SPRAWOZDANIE

Zajęcia: Zbiory Big Data i Eksploracja Danych

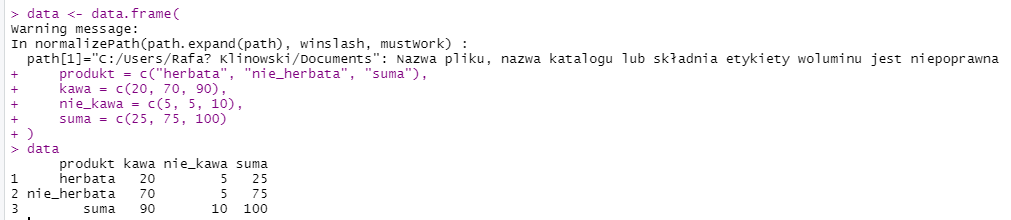
Prowadząca: dr inż. Ruslana Ziubina

|  |  |
| --- | --- |
| Laboratorium nr 5  Data rozpoczęcia: 15.12.2023  Temat: Reguły asocjacyjne | Rafał Klinowski  Informatyka  II stopień, stacjonarne,  Semestr 2, gr. a |

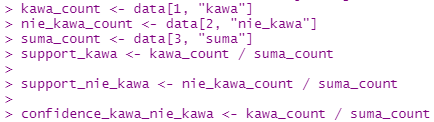
Poszczególne ćwiczenia będą wykonywane w pliku źródłowym edytowanym przy pomocy środowiska RStudio oraz Rattle, opisanego w poprzednich laboratoriach.

# Ćw. 1.

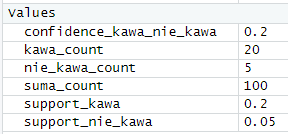
Pierwszym zadaniem było „ręczne” policzenie wsparcia, pewności i liftu dla przykładowego zestawu danych związanych z zakupami kawy i herbaty.



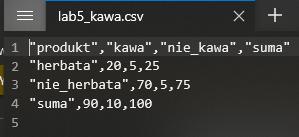
Rysunek . Utworzenie zestawu danych.



Rysunek . Obliczenie wsparcia i pewności dla kawy.



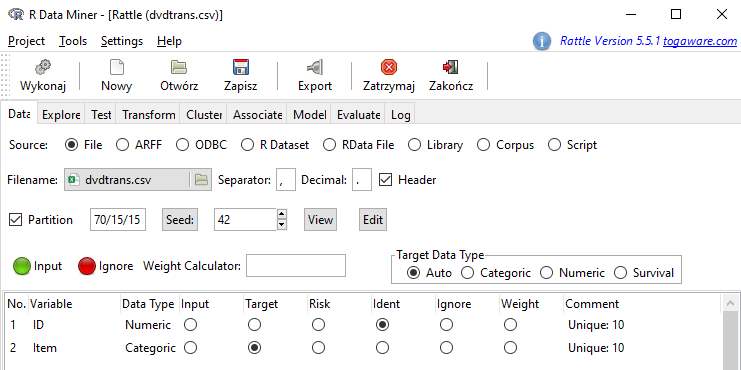
Rysunek . Uzyskane wyniki.



