

Generowanie reguł asocjacyjnych Danologia, Laboratorium

1. Podaj nazwy produktów i wsparcie dla najliczniejszego zbioru częstego spośród zbiorów częstych o minimalnym wsparciu równym 30%. Jeżeli istnieje wiele równolicznych zbiorów, podaj ten o największym wsparciu. Wyniki liczbowe zaokrąglaj do 2 miejsc po przecinku. Zinterpretuj uzyskane wyniki. Podaj kod, dzięki któremu uzyskałeś podane wyniki.
2. Wyrysuj histogram występowania produktów w bazie. Które z produktów występują ponad 2000 razy w transakcjach? Zinterpretuj wyniki. Załącz screeny w formie obrazu *.PNG oraz kod który wykorzystałeś aby wyrysować histogram. Możesz skorzystać z funkcji zdefiniowanych w ramach zajęć.
3. Wygeneruj zbiory często o minimalnym wsparciu 30%, na podstawie tych zbiorów wygeneruj reguły o minimalnej wartości ufności (*confidence*) równej 0.84. Podaj uzyskane reguły i zinterpretuj je. Czy patrząc na wsparcia poprzedników i następników implikacji, jak również na miarę *lift* można powiedzieć, że te reguły są wartościowe? Podaj kod, dzięki któremu uzyskałeś podane wyniki. Wyniki liczbowe zaokrąglaj do 2 miejsc po przecinku.
4. Dla jakich kombinacji produktów, wartość rachunku jest wysoka (jakie jest X z reguły $X \rightarrow \{high\}$)? Zawęż poszukiwania poprzez wygenerowanie w pierwszej kolejności zbiorów częstych o minimalnym wsparciu 20%, a następnie, wygenerowanie reguł o ufności minimalnej 0.55. Dla tak otrzymanych reguł wybierz tylko te, które mają jako następnik implikacji jedynie element high. Podaj i zinterpretuj uzyskane wyniki. Jak oceniasz, czy te reguły są wartościowe (porównaj wartość *lift*)? Liczby zaokrąglaj do 2 miejsc po przecinku. Zamieść kod, dzięki któremu uzyskałeś podane wyniki.