Міністерство освіти і науки, молоді та спорту України

Національний Технічний Університет України

«Київський Політехнічний Інститут»

Навчально-науковий комплекс

«Інститут прикладного системного аналізу»

Кафедра системного проектування

Лабораторна робота №4

з дисципліни «Інтелектуальний аналіз даних»

на тему:« **Знайомство з методами побудови асоціативних правил**»

Варіант №14

Виконав: студент 3 курсу

групи ДА-52

Плотніков С.O.

Перевірив: [Іщенко Г. В.](http://rozklad.kpi.ua/Schedules/ViewSchedule.aspx?v=327eda95-fa17-470c-a431-e67a04d790d4).

Київ 2018

**Мета роботи**: Ознайомитися та набути навички побудови асоціативних правил за допомогою Weka.

**Завдання:** Побудувати асоціативні правила для тестових екземплярів даних за допомогою зазначеного методу.

**Сутність асоціації**

На відміну від методів побудови моделей класифікації, методи пошуку асоціативних правил не потребують вибору атрибуту класу, усі атрибути вважаються атрибутами ознак, класом є комбінація значень окремих атрибутів ознак. Методи побудови асоціативних правил призначені для пошуку комбінацій значень атрибутів, на базі яких за комбінацією значень атрибутів першої множини («умова») можна спрогнозувати значення атрибутів другої множини («наслідок»)*.*

Для пошуку виберемо метод Apriori – класичний метод пошуку закономірностей у значеннях атрибутів у вигляді асоціативних правил. Метод Apriori перебирає усі можливі комбінації значень заданої кількості атрибутів, наприклад, *комбінація1*={*атрибут А* значення *А1*, *атрибут Б значення Б1*}, та вибирає ті комбінації, які перевищують мінімальне значення критерію підтримки (support criterion, задається у параметрах методу).

Метод Apriori починає пошук з закономірностей, які містять лише одне значення атрибуту. Далі, в процесі пошуку, виходячи з того факту, що менші комбінації атрибутів зустрічаються частіше, кількість атрибутів у шуканих закономірностях поступово збільшується. Пошук зупиняється за умови відсутності знайдених екземплярів для поточної кількості атрибутів.

**Зміст параметрів методу Apriori**

Асоціативне правило складається з двох множин *X* («умова») і *Y* («наслідок») у вигляді конструкції IF-THEN: *X → Y*, тобто, якщо знайдені значення атрибутів множини *X*, тоді з ним ймовірно будуть знайдені значення атрибутів множини *Y*.

Критерій підтримки (support criterion, див. Рисунок 2.4.1) – частка кількості екземплярів, яка припадає на кожне асоціативне правило (від 0.0 до 1.0). Для 24 екземплярів підтримка у 20% (minSupport=0.2) дорівнює 5, тобто кожне асоціативне правило має виконуватися щонайменше для 5 екземплярів з усього набору даних.

Критерій достовірності (confidence criterion) – відношення кількості екземплярів множини (*X* U *Y)* до кількості екземплярів множини *X* у знайденному правилі. Відношення характеризує рівень зв’язку атрибутів множин *X* та *Y.*

**Виконання роботи**

*Інформація про дані та атрибути екземплярів даних.*

Множина даних описує голосування кожного конгресмену Палати представників США по 16 ключовим питанням голосування. Конгресмени могли проголосувати за, проти або утриматись.

*Number of Instances*: 435 (267 democrats, 168 republicans)

*Number of Attributes*: 16 + class name = 17 (all Boolean valued)

*Attribute Information:*

1. Название класса: 2 (democrat, republican)

2. Голосование относительно детей-инвалидов: 2 (y,n)

3. Водные проекты: 2 (y,n)

4. Принятие резолюции о бюджете: 2 (y,n)

5. Заморозка оплаты врачам: 2 (y,n)

6. el-salvador-aid: 2 (y,n)

7. Вопрос религионых групп в школах: 2 (y,n)

8. Запрет на антисанитарный тест: 2 (y,n)

