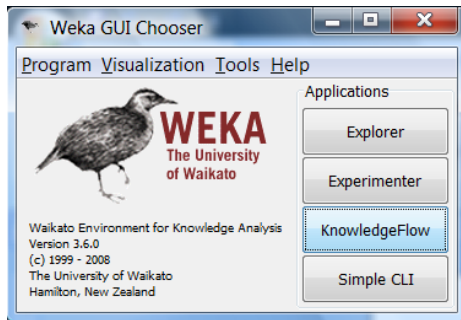


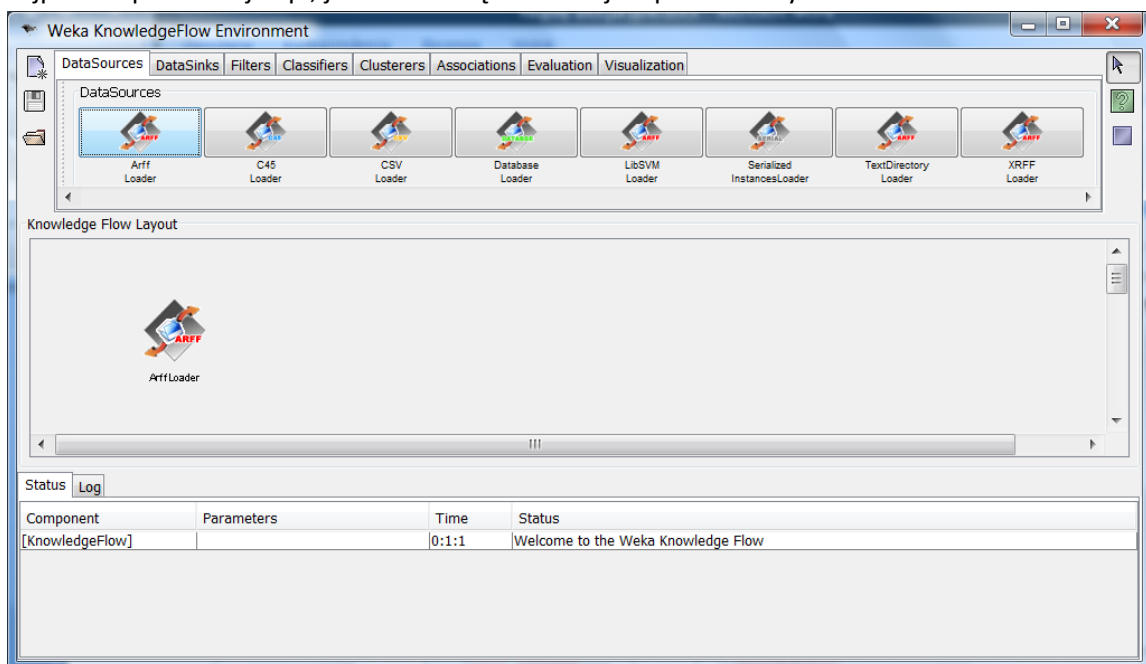
## Reguły asocjacyjne

Niniejsze ćwiczenie demonstruje działanie implementacji algorytmu apriori w systemie WEKA. Ćwiczenie ma na celu zaznajomienie studenta z działaniem systemu WEKA oraz znaczeniem podstawowych parametrów algorytmu priori i ich wpływu na zbiór odkrywanych reguł asocjacyjnych.

