

Klasterizacija – zadatak

1. Pronaći pogodan skup podataka u nekoj od baza podataka sa otvorenim pristupom (ne koristiti skupove podataka date u primerima za računske vežbe):

<https://archive.ics.uci.edu/datasets>

<https://www.kaggle.com/datasets>

<http://www.cs.toronto.edu/~dave/data/datasets.html>

2. Po potrebi prilagoditi pronađeni skup podataka.
3. Korišćenjem biblioteka otvorenog koda izvršiti K-Means algoritam za klasterizaciju nad odabranim podacima. Analizirati dobijene rezultate i pokazati neke od zavisnosti između atributa.
4. Korišćenjem biblioteka otvorenog koda izvršiti algoritam za hijerarhijsku klasterizaciju nad odabranim podacima. Analizirati dobijene rezultate i pokazati neke od zavisnosti između atributa.
5. Kao rešenje zadatka treba predati napisani programski kod, korišćene skupove podataka i kratko objašnjenje dobijenih rezultata (dokument od 1-2 strane). Ukoliko zbog veličine nije moguće postaviti skup podataka na CS, onda treba postaviti samo link u nekom fajlu do skupa podataka.