9. aid-to-nicaraguan-contras: 2 (y,n)

10. mx-missile: 2 (y,n)

11. Иммигарция: 2 (y,n)

12. Сокращение корпорации Synfuels: 2 (y,n)

13. Расходы на образование: 2 (y,n)

14. superfund-right-to-sue: 2 (y,n)

15. Криминал: 2 (y,n)

16. Беспошлинный экспорт: 2 (y,n)

17. export-administration-act-south-africa: 2 (y,n)

Attribute: #Missing Values:

1: 0

2: 0

3: 12

4: 48

5: 11

6: 11

7: 15

8: 11

9: 14

10: 15

11: 22

12: 7

13: 21

14: 31

15: 25

16: 17

17: 28

@relation vote

@attribute 'handicapped-infants' { 'n', 'y'}

@attribute 'water-project-cost-sharing' { 'n', 'y'}

@attribute 'adoption-of-the-budget-resolution' { 'n', 'y'}

@attribute 'physician-fee-freeze' { 'n', 'y'}

@attribute 'el-salvador-aid' { 'n', 'y'}

@attribute 'religious-groups-in-schools' { 'n', 'y'}

@attribute 'anti-satellite-test-ban' { 'n', 'y'}

@attribute 'aid-to-nicaraguan-contras' { 'n', 'y'}

@attribute 'mx-missile' { 'n', 'y'}

@attribute 'immigration' { 'n', 'y'}

@attribute 'synfuels-corporation-cutback' { 'n', 'y'}

@attribute 'education-spending' { 'n', 'y'}

@attribute 'superfund-right-to-sue' { 'n', 'y'}

@attribute 'crime' { 'n', 'y'}

@attribute 'duty-free-exports' { 'n', 'y'}

@attribute 'export-administration-act-south-africa' { 'n', 'y'}

@attribute 'Class' { 'democrat', 'republican'}

@data

'n','y','n','y','y','y','n','n','n','y',?,'y','y','y','n','y','republican'

'n','y','n','y','y','y','n','n','n','n','n','y','y','y','n',?,'republican'

?,'y','y',?,'y','y','n','n','n','n','y','n','y','y','n','n','democrat'

'n','y','y','n',?,'y','n','n','n','n','y','n','y','n','n','y','democrat'

'y','y','y','n','y','y','n','n','n','n','y',?,'y','y','y','y','democrat'

'n','y','y','n','y','y','n','n','n','n','n','n','y','y','y','y','democrat'

'n','y','n','y','y','y','n','n','n','n','n','n',?,'y','y','y','democrat'

'n','y','n','y','y','y','n','n','n','n','n','n','y','y',?,'y','republican'

'n','y','n','y','y','y','n','n','n','n','n','y','y','y','n','y','republican'

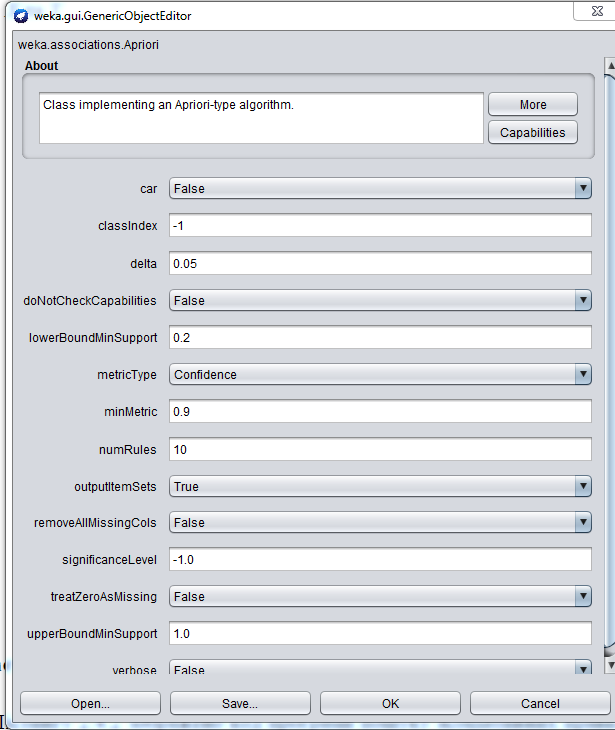
'y','y','y','n','n','n','y','y','y','n','n','n','n','n',?,?,'democrat'

