

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки
Кафедра програмної інженерії

ЗВІТ
Лабораторної роботи № 3
з дисципліни «Інтелектуальний аналіз даних»
на тему «Алгоритм Apriori та асоціативні правила»

Виконав

студент групи ІПЗм-24-2

Голодніков Дмитро

Перевірив

ст. викл. Онищенко К.Г.

Харків 2024

1 МЕТА РОБОТИ

Ознайомитися з алгоритмом Apriori для пошуку частих наборів елементів та генерації асоціативних правил. Вивчити метрики support, confidence, lift. Застосувати алгоритм до аналізу транзакційних даних (market basket analysis).

2 ХІД ВИКОНАННЯ РОБОТИ

2.1 Теоретичні основи

Алгоритм Apriori використовує властивість анти-монотонності: якщо набір елементів не є частим, то жоден його надмножина також не є частою. Це дозволяє ефективно відсікати простір пошуку.

Основні метрики: Support - частка транзакцій з даним набором; Confidence - умовна ймовірність правила; Lift - відношення confidence до очікуваної.

Реалізація алгоритму Apriori з використанням бібліотеки mlxtend:

```
from mlxtend.frequent_patterns import apriori, association_rules
from mlxtend.preprocessing import TransactionEncoder

# Кодування транзакцій у бінарну матрицю
te = TransactionEncoder()
te_array = te.fit_transform(transactions)
df = pd.DataFrame(te_array, columns=te.columns_)

# Пошук частих наборів елементів
frequent_items = apriori(df, min_support=0.2, use_colnames=True)

# Генерація асоціативних правил
rules = association_rules(frequent_items,
                           metric="confidence",
                           min_threshold=0.5)
```