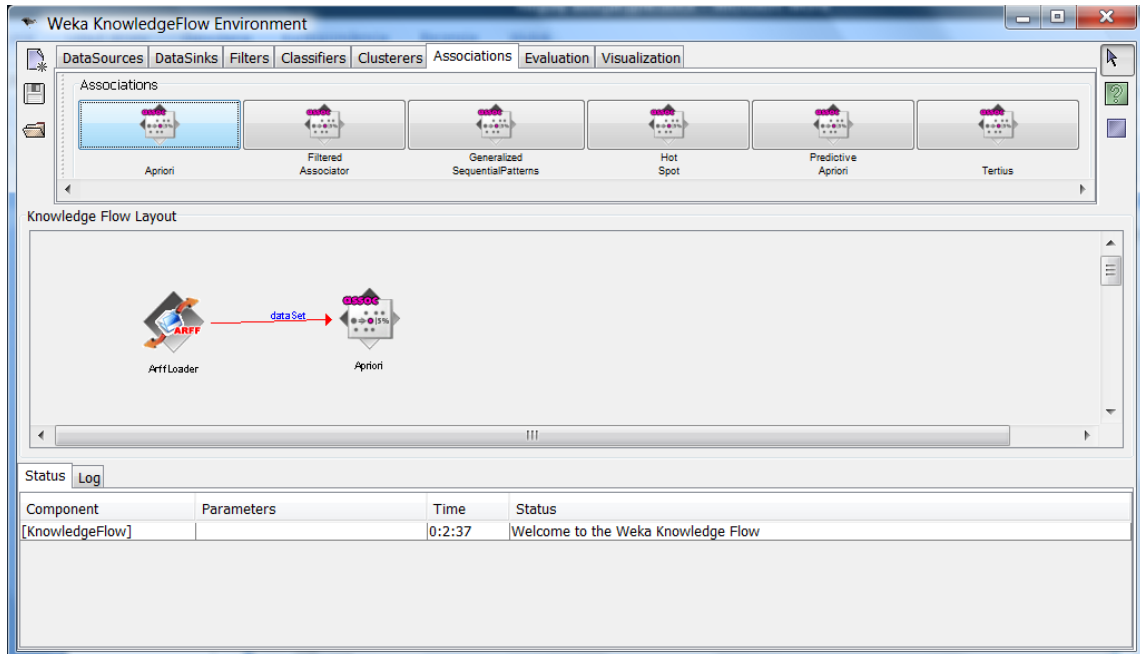
1. Uruchom system weka i wybierz aplikację „Knowledge Flow”.



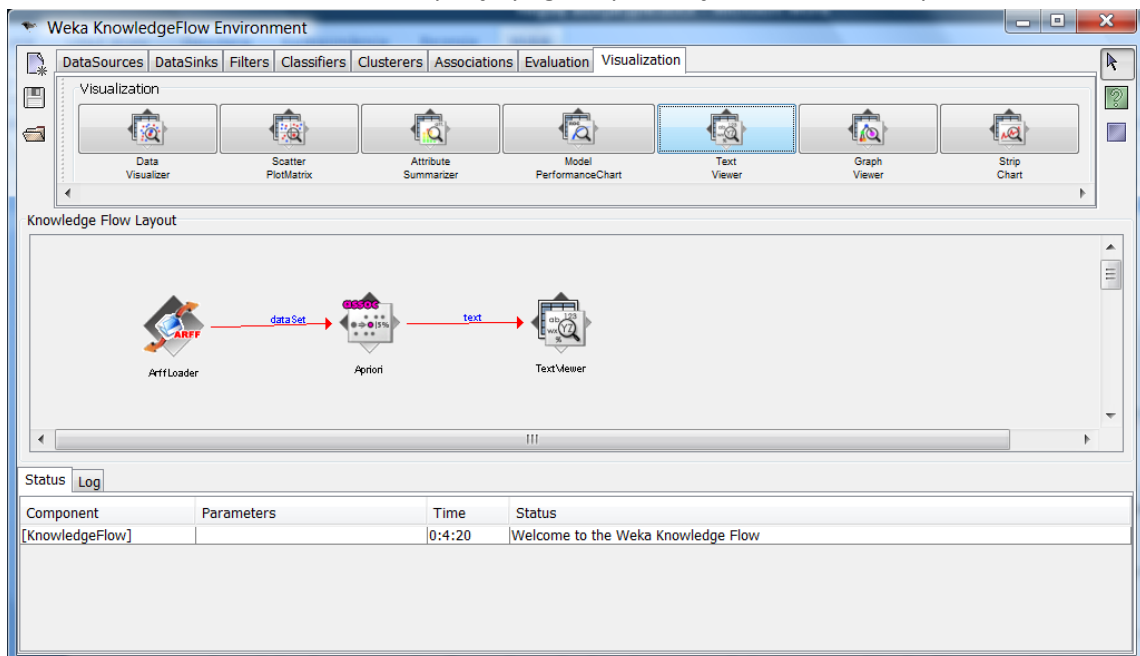
2. Wstaw źródło danych „Arff reader” i w konfiguracji tego źródła wybierz plik zakupy.arff. Zbiór ten będzie naszym przedmiotem odkrywania reguł asocjacyjnych mówiących które towary klient najprawdopodobniej kupi, jeżeli skusi się wcześniej na pewien inny zbiór towarów.



