



Odkrywanie reguł asocjacyjnych

Rapid Miner

Zbiory częste

TS	ID_KLIENTA	Koszyk
12:57	1123	{mleko, pieluszki, piwo}
13:12	1412	{mleko, piwo, butki, masło, pieluszki}
13:55	1425	{piwo, wódka, wino, paracetamol}
14:09	1123	{mleko, pieluszki, butki, piwo}

Minimum support = 0.75

Zbiory częste

TS	ID_KLIENTA	Koszyk
12:57	1123	{mleko, pieluszki, piwo}
13:12	1412	{mleko, piwo, butki, masło, pieluszki}
13:55	1425	{piwo, wódka, wino, paracetamol}
14:09	1123	{mleko, pieluszki, butki, piwo}

Minimum support = 0.75

Sup({mleko})= 0.75

Sup({piwo})=1

Sup({pieluszki})= 0.75

Zbiory częste

TS	ID_KLIENTA	Koszyk
12:57	1123	{mleko, pieluszki, piwo}
13:12	1412	{mleko, piwo, bułki, masło, pieluszki}
13:55	1425	{piwo, wódka, wino, paracetamol}
14:09	1123	{mleko, pieluszki, bułki, piwo}

Minimum support = 0.75

Sup({mleko})= 0.75

Sup({mleko,pieluszki})= 0.75

Sup({piwo})=1

Sup({mleko,piwo})= 0.75

Sup({pieluszki})= 0.75

Sup({piwo,pieluszki})= 0.75

Zbiory częste

TS	ID_KLIENTA	Koszyk
12:57	1123	{mleko, pieluszki, piwo}
13:12	1412	{mleko, piwo, bułki, masło, pieluszki}
13:55	1425	{piwo, wódka, wino, paracetamol}
14:09	1123	{mleko, pieluszki, bułki, piwo}

Minimum support = 0.75

Sup({mleko})= 0.75

Sup({mleko,pieluszki})= 0.75

Sup({piwo})=1

Sup({mleko,piwo})= 0.75

Sup({pieluszki})= 0.75

Sup({piwo,pieluszki})= 0.75

Sup({mleko,piwo,pieluszki})= 0.75

Reguły asocjacyjne

- **poprzednik=>następnik**
- **{mleko,piwo,pieluszki}**
 - {mleko}=>{piwo, pieluszki}
 - {piwo}=>{mleko, pieluszki}
 - {pieluszki}=>{mleko,piwo}
 - {mleko, piwo}=>{pieluszki}
 - {mleko, pieluszki}=>{piwo}
 - {piwo, pieluszki}=>{mleko}

Ocena reguł asocjacyjnych

$$\text{conf}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)}$$

$$\text{lift}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X) * \sup(Y)}$$

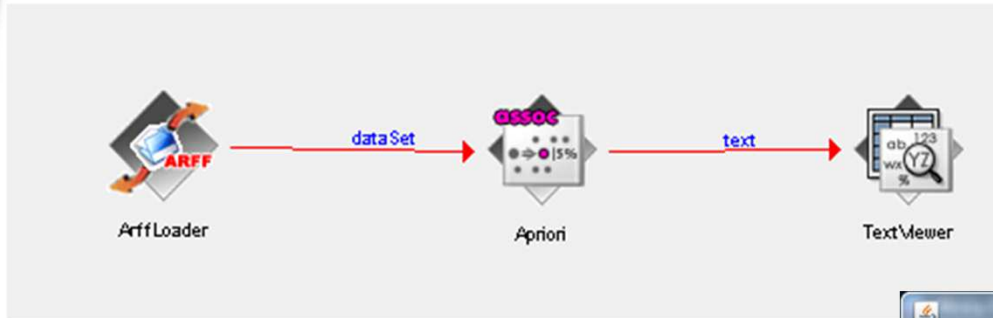
$$\text{conv}(X \Rightarrow Y) = \frac{1 - \sup(Y)}{1 - \text{conf}(X \Rightarrow Y)}$$

$$\text{lev}(X \Rightarrow Y) = \sup(X \cup Y) - \sup(X) * \sup(Y)$$

Ocena reguł asocjacyjnych

- **{piwo} $\Rightarrow$ {mleko, pieluszki}**
 - conf=0.75
 - lift=1
 - conv=1
 - lev=0
- **{mleko,pieluszki} $\Rightarrow$ {piwo}**
 - conf=1
 - lift=1
 - conv=0/0 ??
 - lev=0

Weka (1)



