

Poznań 7 listopada 2008

### CASE STUDY 3

## **Analiza danych medycznych nt. leczenia pacjentów cierpiących na kamice nerkową.**

(*Copyright Jerzy Stefanowski - , Instytut Informatyki PP  
zastrzeżenia dotyczą opisu problemu i ograniczonej dostępności do danych*)

### **Wprowadzenie**

Załóżmy, że współpracujesz ze szpitalem lokalnego uniwersytetu w ramach zastosowań metod informatycznych w medycynie; w szczególności analizy danych. Oddziały tegoż szpitala prowadzą działalność nie tylko w zakresie leczenia pacjentów, szkolenia studentów ale także badawczą. W działalności badawczej często poszukuje się odpowiedzi na temat skuteczności nowych leków, wprowadzania nowych procedur diagnostycznych i terapeutycznych. Ma to znaczenie zarówno badawcze jak i szkoleniowe, np. dla studentów czy "młodszych" lekarzy stażystów.

W ostatnim czasie zwrócił się do Ciebie z prośbą o współpracę - konsultację ordynator jednego z oddziałów. Celem była analiza doświadczenia klinicznego w zakresie leczenia kamicy moczowej metodą kruszenia kamieni za pomocą techniki fali uderzeniowej generowanej poza organizmem (ang. ESWL). Poszukiwano modeli charakteryzujących stan pacjentów po zabiegu (bez komplikacji lub z komplikacjami) oraz wspomagających długoterminowe prognozowanie wyniku leczenia. Opis problemu medycznego w języku angielskim zawarty jest w dodatkowym pliku.

Dane zbierano na podstawie wypełniania ankiet związanych z wykonywanymi zabiegami. Dotyczą one 500 pacjentów opisanych za pomocą 33 atrybutów (nominalnych lub porządkowych). Dostępne są dwie możliwości zdefiniowania klasyfikacji pacjentów. Klasy decyzyjne są względnie zrównoważone

Problem, jaki bada zespół lekarski, dotyczy poszukiwania istotnych zależności w danych i regularności pomiędzy atrybutami, a także ustalania hierarchii ważności atrybutów opisujących pacjentów. Zakładając, że jakość danych jest wystarczająca do oceny stanu pacjentów po zabiegu lub przewidywania przebiegu leczenia (obie klasyfikacje) można dalej postawić pytanie o możliwość zbudowania symbolicznej reprezentacji wiedzy klasyfikacyjnej (pokazującej jawnie związek między wartościami najlepszych atrybutów o klasyfikacją pacjentów).

Dotychczasowa opinia o tych danych wskazuje, że nie jest to „łatwy” zbiór danych, w którym można się spodziewać jednoznacznych wniosków i odnalezienia wysoce skutecznych klasyfikatorów.

W załącznikach otrzymujesz (w języku angielskim - z uwagi na przygotowywaną pracę o charakterze raportu):

## 1. Opis problemu medycznego

## 2. Plik z danymi (format isf)

Powinieneś pamiętać, iż nie masz wpływu na rozmiar dostępnych danych, nie możesz oczekiwać dostarczenia dodatkowych opisów pacjentów; zostało to wykonane przed Twoim udziałem w studium badawczym; Tzn. nie będziesz miał dodatkowych obserwacji lub wprowadzenia dodatkowych atrybutów.

Natomiast, jeśli potrafisz ocenić jakość otrzymanych danych możesz dokonywać przeskalowań lub przedefiniowań atrybutów (np. tworzyć nowe w oparciu o pomierzone), jeśli ich końcowa postać jest akceptowalna dla potencjalnego użytkownika (czytaj ma potencjalnie dogodną interpretację medyczną).

Inne uwagi metodyczne:

Oplaca się badać jakość dostarczonych danych (były one zbierane przez osoby, które nie znają podstaw Twoich metod eksploracji danych);

- Interesujące jest badanie wzajemnych współzależności tkwiących w danych;
- Nie istnieje jasno wyrażony atrybut decyzyjny, co więcej może być ich wiele.
- Warto stosować więcej niż jedną metodę eksploracji danych (ukierunkowanych na różne formy wiedzy i różne ich reprezentacje)
- Uwaga z powodu silnego eksploatowania na zajęciach dotychczas technik budowy klasyfikatorów, zwracam uwagę że automatyczna klasyfikacja pacjentów nie jest zawsze akceptowana przez wielu lekarzy; może być używana jednak jako miara pewności czy wiarygodności wyników;
- Jeśli chcesz rozważać także budowę klasyfikatorów, Lekarze nie skupiają się na globalnej trafności klasyfikacji lecz ważniejsza jest dla nich trafność w poszczególnych klasach, w szczególności osób w gorszym stanie (analiza "confusion matrix" jest bardzo pożądana).