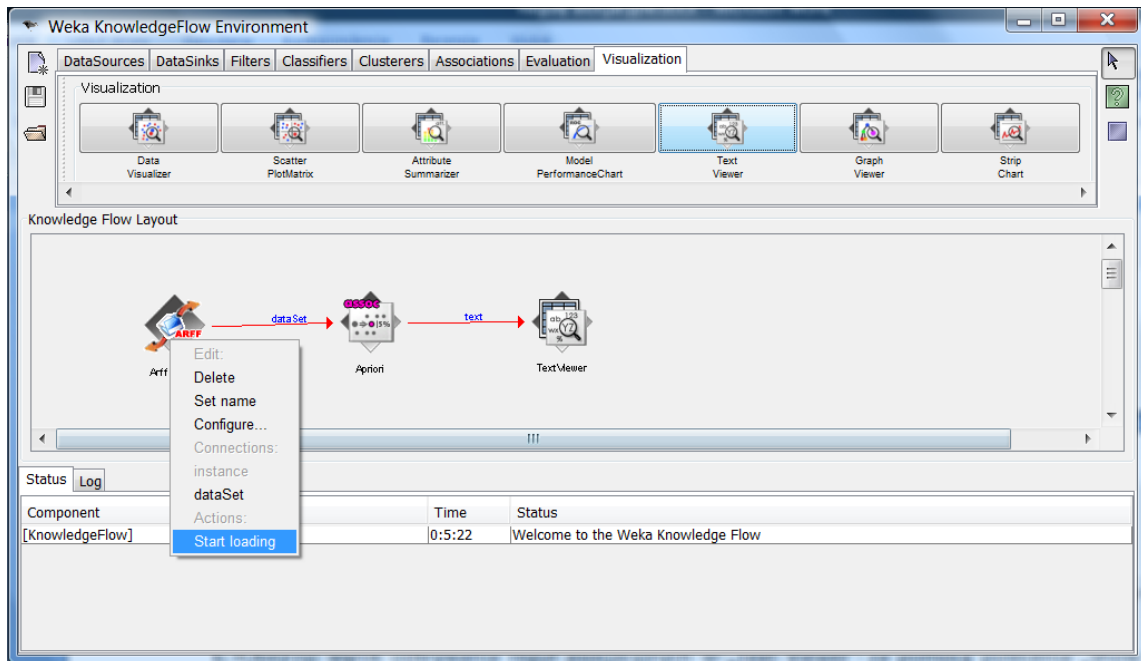
3. Ponieważ algorytmy odkrywania reguł asocjacyjnych są nienadzorowane, nie ma potrzeby zaznaczania który atrybut jest atrybutem decyzyjnym. Wobec braku zdefiniowanego atrybutu decyzyjnego rozróżnienie na zbiór danych, zbiór trenujący i zbiór testowy nie ma sensu. Dlatego też, obecnie należy wstawić algorytm odkrywania reguł asocjacyjnych „Apriori” z zakładki „Associations”, i połączyć go ze źródłem danych za pomocą „data set”.



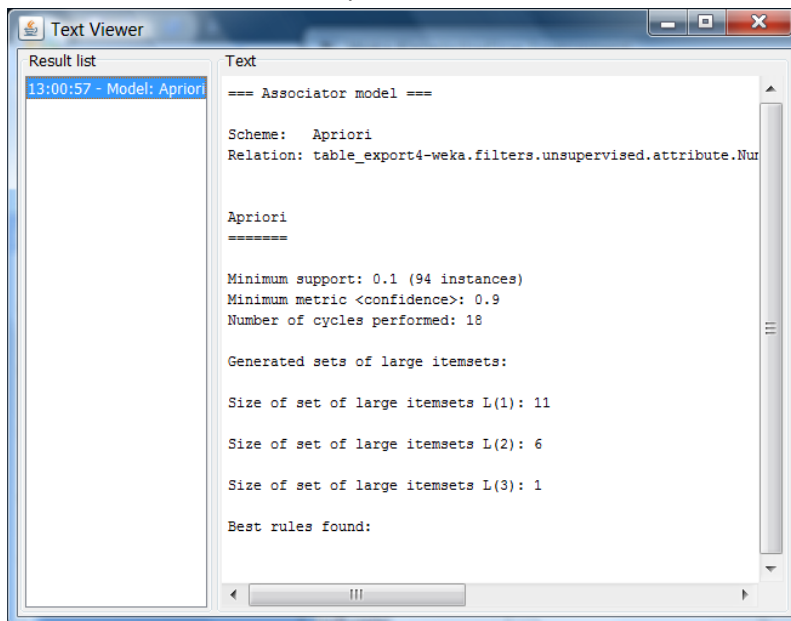
4. Aby obejrzeć rezultat działania algorytmu odkrywania reguł asocjacyjnych, należy wstawić „Text Viewer” z zakładki „Visualizations” i połączyć go za pomocą źródła „text” z „Apriori”.