2.2 Розподіл підтримки

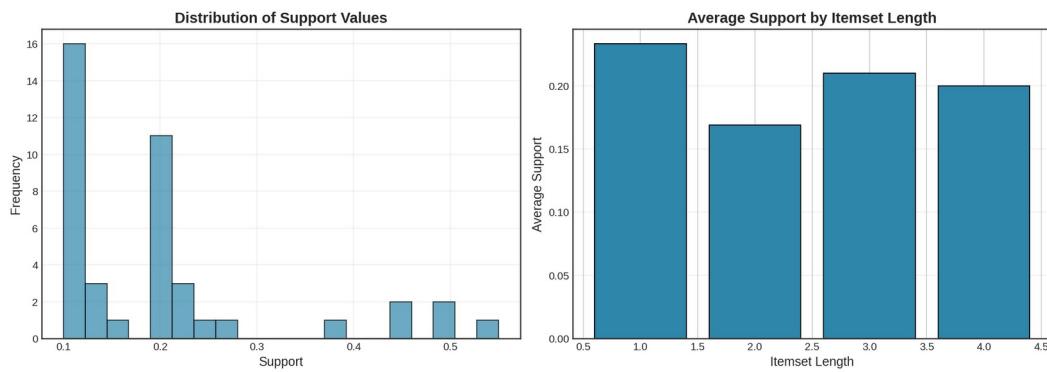


Рис. 2.1 - Розподіл підтримки частих наборів

2.3 Метрики асоціативних правил

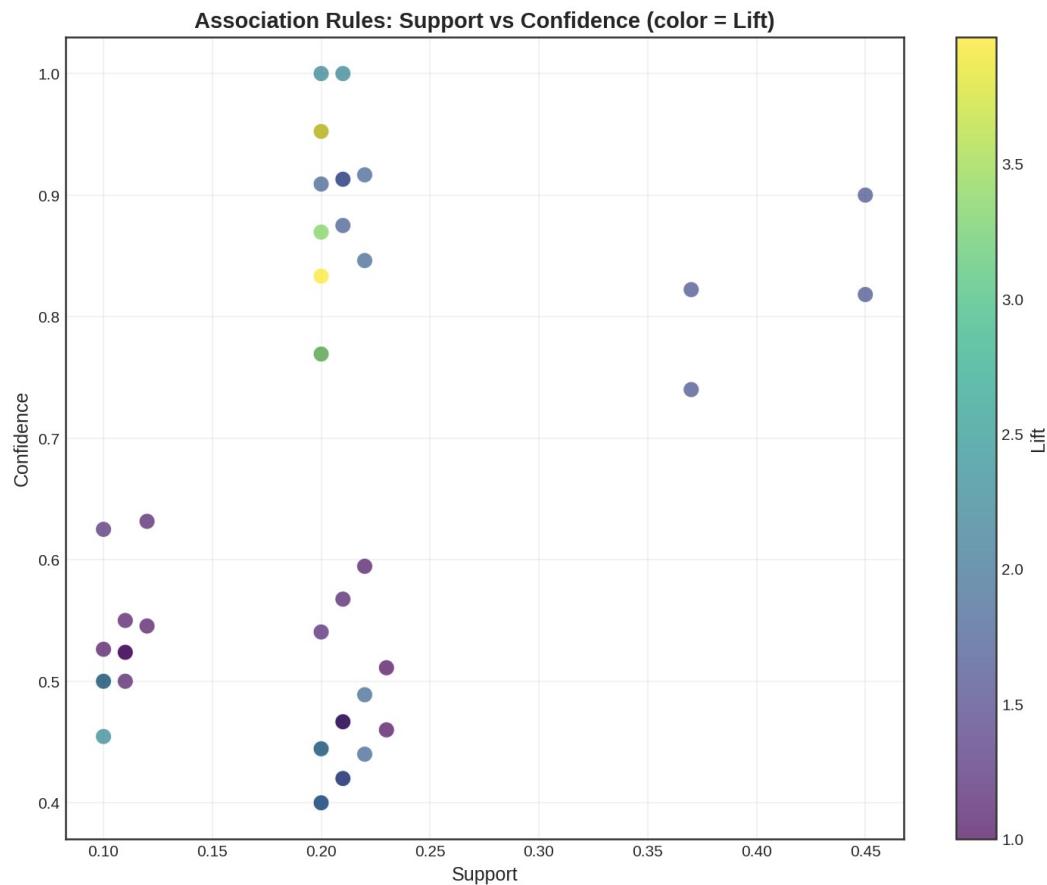


Рис. 2.2 - Метрики знайдених асоціативних правил

2.4 Теплова карта правил

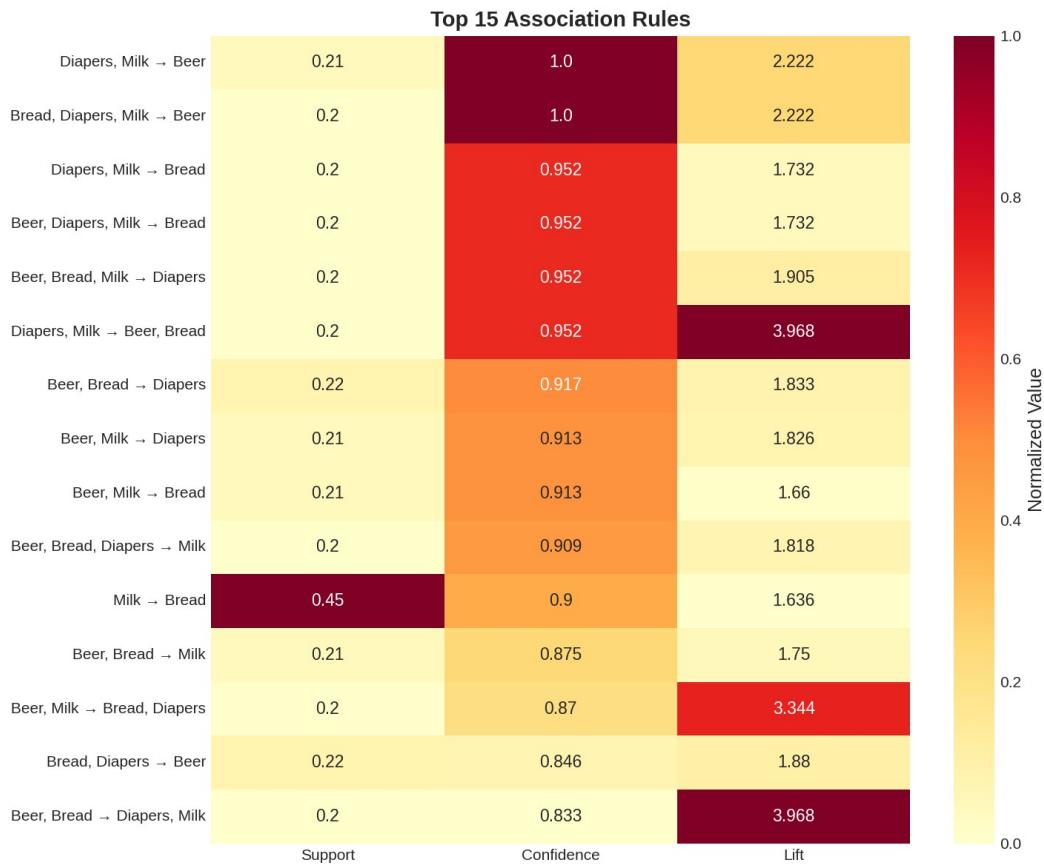


Рис. 2.3 - Теплова карта асоціативних правил

2.5 Мережа асоціацій

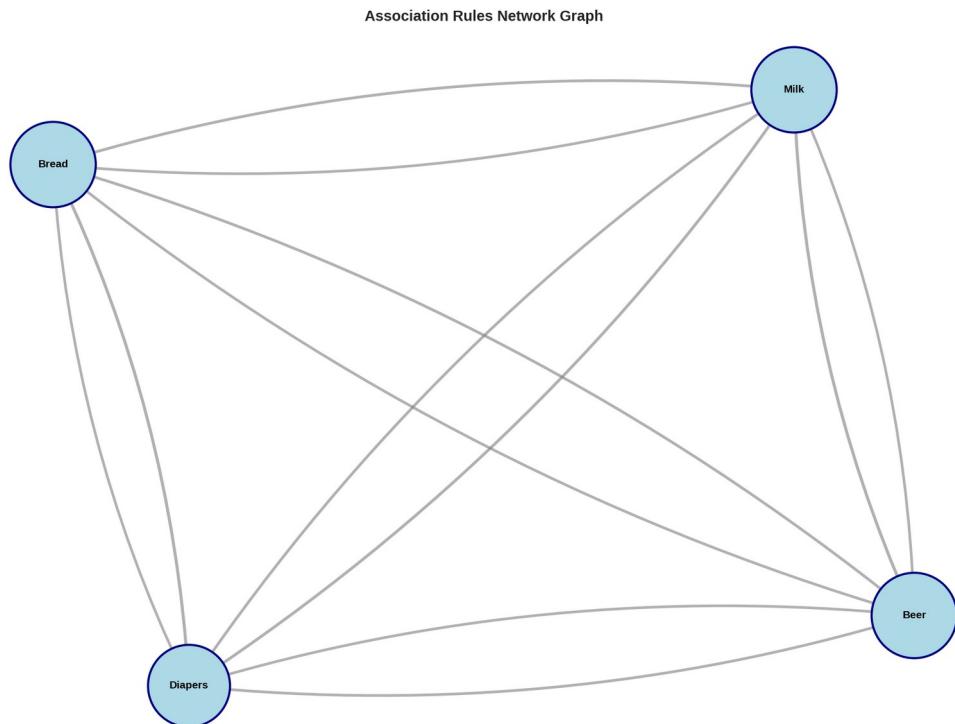


Рис. 2.4 - Мережева візуалізація асоціативних правил

