

# Laboratorium z przedmiotu Data mining

Lab1.b – reguły asocjacyjne

# Pojęcia podstawowe

- Transakcja – podzbiór elementów z pewnej dziedziny elementów.
- Baza danych (BD) - zbiór transakcji.

# Reguły asocjacyjne - przypomnienie

Reguła asocjacyjna to wyrażenie w postaci:

$$A \Rightarrow B$$

gdzie A i B są zbiorami elementów

A – poprzednik reguły

B - następnik reguły

## Przykład

Jeżeli osoba O kupiła smartfona to również kupiła pakiet danych (wsp. 20%, zauf 80%).

# Parametry reguł asocjacyjnych

- Wsparcie (ang. support)
- Zaufanie (ang. confidence)
- Współczynnik podniesienia (ang. lift)
- Ograniczenia dotyczące występowania wybranych elementów w regułach

# Wsparcie 1

Wsparcie mówi jak dana reguła jest powszechna w rozpatrywanej bazie danych.

# Wsparcie 2

**Wsparcie bezwzględne** - liczba transakcji w bazie danych, w których zawarta jest reguła:

wsparcie bezwzględne reguły  $(A \Rightarrow B) =$   
liczba transakcji w BD zawierających  $A \cup B$   
gdzie  $t$  – transakcja

**Wsparcie względne** – częstość występowania reguły w bazie.

wsparcie względne  $(A \Rightarrow B) =$   
wsparcie bezwzględne  $(A \Rightarrow B) /$  liczba trans. w BD

# Zaufanie 1

Zaufanie wskazuje jak mocno poprzednik reguły determinuje pojawienie się następnika reguły.

# Zaufanie 2

**Zaufanie reguły** określa prawdopodobieństwo pojawienia się w transakcji następnika reguły, jeżeli wystąpił poprzednik reguły.

$$\text{zaufanie reguły } (A \Rightarrow B) = \frac{\text{wsparcie}(A \cup B)}{\text{wsparcie}(A)}$$



# Współczynnik podniesienia 1

## **Współczynnik podniesienia (wsp\_pod)**

pokazuje, jak bardzo zaufanie reguły jest większe od zaufanie wynikającego z prawdopodobieństwa występowania w bazie następnika reguły.

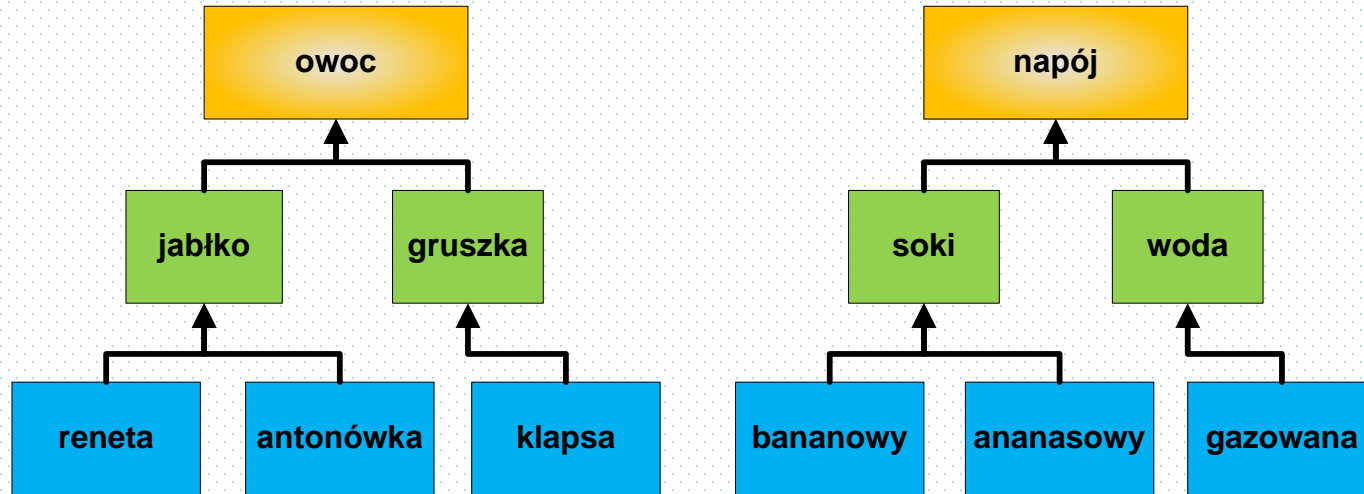
Współczynnik podniesienia  $(A \Rightarrow B) =$

zaufanie reguły  $(A \Rightarrow B)$  / wsparcie względne(B)

# Współczynnik podniesienia 2

- **wsp\_pod > 1** - pokazuje, że wystąpienie poprzednika reguły w transakcji zwiększa prawdopodobieństwo pojawienia się w tej transakcji następnika reguły.
- **wsp\_pod < 1** - pokazuje, że wystąpienie poprzednika reguły w transakcji zmniejsza prawdopodobieństwo pojawienia się w tej transakcji następnika reguły.
- **wsp\_pod = 1** - wskazuje na brak relacji pomiędzy poprzednikiem i następnikiem reguły (występowanie następnika nie zależy od występowania poprzednika).

# Hierarchia



Zastosowanie hierarchii pozwala na odkrycie bardziej ogólnych reguł, odnoszących się do różnych poziomów hierarchii np:

jabłko  $\Rightarrow$  woda, owoc  $\Rightarrow$  gazowana

# Odkrywanie reguł przy użyciu oprogramowania R

- Pakiet **arules** – zapewnia platformę do wykrywania reguł asocjacyjnych
  - oferuje funkcje do odkrywania reguł asocjacyjnych, ale tylko z jednym elementem w następniku.
  - wymaga danych w postaci transakcyjnej: transakcja jest określona jako zbiór binarnych elementów.
- Pakiet **arulesViz** – wizualizacja zbiorów częstych i reguł asocjacyjnych.

# Pakiet arules – wybrane funkcje

- ***itemFrequency*** – wyliczenie częstości występowania elementów
- ***ruleInduction*** – generowanie reguł asocjacyjnych na podstawie zbioru zbiorów częstych
- ***apriori*** , ***eclat*** – funkcje do wykrywania zbiorów częstych/reguł
- ***inspect*** – prezentacje reguł/zbiorów częstych w formie tekstowej
- ***subset*** – funkcja do filtrowania zbioru reguł lub zbioru zbiorów częstych