблюваного додатку буде містити чотири основні модулі:

–модуль імпорту відповідає за конвертування і інтегрування даних з файлів різних форматів (csv, xls, xlsx тощо);

–модуль підготовки до аналізу. Предобробку даних можна розділити на два етапи – очищення і трансформацію. Очищення дозволяє виключити з даних чинники, що заважають їх коректній обробці аналітичними алгоритмами: аномальні і фіктивні значення, пропуски, протиріччя, дублікати, шуми тощо. Трансформація даних виконується для їх узгодження (типів, форматів, діапазонів тощо) із існуючими аналітичними алгоритмами. Вона включає такі операції, як квантування, нормування, перетворення форматів і типів даних, кодування, злиття, угруповання тощо;

–модуль аналізу підготовлених даних - для аналізу вхідних даних будемо використовувати алгоритми data mining, а саме: алгоритм k – найближчих сусідів (kNN)<sup>2</sup> та Apriori<sup>3</sup>. Для вирішення завдання класифікації можуть використовуватися метод k-найближчого сусіда (k-Nearest Neighbor). Класифікація найбільш проста і поширена задача Data Mining. В результаті рішення задачі класифікації виявляються ознаки, які характеризують групи об'єктів досліджуваного набору даних - класи; за цими ознаками новий об'єкт можна віднести до того чи іншого класу. Найбільш відомий алгоритм рішення задачі пошуку асоціативних правил - алгоритм Apriori. В результаті виконання завдання пошуку асоціативних правил відшукуються закономірності між пов'язаними подіями в наборі даних. Пошук закономірностей здійснюється не на основі властивостей аналізованого об'єкта, а між кількома подіями, які відбуваються одночасно;

–модуль візуалізації за допомогою якого користувач формує найбільш інформативне подання для певного виду даних (таблиці, графіки, діаграми, текстові документи тощо).

Архітектура додатку<sup>4</sup> зображена на рисунку 1.

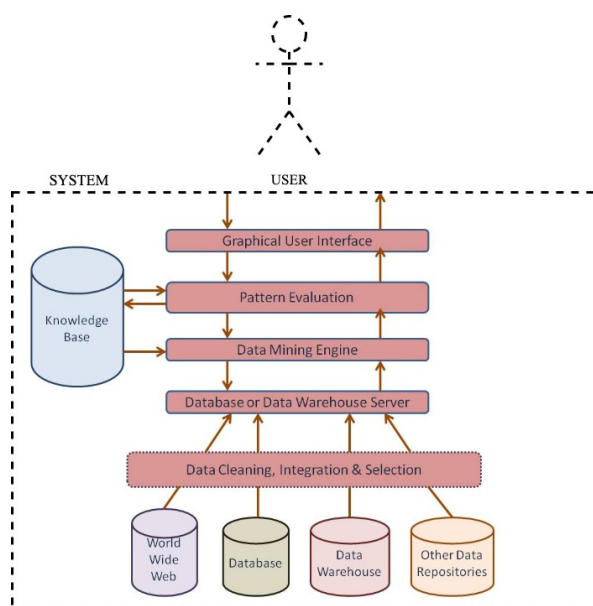


Рисунок 1 – Архітектура додатку

Основні модулі архітектури додатку:

–модуль GUI взаємодіє між користувачем та системою обміну даними. Цей модуль допомагає користувачеві легко та ефективно використовувати систему, не знаючи реальної складності процесу;

–модуль pattern evaluation головним чином відповідає за міру цікавості шаблону, використовуючи порогове значення. Він взаємодіє з механізмом обміну даними, щоб зосередити пошук на цікавих зразках;

–database or data warehouse server містить фактичні дані, готові до обробки. Отже, сервер відповідає за отримання відповідних даних на основі запиту користувача щодо вилучення даних;

–data mining engine є основним компонентом будь-якої системи обміну даними. Він складається з ряду модулів для виконання завдань з пошуку даних, включаючи об'єднання, класифікацію, характеристику, кластеризацію, прогнозування, аналіз часових рядів тощо;

–knowledge base (База знань) є корисною для всього процесу вилучення даних. База знань може навіть містити переконання користувачів та дані з досвіду користувачів, які можуть бути корисними в процесі пошуку даних. Система data mining engine може отримати дані з бази знань, щоб зробити результат більш точним та надійним. Модуль pattern evaluation регулярно взаємодіє з базою знань, щоб отримати вхідні дані, а також оновити їх.

Отже, розробка додатку для прогнозування продажів, аналізу клієнтів та продуктів або послуг на мові програмування Python з використанням графічного фреймворку Qt та модулем PyQt5 значно підвищить ефективність прийняття рішень при веденні бізнесу.

#### Список використаних джерел

1. Data Mining — добыча данных. BaseGroup Labs: веб-сайт. URL: <https://basegroup.ru/community/articles/data-mining>
2. Nearest Neighbors. Scikit learn: веб-сайт. URL: <https://scikit-learn.org/stable/modules/neighbors.html>
3. Apriori Algorithm. GeeksforGeeks: веб-сайт. URL: <https://www.geeksforgeeks.org/apriori-algorithm/>
4. Data Mining Architecture. URL: <https://www.wideskills.com/data-mining-tutorial/data-mining-architecture>