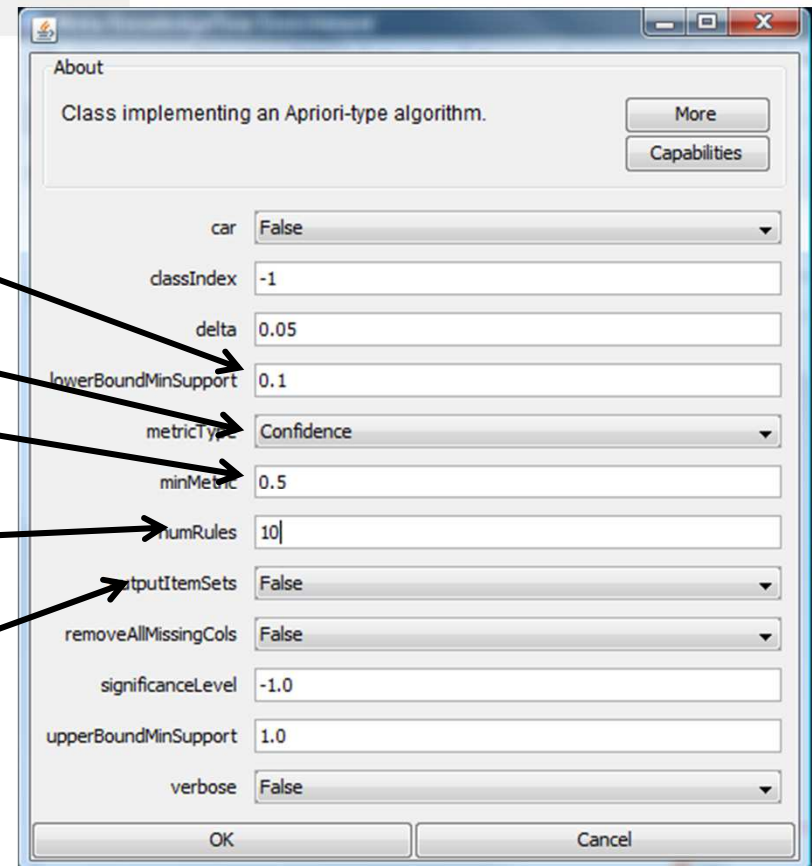
Minimalne wsparcie

Miara oceny reguł

Minimalna wartość miary

Liczba reguł do wygenerowania

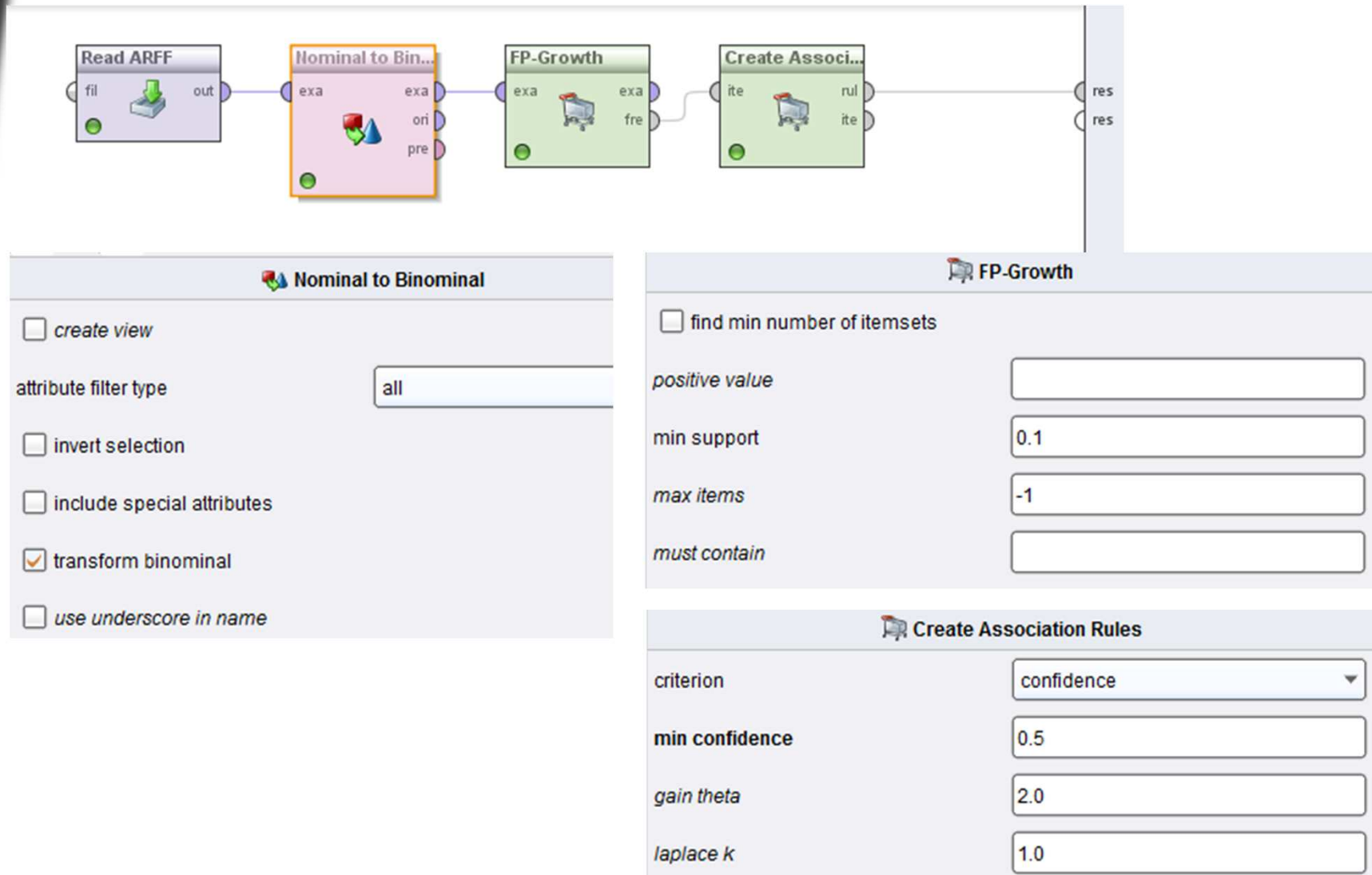
Czy zwrócić też zbiory częste



Zadanie

- Przeczytaj opis zestawu danych zakupy.arff
- Pobierz plik reguly_asocjacyjne.pdf i postępuj zgodnie z poleceniami.

Rapid Miner



Zadanie

- Przeprowadź podobną analizę jak w programie WEKA, ale wykorzystując RapidMiner.
- Wykorzystaj plik zakupy.binominal.arff
- Obejrzyj wygenerowane reguły w Graph View. Kiedy taki widok może się przydać?

Ilościowe reguły asocjacyjne

BÓL	WYSYPKA	GARDŁO	DIAGNOZA
T	BRAK	NORMA	NIESTRAWNOŚĆ
N	MAŁA	PRZEKR.	ALERGIA
N	BRAK	NORMA	PRZECHŁODZENIE
T	DUŻA	PRZEKR.	CZARNA OSPA
...	...	...	...

BÓL=T	BÓL=N	WYSYPKA =BRAK	WYSYPKA =MAŁA	WYSYPKA =DUŻA	GARDŁO =NORM.	GARDŁO =PRZEKRW.	...
1	0	1	0	0	1	0	
0	1	0	1	0	0	1	
0	1	1	0	0	1	0	
1	0	0	0	1	0	1	

{ból=T,Wysypka=brak,gardło=Norma,diagnoza=niestrawność}

{ból=N,Wysypka=mała,gardło=Przekr,diagnoza=alergia}

{ból=N,Wysypka=brak,gardło=Norma,diagnoza=przechłodzenie}

{ból=T,Wysypka=duża,gardło=Przekr,diagnoza=czarna ospa}