5. Uruchom proces odkrywania reguł za pomocą „Start loading” w menu kontekstowym źródła danych.

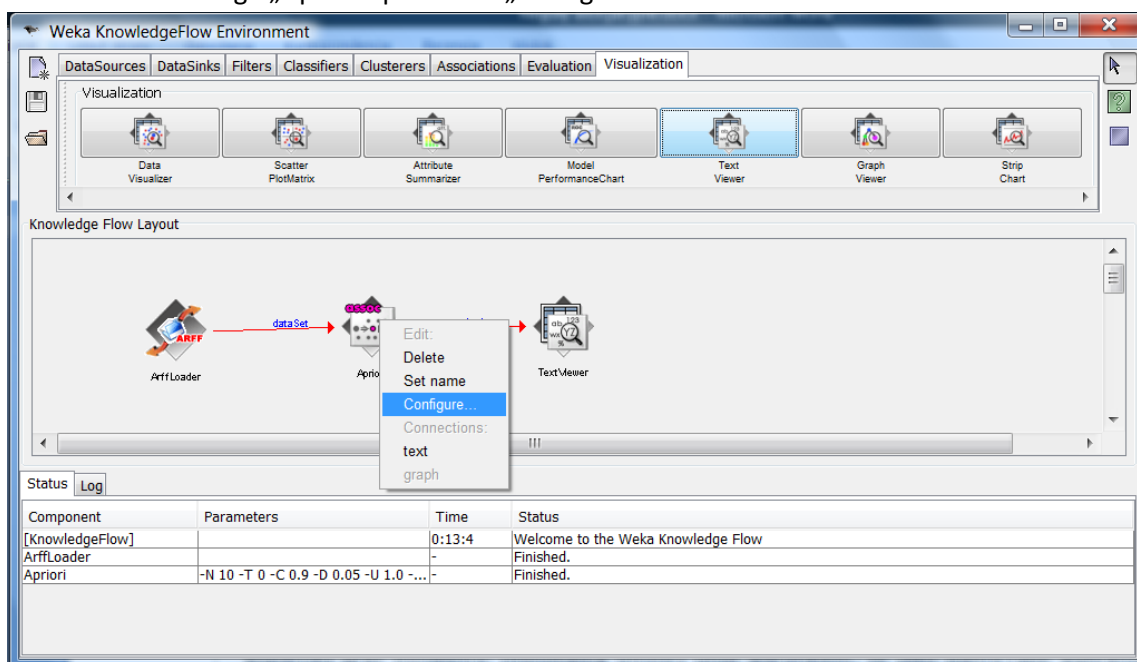


6. Obejrzyj wynik odkrywania reguł asocjacyjnych w „Text Viewer” za pomocą polecenia „Show results” w menu kontekstowym „Text Viewer”.

