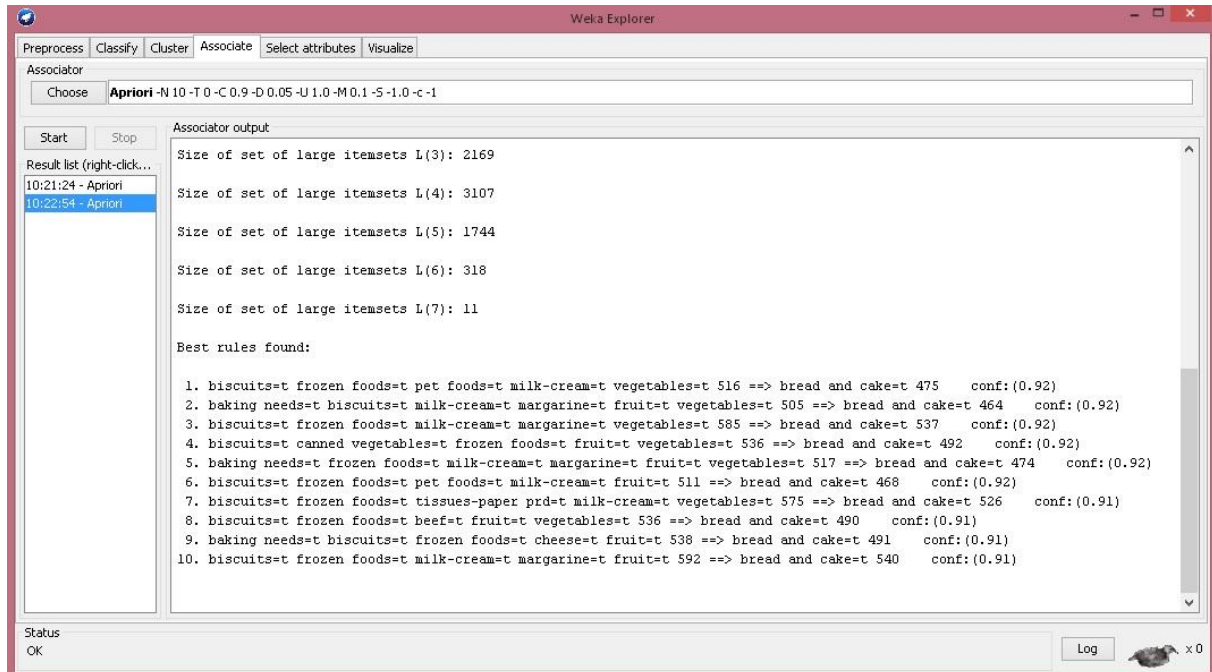


Reguły asocjacyjne – Supermarket (Weka / Explorer)

Zadanie 2.

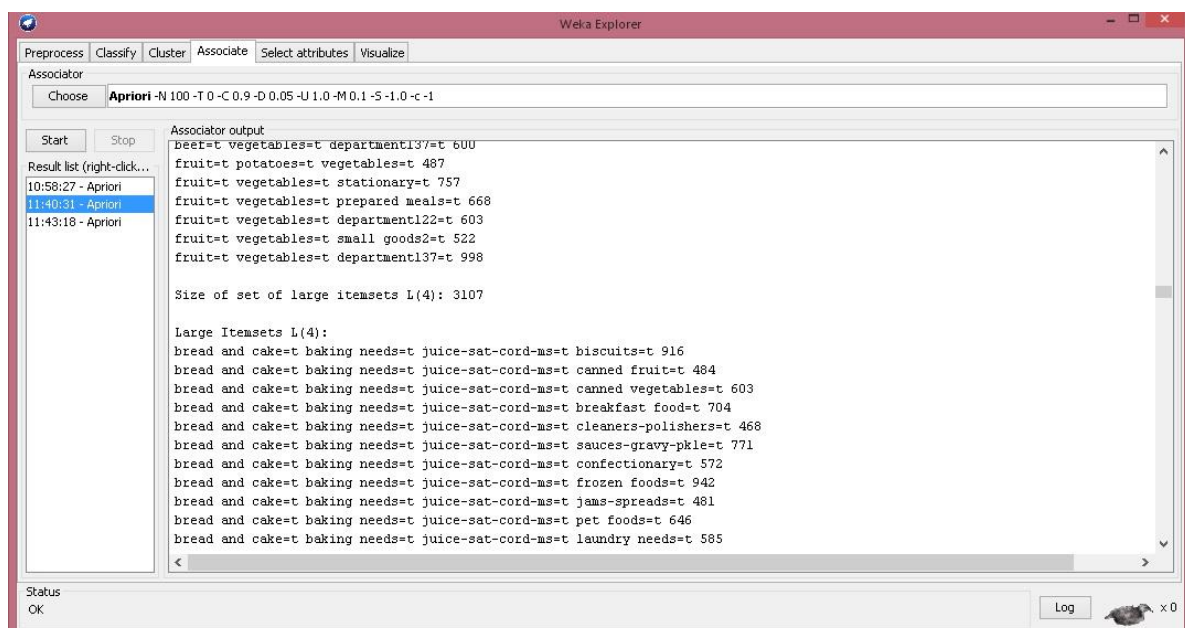
Na początek, uruchamiamy algorytm Apriori z domyślnymi ustawieniami. Dla jakiej kategorii produktów zostały znalezione reguły?

