МІНІСТЕРСТВО НАУКИ І ОСВІТИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗВІТ З ЛАБОРАТОРНОЇ РОБОТИ № 4 ЗА ТЕМОЮ:

**Пошук асоціативних правил в середовищі аналізу даних**

Група \_\_\_\_\_11\_\_\_\_\_\_

Курс \_\_\_\_\_1\_\_\_\_\_\_\_\_\_

Студент (ка) \_\_\_\_Супруненко М. І. \_\_\_\_

Дата оформлення \_\_\_\_\_\_05.03.2025\_\_\_\_\_\_\_

Перевірив \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

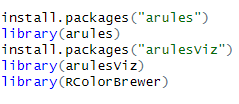
Дата \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

МЕТА РОБОТИ: Метою лабораторної роботи є формування професійних вмінь та навичок щодо використання **апріорного алгоритму (A-priori) для обчислення частих наборів елементів (товарів)**, вміння застосовувати отримані знання на практиці в практичних задачах аналізу великих даних.

ТЕОРІЯ: Теоретичні питання щодо інструментів та методів побудов асоціативних правил детально розглянуті в лекційних матеріалах за темою 5 «Оброблення документів у великих даних. Індустрія продажів. Індустрія розваг. Перспективи розвитку моделей і методів обробки великих даних».

ХІД РОБОТИ:

1. Завантажте необхідні бібліотеки



1. Завантажте набір даних ‘**Market\_Basket\_Optimisation.csv**‘. Опишіть цей набір даних: кількість записів, які дані в ньому збережені тощо.

dataset = read.transactions('Market\_Basket\_Optimisation.csv',

sep = ',', rm.duplicates = TRUE)

1. Використайте функцію ‘**apriori()**‘. Опишіть параметри функції та можливі налаштування.
2. Застосуйте функцію **inspect()**. Опишіть параметри функції та можливі налаштування. Опишіть отримані асоціації.
3. Візуалізуйте отримані результати щодо підтримки різних продуктів, використовуючи функцію itemFrequencyPlot(). Опишіть параметри налаштування функції.
4. Побудуйте візуалізацію правил за допомогою функції:

plot(associa\_rules, method = "graph",

measure = "confidence", shading = "lift")

1. Встановлення та імпорт бібліотек

|  |
| --- |
| # Встановлення необхідних бібліотек (виконується один раз)  install.packages("arules")  install.packages("arulesViz")  install.packages("RColorBrewer")  # Імпортуємо бібліотеки  library(arules)  library(arulesViz)  library(RColorBrewer) |

2. Завантаження та опис набору даних

|  |
| --- |
| # Завантаження набору даних у формат транзакцій  dataset <- read.transactions("Market\_Basket\_Optimisation.csv",                               sep = ",",                               rm.duplicates = TRUE)  # Перегляд основної інформації про набір даних  summary(dataset) |
|  |

3. Візуалізація популярності товарів

|  |
| --- |
| # Графік популярності товарів  itemFrequencyPlot(dataset, topN = 20, col = brewer.pal(8, "Pastel2"), main = "Топ-20 найпопулярніших товарів") |
|  |

4. Застосування алгоритму A-priori

|  |
| --- |
| # Налаштування параметрів та пошук асоціативних правил  associa\_rules <- apriori(dataset, parameter =                             list(supp = 0.001, conf = 0.7, minlen = 4))  # Виведення підсумку  summary(associa\_rules) |
|  |

5. Аналіз отриманих правил сортуючи по ліфту

|  |
| --- |
| inspect(sort(associa\_rules, by = "lift")[1:10]) |

6. Візуалізація асоціативних правил

|  |
| --- |
| strong\_rules <- subset(associa\_rules, lift > 3)  plot(strong\_rules, method = "graph", measure = "confidence", shading = "lift") |

**ВИСНОВКИ:**

У ході лабораторної роботи було досліджено апріорний алгоритм для пошуку асоціативних правил у великих даних. Проаналізовано набір Market\_Basket\_Optimisation.csv, визначено найбільш популярні товари та знайдено асоціативні зв’язки між ними. Використання функції apriori() дозволило отримати правила, що вказують на товари, які часто купуються разом. Візуалізація результатів допомогла виявити основні зв’язки між продуктами. Отримані знання можна застосовувати в маркетингових стратегіях та рекомендаційних системах.

Граючись з параметрами алгоритму, я дійшов до того, що при параметрах високого конфіденсу = 0,8 і при супорті = 0,002, то майже в усіх наборох присутня вода, що свідчить про велику кількість транзакцій з водою. Для подальшого аналізу треба видалити цей продукт.

Якщо в нас значення супорт меньше = 0,001, то правил стає більше беруться більше товарів і можно зробити краще аналіз.

Саме відсортувавши значення по ліфту ми отримали цікаві значення, бо ця метрика каже про те, що саме ці товари разом зустрічаються частіше.