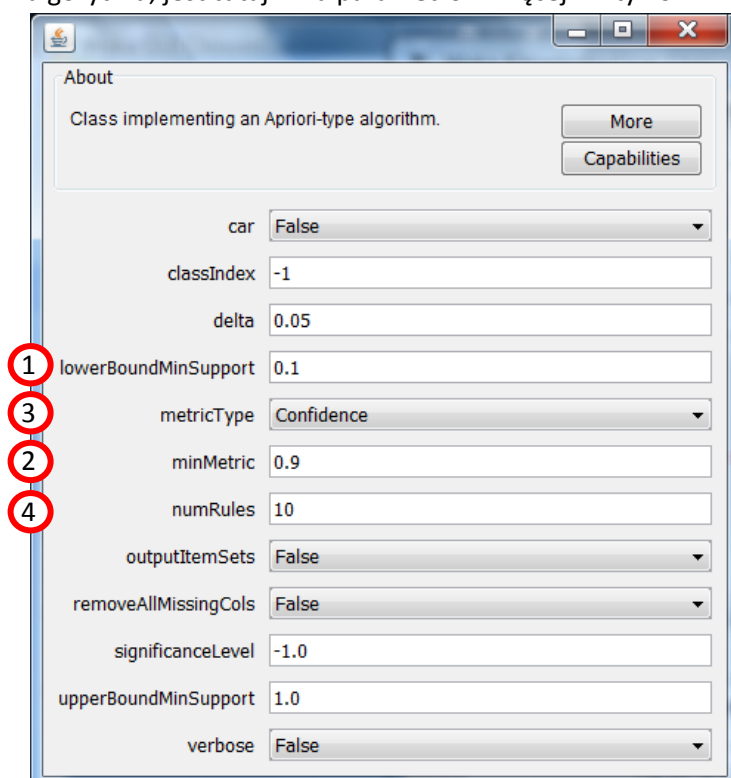


7. Jak łatwo zauważyć, nie odnaleziono żadnych reguł. Może tak się dzieć z dwóch powodów: minimalne wsparcie dla szukanych zbiorów częstych jest za wysokie, bądź za wysoki jest próg minimalnej ufności.

8. W pierwszym kroku spróbujemy obniżyć próg minimalnego wsparcia. Aby to zrobić, wybierz z menu kontekstowego „Apriori” polecenie „Configure”.

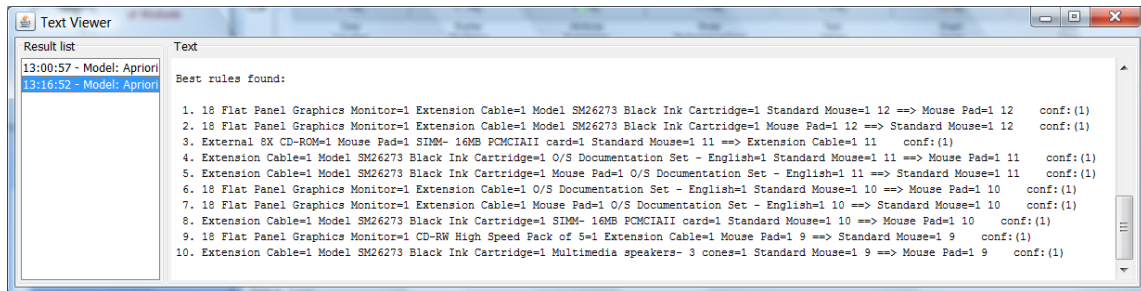


Powinno pojawić się okno dialogowe z kilkoma parametrami algorytmu. Ponieważ wersja Apriori zaimplementowana w systemie Weka jest rozszerzona w stosunku do podstawowej wersji tego algorytmu, jest tutaj kilka parametrów więcej niż tylko minimalne wsparcie i minimalna ufność.



Odpowiednikami tych parametrów są tutaj pozycje: lowerBoundMinSupport (1) (minimalne wsparcie) oraz minMetric (2) (minimalna ufność) (pod warunkiem, że jako metricType (3) wybrano „confidence”). Dodatkowym parametrem na który należy zwrócić uwagę, to numRules (4), który określa maksymalną liczbę reguł, jakie mają zostać wygenerowane. Wypróbuj minimalne wsparcia 0.07, 0.05 i 0.01. Czy udało się Ci uzyskać jakieś reguły?

9. Reguły uzyskano jedynie dla minimalnego wsparcia 0.01, co przy 1000 zbiorów oznacza, że jedynie 10 zbiorów musi spełniać regułę, aby mogła ona znaleźć się w rozwiązaniu. Można natomiast zauważyć, że uzyskane reguły mają ufność 1, co znaczy, że sprawdzają się w całym zbiorze danych.