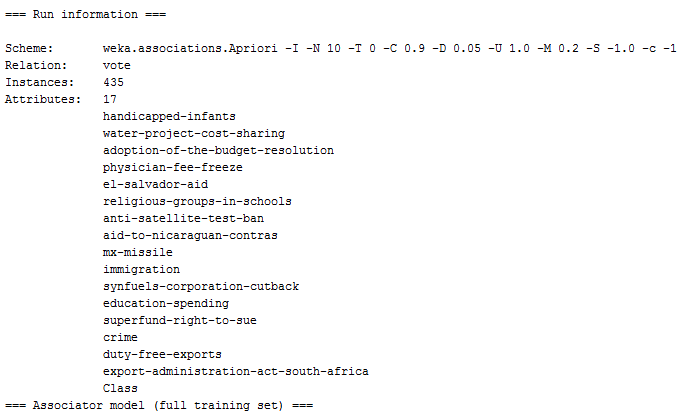
'n','y','n','y','y','n','n','n','n','n',?,?,'y','y','n','n','republican'

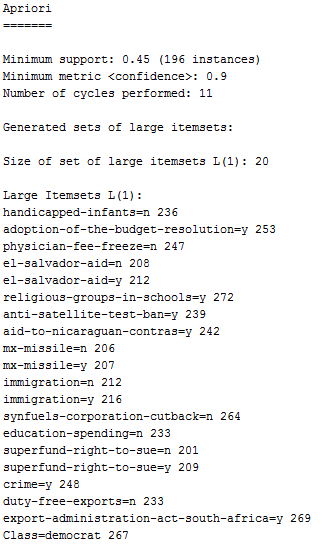
'n','y','n','y','y','y','n','n','n','n','y',?,'y','y',?,?,'republican'

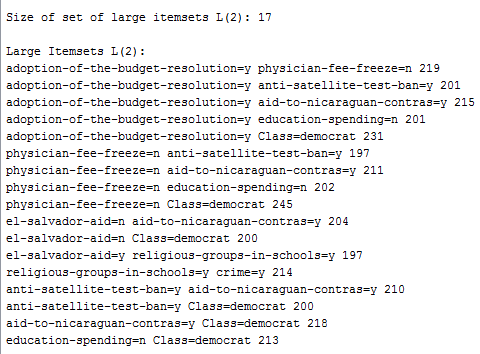
'n','y','y','n','n','n','y','y','y','n','n','n','y','n',?,?,'democrat'

