

8.6. Лабораторна робота №6 «Аналітичні рішення на основі асоціативних правил».

Завдання до лабораторної роботи.

1. Відповідно до заданого варіанту підготувати необхідні дані у вигляді таблиць MS Excel і зберегти їх як персональні файли. Для підготовки даних використовувати тематичні сайти Інтернет, результати проходження практик, довідники і каталоги.

2. Виконати завантаження даних з таблиць MS Excel за допомогою Майстра імпорту відповідно до методики, представленої в розділі «Порядок виконання роботи».

3. Отримати аналітичне рішення на основі асоціативних правил відповідно до методики, представленої в розділі «Порядок виконання роботи».

4. Виконати економічний аналіз різних ситуацій, застосовуючи основні візуалізатора, відповідно до методики, представленої в розділі «Порядок виконання роботи».

Варіанти завдань:

N варіанту	Галузь
1	Продажі продовольчих товарів (шаблон покупок)
2	Продажі промислових товарів (шаблон покупок)
3	Продажі будівельних товарів (шаблон покупок)
4	Продажі комп'ютерної техніки (шаблон покупок)
5	Зручне розміщення товарів на прилавках супермаркетів
6	Зручне розміщення товарів на прилавках промислових магазинів
7	Зручне розміщення товарів на вітринах книжкових електронних магазинів
8	Зручне розміщення товарів на вітринах електронних магазинів загального призначення
9	Зручне розміщення товарів на вітринах комп'ютерних електронних магазинів
10	Стимулювання продажів промислових товарів
11	Стимулювання продажів продукції ЗМІ
12	Стимулювання продажів продовольчих товарів
13	Стимулювання продажів комп'ютерної техніки

Теоретична частина.

Асоціативні правила дозволяють знаходити закономірності між зв'язаними подіями. Прикладом такого правила, служить твердження, що покупець, який купує «Хліб», придбає і «Молоко» з вірогідністю 75%. Вперше

ця задача була запропонована для пошуку асоціативних правил при знаходженні типових шаблонів покупок, що здійснюються в супермаркетах, тому інколи її ще називають *аналізом ринкової кошику* (market basket analysis).

Транзакція – це множина подій, відбувшись одночасно. Хай є база даних, що складається з купівельних транзакцій. Кожна транзакція – це набір товарів, куплених покупцем за один візит. Таку транзакцію ще називають *ринковою кошиком*. Після визначення поняття транзакція, можна перейти до визначення поняття асоціативного правила.

Хай є список транзакцій. Необхідно знайти закономірності між цими подіями. Як в умові, так і внаслідку правила повинні знаходитися елементи транзакцій. Хай $I = \{i_1, i_2, \dots, i_n\}$ – множина елементів, що входять в транзакції. *Асоціативним правилом називається імплікація $X \Rightarrow Y$, де $X \subset I, Y \subset I$ та $X \cap Y = \emptyset$. Правило $X \Rightarrow Y$ має підтримку s (support), якщо $s\%$ транзакцій з D , містять $X \cup Y$, $\sup p(X \Rightarrow Y) = \sup p(X \cup Y)$. Достовірність правила показує, яка вірогідність того, що з X виходить Y . Правило $X \Rightarrow Y$ справедливо з достовірністю (confidence) c , якщо $c\%$ транзакцій з D , що містять X , також містять Y , $\text{conf}(X \Rightarrow Y) = \sup p(X \cup Y) / \sup p(X)$.*

Покажемо це на конкретному прикладі: «75% транзакцій, що містять хліб, також містять молоко. 3% від загального числа всіх транзакцій містять оба товари». 75% – це достовірність (confidence) правила, 3% це підтримка (support), або «Якщо «Хліб», то «Молоко»» з вірогідністю 75%. Іншими словами, метою аналізу є встановлення наступних залежностей: якщо в транзакції зустрівся деякий набір елементів X , то на підставі цього можна зробити висновок про те, що інший набір елементів Y також же повинен з'явитися в цій транзакції. Встановлення таких залежностей дає нам можливість знаходити дуже прості і інтуїтивно зрозумілі правила.

Алгоритми пошуку асоціативних правил призначені для знаходження всіх правил $X \Rightarrow Y$, причому підтримка і достовірність цих правил мають бути вище за деякі наперед задані пороги, названі відповідно *мінімальною підтримкою* (minsupport) і *мінімальною достовірністю* (minconfidence).