Przeanalizujmy uzyskane reguły. Jedną z uzyskanych reguł jest:

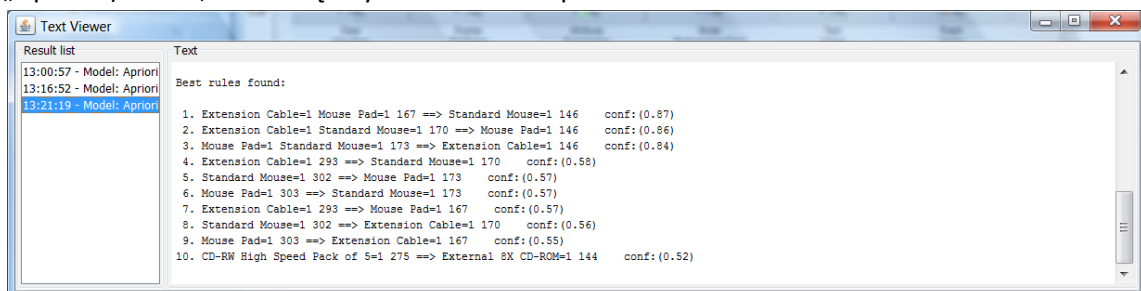
18 Flat Panel Graphics Monitor=1 Extension Cable=1 Model SM26273 Black Ink Cartridge=1 Standard Mouse=1 12 ==> Mouse Pad=1 12 conf:(1)

Reguła ta mówi, że jeżeli klient kupi 18 calowy monitor, przedłużacz, cartridge do drukarki i myszkę, to najprawdopodobniej kupi również podkładkę pod mysz. Liczby 12 przed „==>” oraz na końcu reguły oznaczają wsparcie (liczbę zbiorów) zawierających poprzednik reguły (pierwsza 12stka) oraz oraz zbiorów zawierających zarówno poprzednik, jak i następnik (druga 12stka). Stosunek tych dwóch liczb określa ufność reguły ( $12/12=1$ ), co jest zresztą wskazane w nawiasie („conf:(1)”). Uzyskana reguła ma wysoką ufność – sprawdza się w całym zbiorze danych, ale jak łatwo zauważyć, sprawdza się jedynie w 12 zbiorach na 1000. Taka reguła jest mało interesująca z punktu widzenia marketingu w sklepie, gdyż bardzo rzadko znajduje zastosowanie. Dodatkowo, można mieć pewne wątpliwości co do sensowności odkrytej reguły. Duża bardziej prawdopodobna wydaje się być okrojona wersja wyżej wymienionej reguły:

Standard Mouse=1 ==> Mouse Pad=1

Czyli reguła, według której jeżeli ktoś kupi myszkę, to kupi również podkładkę. Ponieważ z oczywistych powodów nie zawsze taka reguła musi się sprawdzić, nie może ona mieć ufności 1 lub bliskiej. Ponieważ minimalna ufność szukanych reguł dla domyślnych ustawień musi być większa od 0.9, jest to przyczyną, dla której taka reguła nie znalazła się w wyniku.

10. Spróbuj teraz zmniejszyć minimalną ufność szukanych reguł (parametr „minMetric” w konfiguracji „Apriori”) do 0.5, oraz zwiększyć minimalne wsparcie do 0.1.



11. Wśród uzyskanych reguł można znaleźć na 5tym miejscu regułę Standard Mouse=1 ==> Mouse Pad=1, o której wspomniano w punkcie 9. Reguła ta posiada dość duże wsparcie (173, czyli ok. 17% wszystkich transakcji zawieranych w sklepie), ale dość niską ufność: 0.57, co oznacza, że tylko w 57% przypadków, kiedy ktoś kupi myszkę, to kupi również podkładkę. Co jest ciekawe, na drugim miejscu odnaleziono regułę: Extension Cable=1 Standard Mouse=1 ==> Mouse Pad=1, która mówi: jeżeli ktoś kupi przedłużacz i myszkę, to kupi również podkładkę do myszy. Ponieważ reguła posiada ufność 0.86, a zatem w 86% przypadków, jeżeli klient kupi przedłużacz i mysz, to kupi również podkładkę do myszy. Oznacza to, że (o ile reguła jest

prawdziwa) sam fakt kupienia przedłużacza, zwiększa szansę na zakup podkładki do myszki o prawie 30%! Stanowi to świetną okazję do stworzenia promocji: „kup myszkę, przedłużacz dostaniesz za darmo”. Oczywiście cena przedłużacza jest wliczona w koszt myszki, ale firma zyskuje na 30% wzroście zakupów podkładek pod mysz ;).