Odp. Reguły zostały odnalezione dla kategorii „bread and cake”



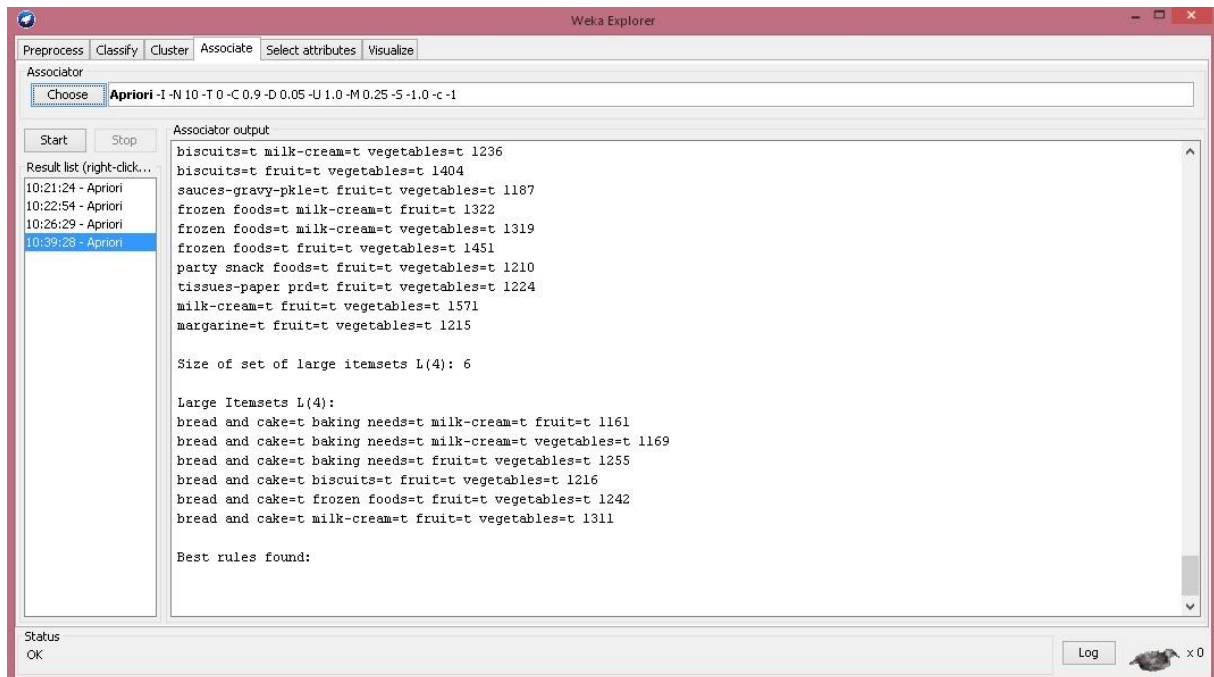
Następnie, klikamy w listę parametrów algorytmu Apriori aby zmienić ustawienia. Ustawiamy opcję `outputItemSets` na „true”, co powoduje wyświetlenie listy zbiorów częstych. Jaka jest największa liczba elementów w zbiorze częstym?

Odp. Największa liczba elementów to 3107



Na liście zbiorów częstych pod nagłówkiem “Large Itemsets L(3)” znajdujemy liczbę wystąpień każdego ze zbiorów częstych. Czy możemy oczekiwać znalezienia zbiorów częstych dla minSupport = 25% (wskazówka: porównaj liczbę wystąpień zbiorów częstych 3-elementowych z całkowitą liczbą transakcji widoczną np. w zakładce Preprocess)?

Odp. Przy ustawieniach minSupport na 25% zostały znalezione zbiory często 1,2,3 oraz 4-elementowe. Nie zostały natomiast znalezione żadne reguły.



Ponownie klikamy w listę parametrów algorytmu Apriori w celu zmiany ustawień. Ustawiamy minimalne wsparcie na 0.25 i uruchamiamy odkrywanie reguł. Następnie robimy to samo dla minSupport = 50% oraz minSupport = 75%. Jakie są wielkości odkrytych zbiorów częstych przy tych ustawieniach? Czy odkryte zostały jakieś reguły?

minSupport	Największy zbiór częsty	????	Liczba reguł	Długość reguły		Confidence	
				min	max	min	max
10%	L4 - 3107	default	10	5	6	0,91	0,92
25%	L2 - 115		0	-	-	-	-
50%	L1 - 9		0	-	-	-	-
75%	brak		0	-	-	-	-

Odp. Tylko w przypadku ostawień minSupport na 10% zostały odnalezione reguły. Przy progu 10%,25% oraz 50% występowały zbiory częste.