Значення для цих параметрів вибираються так, щоб обмежити кількість знайдених правил. Якщо підтримка має велике значення, то алгоритми знаходитимуть правила, добре відомі аналітикам або настільки очевидні, що немає жодного сенсу проводити такий аналіз. З іншого боку, низьке значення підтримки веде до генерації величезної кількості правил, що, звичайно, вимагає істотних обчислювальних ресурсів. Проте, більшість цікавих правил знаходяться саме при низькому значенні порогу підтримки. Хоча дуже низьке значення підтримки веде до генерації статистично необґрунтованих правил.

Пошук асоціативних правил зовсім не тривіальна задача, як може здатися на перший погляд. Одна з проблем – алгоритмічна складність при знаходженні часто зустрічаючих наборів елементів, оскільки із зростанням числа елементів експоненціально зростає число потенційних наборів елементів. Звичайні асоціативні правила – це правила, в яких як в умові, так і в слідстві присутні лише елементи транзакцій і при обчисленні яких використовується лише інформація про те, чи присутній елемент в транзакції чи ні. Фактично всі наведені вище приклади відносяться до звичайних асоціативних правил. Для пошуку звичайних асоціативних правил в програмі служить технологія *«Ассоциативные правила»*.

Налаштування. По-перше необхідний вказати, що є ідентифікатором (ID) транзакції, а що елементом транзакції. Наприклад, ідентифікатор транзакції – це номер чека або код накладної. А елемент – це найменування товару в чеці. Потім слідує налаштування параметрів пошуку правил. Застосовуються слідує параметри:

- мінімальна і максимальна підтримка. Асоціативні правила шукаються лише в деякій множині всіх транзакцій. Для того, щоб транзакція увійшла до цієї множини вона повинна зустрітись у вхідній вибірці декілька разів, більше мінімальної підтримки і менше максимальною. Наприклад, мінімальна підтримка дорівнює 1%, а максимальна – 20%. Кількість елементів «Хліб» і «Молоко» стовпця «Товар» з однаковим значенням стовпця «Номер чека»

зустрічаються в 5% всіх транзакцій (номерів чека). Тоді ці два рядки увійдуть до шуканої множини;

- мінімальна і максимальна достовірність. Це процентне відношення кількості транзакцій, що містять всі елементи, які входять в правило, до кількості транзакцій, що містять елементи, які входять в умову. Якщо транзакція – це замовлення, а елемент – товар, то достовірність характеризує, наскільки часто купуються товари, що входять в слідство, якщо замовлення містить товари, що увійшли до всього правила.

Порядок виконання роботи.

1. Розглянемо механізм пошуку асоціативних правил на прикладі даних про продажі товарів в деякій торгівельній точці. Дані знаходяться у файлі «Supermarket.xls», підготовленому за допомогою MS Excel (НЕ МЕНШЕ 30 ТРАНЗАКЦІЙ) (рис. 8.26). У таблиці представлена інформація по покупках продуктів декількох груп. Вона має всього два поля «НОМЕР ЧЕКА» і «ТОВАР». Необхідно вирішити задачу аналізу споживчої корзини з метою подальшого вживання результатів для стимулювання продажів.

Номер чека	Товар
160698	КЕТЧУПЫ, СОУСЫ, АДЖИКА
160698	МАКАРОННЫЕ ИЗДЕЛИЯ
160698	ЧАЙ
160747	МАКАРОННЫЕ ИЗДЕЛИЯ
160747	МЕД
160747	ЧАЙ
161217	КЕТЧУПЫ, СОУСЫ, АДЖИКА
161217	МАКАРОННЫЕ ИЗДЕЛИЯ
161217	СЫРЫ

Рис. 8.26. Дані файлу «Supermarket.xls».

2. За допомогою «Мастера импорта» імпортуємо дані з файлу в пакет Deductor і проглянемо їх у вигляді таблиці.

3. Для пошуку асоціативних правил запустимо майстер обробки. У ньому виберемо тип обробки «Ассоциативные правила». На другому кроці майстра необхідно вказати, який стовпець є ідентифікатором транзакції («Номер чека»), а який є елементом транзакції («Товар»).

4. Наступний крок дозволяє налаштувати параметри побудови асоціативних правил: мінімальну і максимальну підтримку, мінімальну і максимальну достовірність, а також максимальну потужність множини. Виходячи з характеру наявних даних, слід вказати кордони підтримки – 13% і 80%, та достовірності 60% і 90%.

5. Наступний крок дозволяє запустити процес пошуку асоціативних правил. На екрані відображається інформація про кількість множин, кількість знайдених правил, а також гістограма розподілу знайденої множини, що часто зустрічається по потужності.