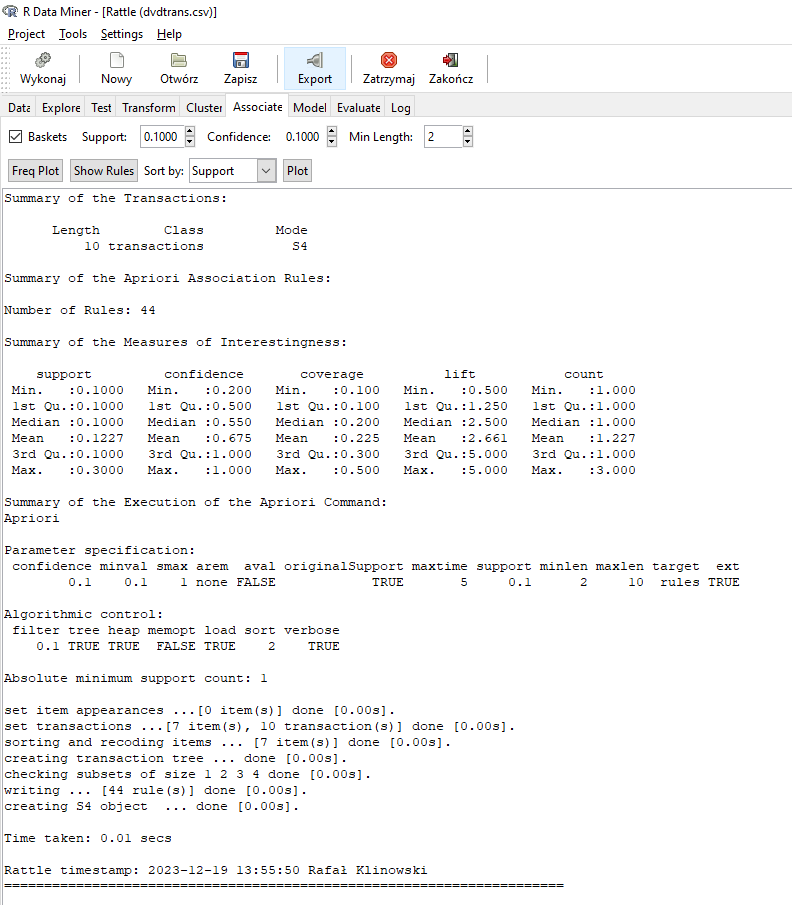
Rysunek . Zbiór danych zapisany do pliku CSV.

# Ćw. 2.

W tym ćwiczeniu przeprowadzono analizę reguł asocjacyjnych w Rattle na podstawie zbioru danych DVDtrans.



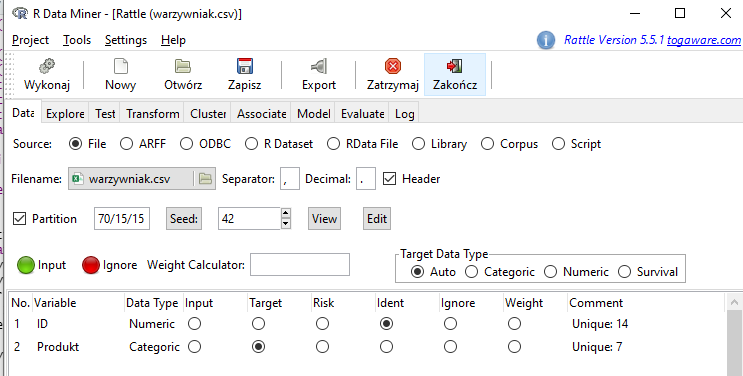
Rysunek . Załadowanie zbioru danych DVDtrans.



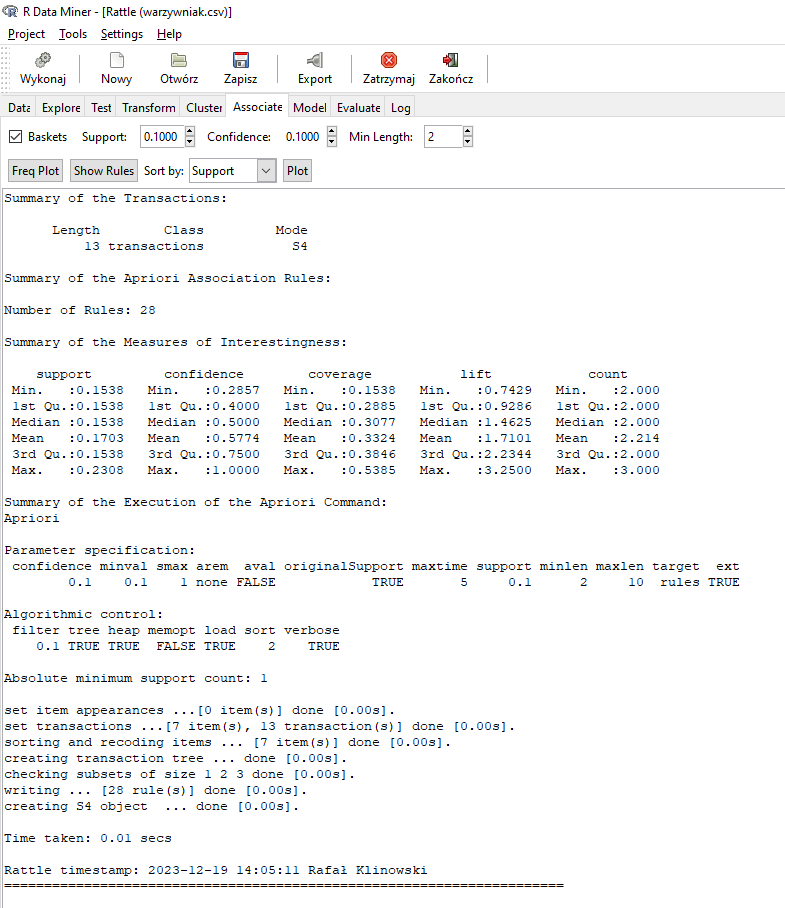
Rysunek . Reguły asocjacyjne dla tego zbioru danych w Rattle.