Który parametr należy zmienić (i w jaki sposób) jeśli nie zostały odkryte żadne zbiory częste, a który jeśli odkryte zostały zbiory częste, ale nie zostały odkryte żadne reguły? Dobierz wartości tych parametrów tak, aby odkryte zostały reguły asocjacyjne.

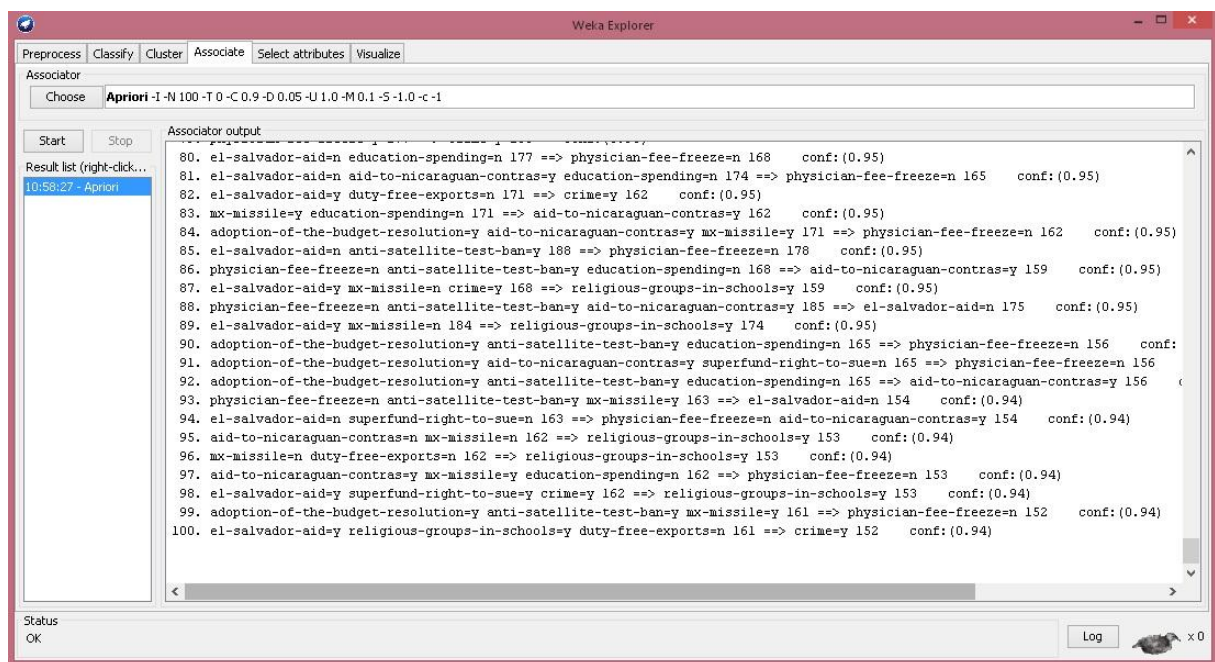
Odp. Aby zostały odkryte reguły oraz zbiory częste wartość minSupport musi być stosunkowo niska. Przykładowo przy poziomie 0,05 można wykryć reguły oraz zbiory częste co przedstawia tabela poniżej.

minSupport	Największy zbiór częsty	????	Liczba reguł	Długość reguły		Confidence	
				min	max	min	max
5%	L4 - 3107	default	100	3	6	0,90	0,92

Reguły asocjacyjne – Votes (Weka / Explorer)

Zadanie 3.

Wykonaj odkrywanie reguł asocjacyjnych w pliku vote.arff. Na początek ustawiamy parametry algorytmu Apriori na wartości domyślne (minSupport = 0.1 oraz minConfidence = 0.9). Ustawiamy liczbę wyświetlanych reguł na np. 100.



Czy wiarygodność odkrywanych reguł jest wysoka? Wyniki których głosowań można przewidzieć jeśli znamy wyniki innych głosowań? Podsumuj swoje obserwacje w poniższej tabeli.

Temat	Głos na tak	Głos na nie
handicapped-infants	187	236
water-project-cost-sharing	195	192
adoption-of-the-budget-resolution	253	171
physician-fee-freeze	177	247
el-salvador-aid	212	208
religious-groups-in-schools	272	152
anti-satellite-test-ban	239	182
aid-to-nicaraguan-contras	242	178
mx-missile	207	206
immigration	216	212
synfuels-corporation-cutback	-	264
education-spending	171	233
superfund-right-to-sue	209	201
crime	248	170

Duty free exports	174	233
Export-administration-act-south-africa	269	-

Wiarygodność reguł oscyluje na poziomie 0,94 do 0,99, jest więc bardzo wysoka.

Na podstawie odkrytych reguł jesteśmy w stanie przewidzieć, że głosy zostaną oddane na następujące kategorie:

crime=y 152,

physician-fee-freeze=n 152

religious-groups-in-schools=y 153

physician-fee-freeze=n aid-to-nicaraguan-contrast=y 154

el-salvador-aid=n 154

aid-to-nicaraguan-contras=y 156

adoption-of-the-budget-resolution=y 152

Edyta Karasek

157062

lwB

Rok 1, semestr 2

Stopień 2