2.6 Матриця частот

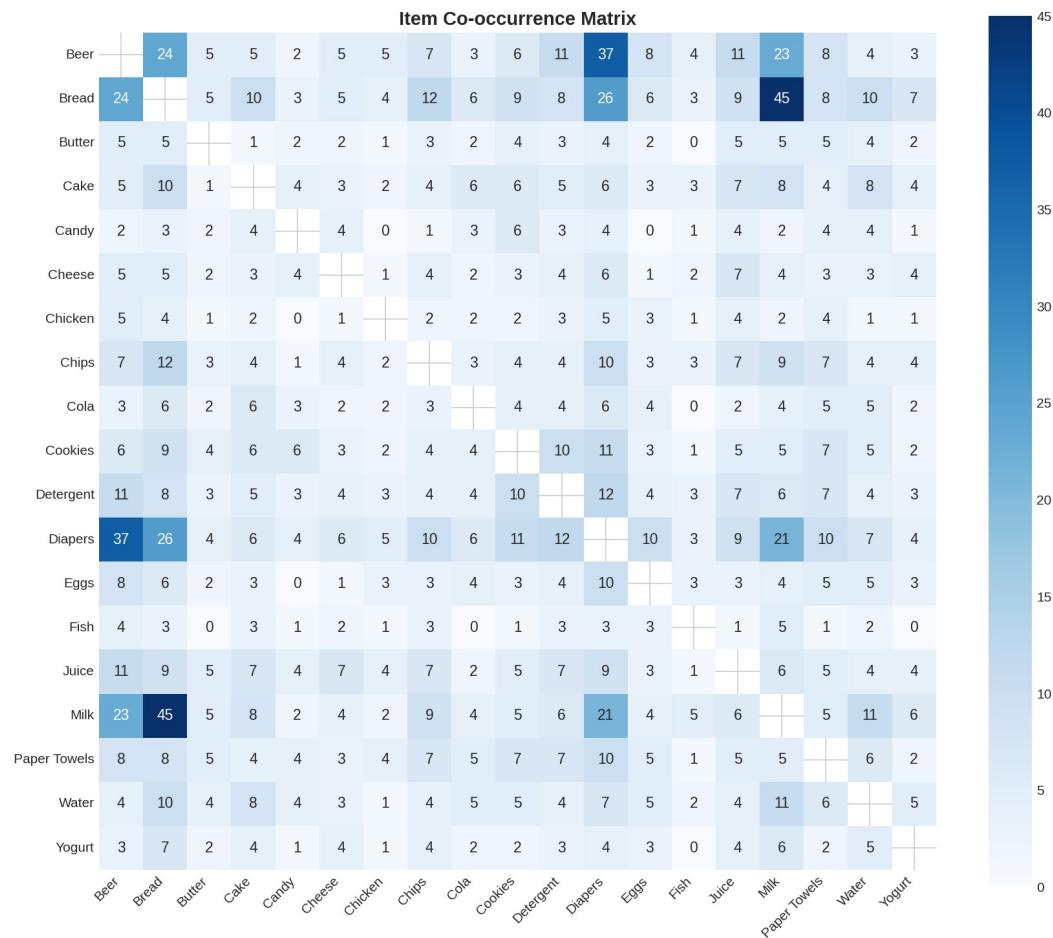


Рис. 2.5 - Матриця спільної появи товарів

2.7 Аналіз чутливості

Проведено аналіз чутливості кількості знайдених правил до порогових значень min_support та min_confidence.

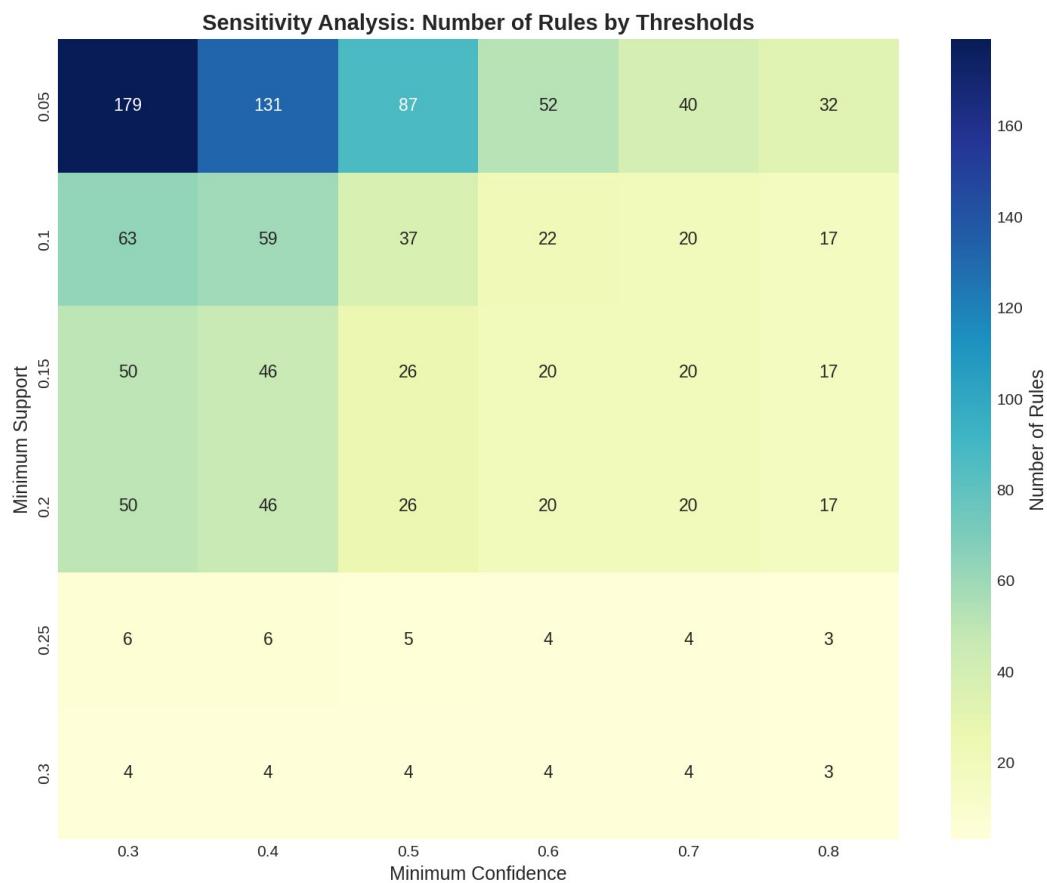


Рис. 2.6 - Аналіз чутливості до порогових значень

3 ВИСНОВКИ

У ході виконання лабораторної роботи було реалізовано алгоритм Apriori для пошуку частих наборів елементів та генерації асоціативних правил. Проаналізовано транзакційні дані супермаркету та виявлено закономірності у покупках. Візуалізовано результати у вигляді теплових карт, мережевих графів та матриць частот. Досліджено вплив порогових значень support та confidence на кількість знайдених правил. Отримані знання застосовуються у рекомендаційних системах та маркетинговому аналізі.

ПОСИЛАННЯ

Код проекту доступний у репозиторії GitHub: <https://github.com/na-naina/data-analysis-khnure>