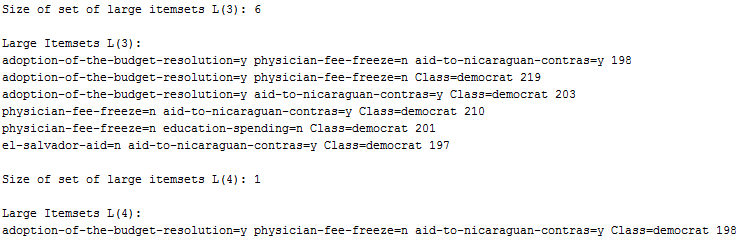
………………………………………………………

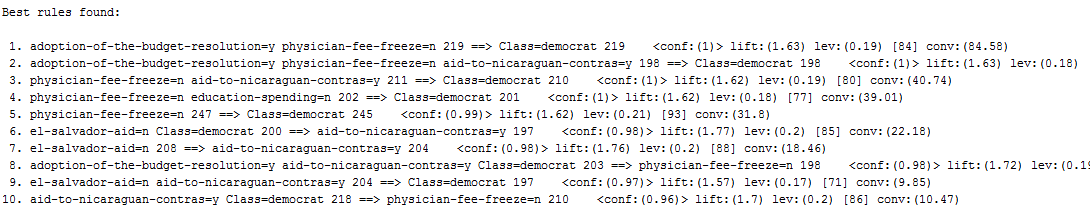




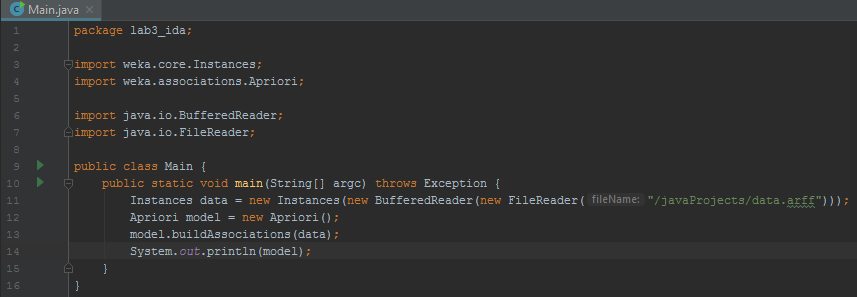


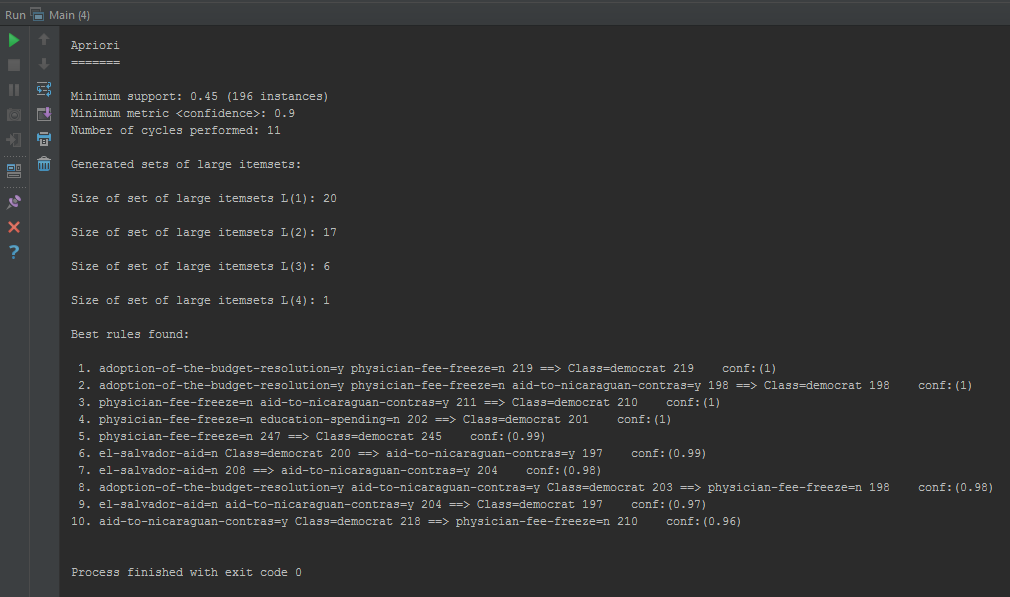






**Застосування Weka API в JAVA**





**Побудуємо асоціативні правила для іншого набору даних**

@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}

@attribute temperature {hot, mild, cool}

@attribute humidity {high, normal}

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

@data

sunny,hot,high,FALSE,no

sunny,hot,high,TRUE,no

overcast,hot,high,FALSE,yes

rainy,mild,high,FALSE,yes

rainy,cool,normal,FALSE,yes

rainy,cool,normal,TRUE,no

overcast,cool,normal,TRUE,yes

sunny,mild,high,FALSE,no

sunny,cool,normal,FALSE,yes

rainy,mild,normal,FALSE,yes

sunny,mild,normal,TRUE,yes

overcast,mild,high,TRUE,yes

overcast,hot,normal,FALSE,yes

rainy,mild,high,TRUE,no

