Predviđanje srčanog napada u R programskom jeziku

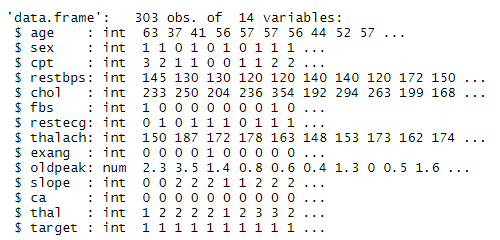
1. Analiza skupa podataka

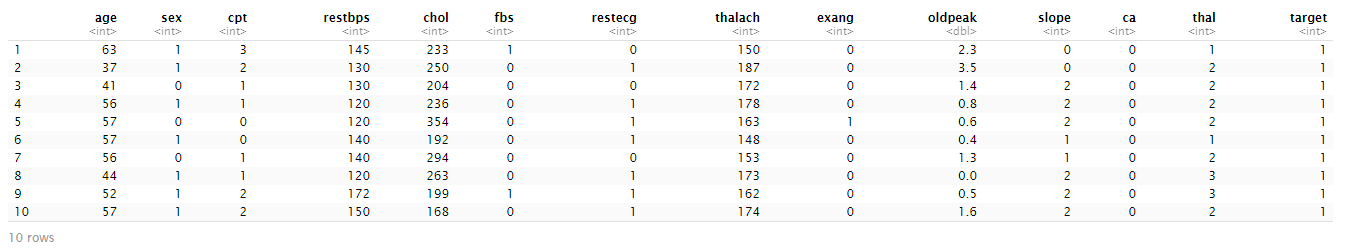
Skup podataka je vezan za dijagnostičke podatke i analiza krvi i srca. Na osnovu ovih podataka se može izvršiti predikcija ili verovatnoća da će osoba imati srčani napad. Svi atributi su već bili brojčane vrednosti tako da nije bilo potrebe za numeričkom kategorizacijom atributa ili dodeljivanjem nekih numeričkih vrednosti, kao niti bilo kakvog skaliranja vrednosti.

Srčani napad je uzrokovan otkazom dela srčanog mišića usled prestanka dovoda krvi. Dovod krvi je obično prekinut usled stvaranja ugruška u krvi u arteriji koja snabdeva srčani mišić. Ako neki od delova srčanog mišića otkaže, osoba oseća jak bol u grudima i električnu nestabilnost tkiva srčanog mišića.

**Atributi:**  
1) age – **godine starosti,**  
2) sex – **pol (1 = muško, 0 = žensko),**  
3) cpt = chest pain type (4 vrednosti) – **tip bola u grudima,**  
4) restbps = resting blood pressure – **krvni pritisak u mirovanju,**  
5) chol = serum cholestoral in mg/dl – **holesteralni serum,**6) fbs = fasting blood sugar > 120 mg/dl – **visok šećer u krvi,**  
7) restecg = resting electrocardiographic results (vrednosti 0,1,2) – **ostalih elektrokardiografskih rezultata,**8) thalach = maximum heart rate achieved – **maksimalni broj otkucaja srca,**  
9) exang = exercise induced angina – **angina izazvana vežbanjem,**  
10) oldpeak = ST depression induced by exercise relative to rest – **ST depresija izazvana vežbanjem u odnosu na odmor,**  
11) slope = the slope of the peak exercise ST segment   
12) ca = number of major vessels (0-3) colored by flourosopy  
13) thal: 0 = **normalno**; 1 = **fiksni defekt**; 2 = **reverzibilni defekt**  
14) target: 0= **manja šansa za srčani napad**; 1= **veća šansa za srčani napad**

Ispod je prikaz strukture skupa podataka kao i prvog head skupa.





Na osnovu predstavljenih podataka, može se izvršiti predikcija ili verovatnoća za šansu dobijanja srčanog napada upotrebom stabala odlučivanja (Random Forest), Metoda nosećih vektora, Logističkom regresiju ili kako su objavljeni rezultati na internetu najbolje daje Neuronska mreža što je oko 87% tačnosti predikcije na osnovu datih podataka.