6. Після завершення процесу пошуку отримані результати можна проглянути, використовуючи спеціальну візуалізацію: «Популярні набори», «Правила», «Дерево правил», «Що-якщо».

«Популярні набори» - це множина, що складається з одного і більше елементи, які найчастіше зустрічаються в транзакціях одночасно. На скільки часто зустрічається множина у вхідному наборі транзакцій можна судити по підтримці. Дана візуалізація відображує множину у вигляді списку (рис. 8.27).

N	↓ Множество	Поддержка	
		%	Кол-во
1	КЕТЧУПЫ, СОУСЫ, АДЖИКА	66,67	2
5	КЕТЧУПЫ, СОУСЫ, АДЖИКА И СЫРЫ	33,33	1
6	КЕТЧУПЫ, СОУСЫ, АДЖИКА И ЧАЙ	33,33	1
2	МЕД	33,33	1
7	МЕД И ЧАЙ	33,33	1
3	СЫРЫ	33,33	1
4	ЧАЙ	66,67	2

Рис. 8.27. Візуалізатор «Популярні набори».

Само назва візуалізації говорить про те, як застосувати результати на практиці. Набори товарів, що вийшли, найчастіше купують в даній торговельній точці, отже можна приймати рішення про постачання товарів, їх розміщенні і так далі.

Візуалізація «Правила» відображає асоціативні правила у вигляді списку правил. Цей список представлений таблицею із стовпцями: «номер правила», «умова», «слідство», «підтримка %», «підтримка, кількість», «достовірність» (рис. 8.28).

N	Условие	Следствие	Поддержка		Достоверность, %
			%	кол-во	
1	ВАФЛИ	СУХАРИ	22,73	10	71,43
2	СУХАРИ	ВАФЛИ	22,73	10	71,43
3	КЕТЧУПЫ, СОУСЫ, АДЖИКА	МАКАРОННЫЕ ИЗДЕЛИЯ	45,45	20	86,96
4	МАКАРОННЫЕ ИЗДЕЛИЯ	КЕТЧУПЫ, СОУСЫ, АДЖИКА	45,45	20	83,33
5	МЕД	ЧАЙ	40,91	18	91,82
6	СЫРЫ	ЧАЙ	29,55	13	68,42
7	ВАФЛИ И СУХАРИ	ЧАЙ	20,45	9	90,00
8	ВАФЛИ И ЧАЙ	СУХАРИ	20,45	9	68,23

Рис. 8.28. Візуалізатор «Правила».

Таким чином, експертові надається набір правил, які описують поведінку покупців. Наприклад, якщо покупець купив вафлі, то він з вірогідністю 71% також купить і сухарі.

Візуалізація «Дерево правил» - це завжди дворівневе дерево. Воно може бути побудоване або по умові, або по слідству. При побудові дерева правил по умові, на першому (верхньому) рівні знаходяться вузли з умовами, а на другому рівні - вузли із слідством. Другий варіант дерева правил - дерево, побудоване по слідству. Тут на першому рівні розташовуються вузли із слідством.

Праворуч від дерева знаходиться список правил, побудований по вибраному вузлу дерева. Для кожного правила відображаються підтримка і достовірність. Якщо дерево побудоване по умові, то зверху списку відображується умова правила, а список складається з його слідств. Тоді

правила відповідають на питання, що буде за такої умови. Якщо ж дерево побудоване по слідству, то зверху списку відображується слідство правила, а список складається з його умов. Ці правила відповідають на питання, що потрібно, аби було задане слідство. Дана візуалізація відображає ті ж самі правила, що і попередня, але в зручнішій для аналізу формі (рис. 8.29).

