

# Implementacja algorytmu do odkrywania reguł asocjacyjnych

Bartosz Latosek





# 1. Omówienie problemu

- Pobieranie i preprocessing danych
- Implementacja algorytmu Apriori
- Implementacja sposobu na wyznaczanie miar
- Przeprowadzenie testów na różnych zbiorach danych
- Porównanie wyników



## 2. Technologie

Projekt zostanie w całości zrealizowany w środowisku Python. Stworzony zostanie skrypt konsolowy, na podstawie którego przeprowadzone zostaną badania. Przy implementacji rozwiązania zachowane zostaną wszystkie najnowocześniejsze i specyficzne dla języka Python techniki i paradygmaty programowania.



## 3. Dane

Dane będą pobierane z **UCI Machine Learning Repository** (<https://archive.ics.uci.edu>) za pośrednictwem biblioteki *ucimlrepo*. Zostanie stworzona klasa udostępniająca interfejs pozwalający na pobranie jednego z wybranych zbiorów danych. Z początku wybrane zostaną dyskretne zbiory danych, odpowiednie do problemu. W dalszej części rozwiązania, jeżeli będzie to wymagane przez eksperymenty, stworzone zostanie rozwiązanie pozwalające na dyskretyzację dowolnych danych.



## 4. Wybrany algorytm

Do zadania wybrany został algorytm Apriori. Zostanie on zaimplementowany jako klasa udostępniająca interfejs pozwalający na wyznaczanie reguł asocjacyjnych.



## 5. Wybrane miary

- Wsp. podniesienia (lift)
- Współczynnik pewności (certanity factor)
- Jaacard / coherence
- iloraz szans (odds-ratio)



## 6. Testy

Celem testów projektu jest porównanie zachowania się wybranych miar w odniesieniu do wartości współczynnika podniesienia. Algorytm zostanie uruchomiony na 3 wybranych zbiorach danych, z których wybrane zostaną reguły asocjacyjne o różnych (zbliżonych) wartościach wsp. podniesienia - lift. Następnie zbadana zostanie zależność między wartościami wybranych miar a wartością wybranego wsp. podniesienia lift. Wnioski zostaną dokładnie opisane i zestawione w sprawozdaniu końcowym.