# Otkrivanje Zakonitosti u Podacima

**Zahtevi za drugi domaći zadatak**

1. Učitati podatke o vinima (wine\_quality.csv), sa ciljem prepoznavanja "top\_quality" klase. 1.2. Kreirati minimalno 3 modela za klasifikaciju (sa default parametrima) - proceniti grešku generalizacije za sve modele, na osnovu nekoliko metrika - prikazati i standardnu devijaciju procene greške (od 10 kros-validacionih uzoraka) - odgovoriti da li su modeli pretrenirani; - prodiskutovati koja metrika je najadekvatnija za ovaj problem, i koji tip grešaka je skuplji. modifikovati granicu odlučivanja, tako da model pravi više jeftinijih grešaka

2. Učitati insurance.csv skup podataka.

2.1. Kreirati tok podataka (eng. pipeline) u kome ćete izvršiti:

- Pretvaranje kategoričkih u numeričke atribute

- Popunjavanje nedostajućih vrednosti

- Modelovanje (uz pomoć regresionog stabla odlučivanja)

2.2. Korišćenjem kros validacije oceniti kvalitet regresionog modela (na osnovu RMSE) koji je kreiran kroz prethodno definisan tok podataka.

2.3 Korišćenjem rezultata iz 2.1 i 2.2. optimizovati regresiono stabla odlučivanja (minumum 2 parametra i minimum 10 modela treba da bude evaluirano). Da li je optimizacija poboljšala model i ako jeste koliko?

2.4. Odaberite najbolji model