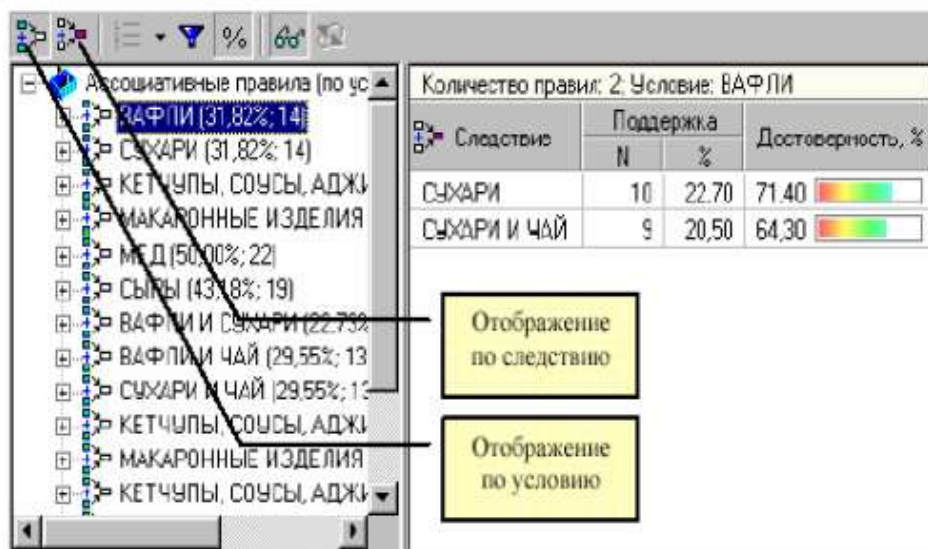


Рис. 8.29. Візуалізатор «Дерево правил».

В даному випадку правила відображують по умові. Тоді результат, що відображається в даний момент, можна інтерпретувати як 2 правила:

- Якщо покупець придбав вафлі, то він з вірогідністю 71% також придбає сухарі.
- Якщо покупець придбав вафлі, то він з вірогідністю 64% також придбає, сухарі і чай.

Аналогічно інтерпретуються і інші правила.

Аналіз "Що-якщо" в асоціативних правилах дозволяє відповісти на питання: що отримаємо як слідство, якщо виберемо дані умови? Наприклад, які товари купуються спільно з вибраними товарами. У вікні зліва розташований список всіх елементів транзакцій. Праворуч від кожного елемента вказана підтримка – скільки разів даний елемент зустрічається в транзакціях. У правому

верхньому кутку розташований список елементів, що входять в умову. Це, наприклад, список товарів, які придбав покупець. Для них потрібно знайти слідство. Наприклад, товари, що вживаються спільно з ними. Аби запропонувати людині те, що він можливо забув купити. У правому нижньому кутку розташований список слідств. Праворуч від елементів списку відображається підтримка і достовірність.

Хай необхідно проаналізувати, що, можливо, забув покупець придбати, якщо він вже купив вафлі і мед? Для цього необхідно додати в список умов ці товари (наприклад, за допомогою подвійного клацання миші) і потім натискувати на кнопку «*Вычислить правила*». При цьому в списку слідств з'являться товари, що спільно купуються з даними. В нашому випадку з'являться «СУХАРИ», «ЧАЙ», «СУХАРИ І ЧАЙ». Тобто можливо, покупець забув придбати сухарі або чай або і те і інше (рис. 8.30).

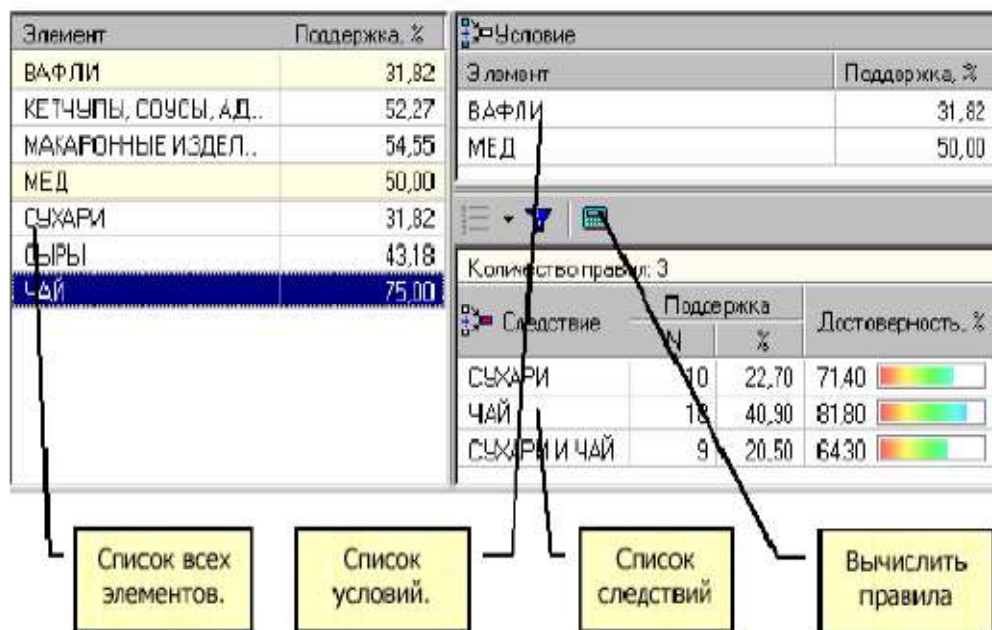


Рис. 8.30. Візуалізатор «Що-якщо».