# Ćw. 3.

Przeanalizowano reguły asocjacyjne dla sklepu warzywnego przy pomocy Rattle.



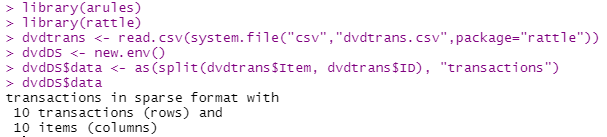
Rysunek . Wczytanie danych z pliku warzywniak.csv.



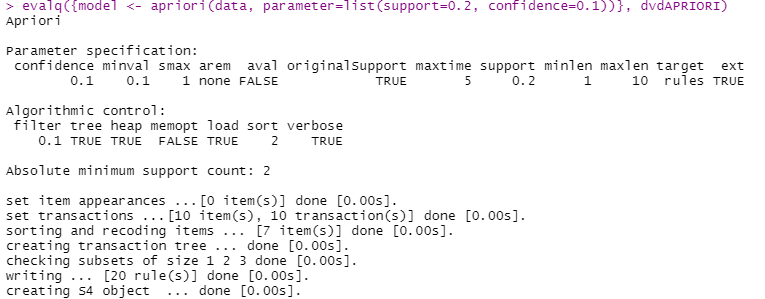
Rysunek . Reguły asocjacyjne wygenerowane dla tego zbioru danych.

# Ćw. 4.

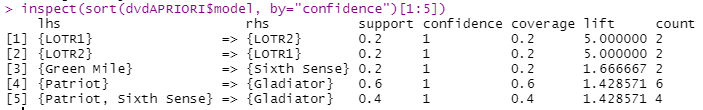
Reguły asocjacyjne w R – wykorzystując zbiór DVDtrans wygenerowano reguły bezpośrednio w R przy pomocy algortymu „apriori”.



Rysunek . Wczytanie danych DVDtrans.



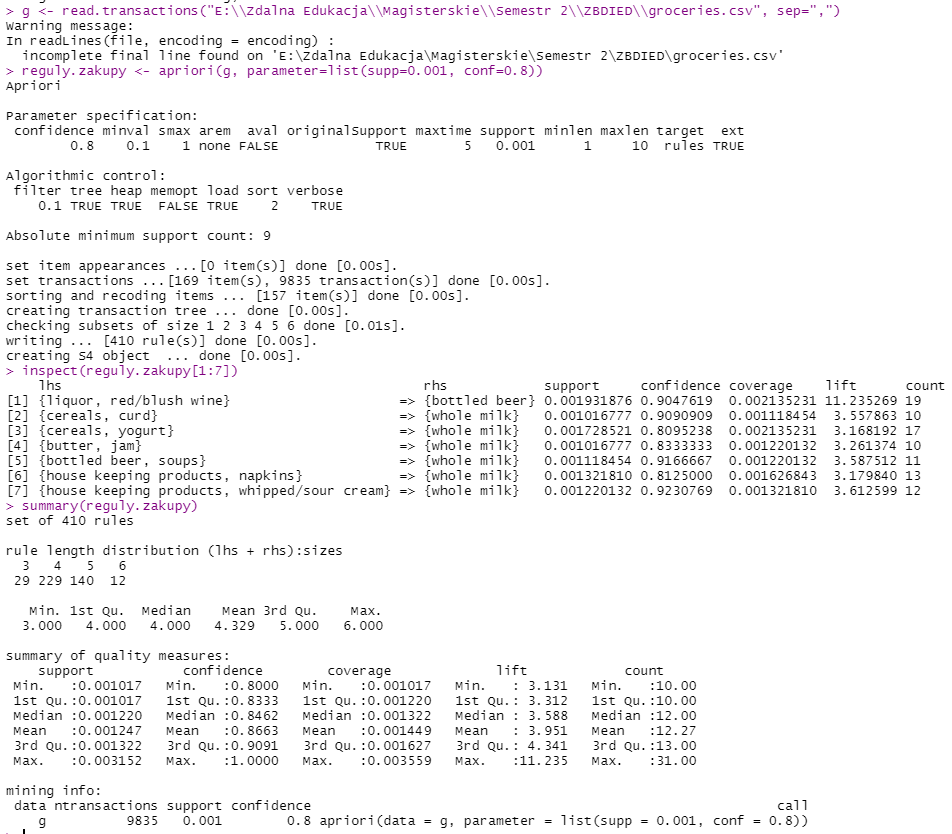
Rysunek . Wygenerowanie reguł asocjacyjnych przy pomocy apriori.



