

Istraživanje podataka - vežbe 4

1. Klasifikacija nad studentskim tabelama

Klasifikovati ispite prema oznaci ispitnog roka. Za sve polagane ispite izdvojiti oznaku ispitnog roka, godinu ispitnog roka, naziv smera studenta koji je polagao ispit, naziv predmeta koji je polagan i dobijenu ocenu. Ako je oznaka roka oznaka klase, primeniti klasifikaciju po ostalim atributima bez godine roka.

2. Klasifikacija nad tabelom Adult

Zadatak:

- Uraditi data exploration
- Da li se na osnovu histograma može pretpostaviti zavisnost klase od nekog od atributa?
- Uraditi klasifikaciju i diskutovati rezultate
- Da li ima preprilagodjavanja (overfitting)?
- Koji su atributi važni za klasifikaciju?
- Da li se to slaže sa predviđanjima na osnovu histograma?
- Proveriti korelacije
- Kolika se preciznost klasifikacije može postići koristeći samo mali skup važnih atributa?
- U promociji nekog proizvoda marketinška kompanija cilja dobrostojeće ljude. Pri tom, dosta je važnije da neki dobrostojeći covek ne bude propušten nego da li će akcija uključivati i nekog lošije stojećeg. Uraditi cost sensitive klasifikaciju i diskutovati rezultate.