Zadanie (WEKA)

- Wczytaj plik flags.arff i wyszukaj ilościowe reguły asocjacyjne.
 - Wygeneruj reguły o wysokiej ufności, ale niskim wsparciu, oraz reguły o wysokim wsparciu, ale niekoniecznie bardzo wysokiej ufności.
 - Spróbuj zwiększyć limit na liczbę generowanych reguł (parametr „numRules” w „Apriori”). Przejrzyj uzyskane reguły. Czy potrafisz wyjaśnić niektóre z tych reguł? Którą regułę uważasz za najciekawszą?
 - Wykorzystaj opcję „outputItemsets” w konfiguracji „Apriori”, aby wygenerować, prócz reguł asocjacyjnych, zbiory częste. Spróbuj wskazać zbiór częsty dla reguły o najwyższej ufności, oraz zbiór częsty poprzednika tej reguły.
 - Potwierdź swoje przypuszczenia pod adresem:
<https://www.cia.gov/library/publications/the-world-factbook/docs/flagsoftheworld.html>

Zadanie (Rapid Miner)

- Zbuduj ilościowe reguły asocjacyjne dla zbioru danych weather_nominal.arff
- Pamiętaj aby każdy atrybut nominalny zmienić w zbiór atrybutów binominalnych.
- Ustaw minsup=0.3 i minconf=0.5 i zobacz wygenerowane reguły. Jak pewne są wygenerowane reguły? Co zrobić by uzyskać reguły o większym confidence?