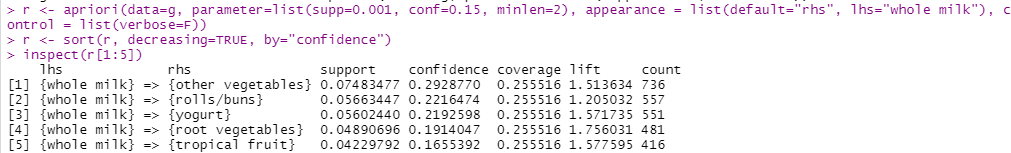
Rysunek . Uzyskane reguły.

# Ćw. 5.

Reguły asocjacyjne w R dla zbioru danych „groceries”. Zapisanie reguł do pliku.



Rysunek . Wczytanie danych i przeprowadzenie analizy reguł asocjacyjnych dla zbioru groceries.csv.

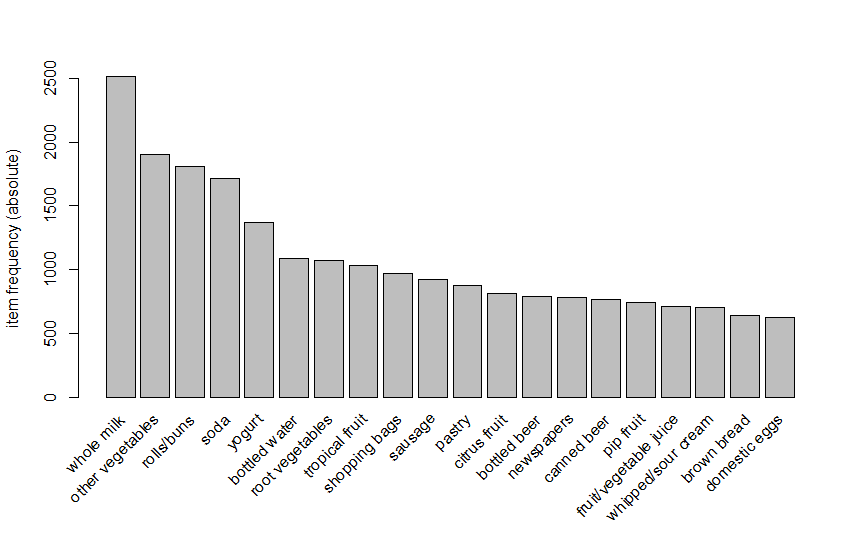


Rysunek . Wyświetlenie posortowanych reguł asocjacyjnych.

# Ćw. 6. – Praca Domowa

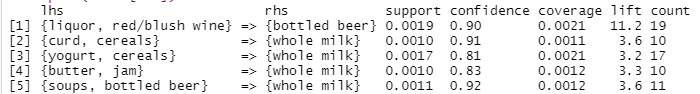
W ramach pracy domowej konieczne było dokonanie analizy reguł dla zbioru groceries, a następnie utworzenia dwóch plików zawierających konkretne reguły.

Najpierw przeprowadzono analizę zgodnie ze stroną „Basket Analysis with R”.

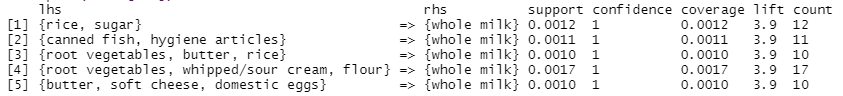


Rysunek . Wykres zawartości koszyka dla danych groceries.

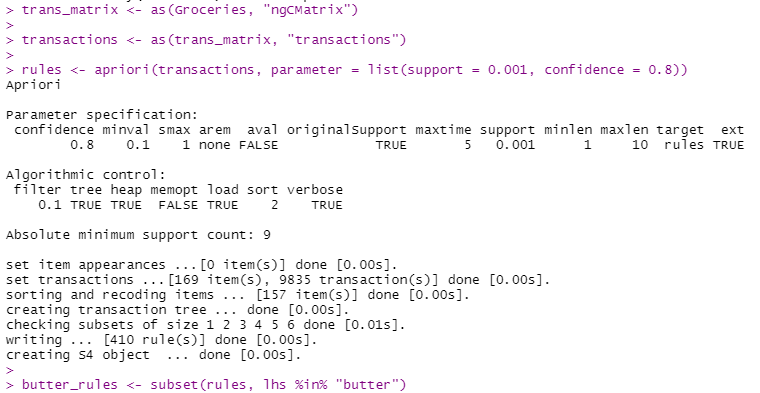
Przeprowadzono analizę apriori dla tego zbioru danych.



Posortowano wyniki po wartości confidence (ufność).

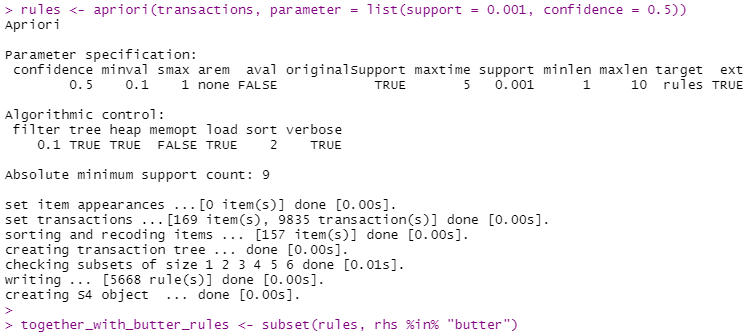


Najpierw należało wskazać reguły mówiące o tym, co kupują klienci, którzy kupili masło. W związku z tym posortowano reguły w taki sposób, aby po lewej stronie (lhs) występowało masło.



Rysunek . Uzyskanie reguł zawierających produkty, w tym masło.

Drugim przypadkiem były reguły wskazujące, które towary kupują klienci wraz z masłem. Wobec tego wskazano, by po prawej stronie (rhs) występowało masło. Przy poziomie ufności 0.8 wskazanym w algorytmie apriori nie uzyskano żadnych wyników – wobec tego powtórzono algorytm z poziomem ufności 0.5.



Rysunek . Uzyskanie reguł zawierających masło po stronie prawej, aby uzyskać inne produkty kupowane wraz z nim.

Oba pliki z regułami zostały załączone wraz z niniejszym sprawozdaniem.

# Wnioski.

Środowisko Rattle umożliwia w prosty sposób pracę nad danymi, w tym tworzenie wykresów, eksplorację danych czy uzyskiwanie podsumowań. Rattle zawiera również sporą ilość narzędzi wycelowanych w uczenie maszynowe czy grupowanie danych. Środowisko współpracuje z wieloma dodatkowymi pakietami i zawiera przejrzysty interfejs, w którym dość łatwo znaleźć wszystkie interesujące nas opcje. W szczególności pozwala ono w łatwy sposób przeprowadzać analizę reguł asocjacyjnych bez konieczności ręcznego wpisywania poszczególnych poleceń z odpowiednimi parametrami.

Podczas analizy reguł asocjacyjnych w praktyczny sposób dowiedziałem się, jak taką analizę przeprowadzać oraz jaka jest jej użyteczność. W szczególności dobrym przykładem okazała się analiza zbioru Groceries, gdzie w praktyczny sposób udało się zapoznać z sensem i interpretacją takich reguł.