

1 Podatki.

Dobili smo podatke o državah tekmovalkah na Evroviziji od leta 1998 do leta 2009 in sicer 291 primerov, vsakega s 63 atributi. V nalogi je bilo bistveno analizirati glasovanja med državami, kar pomeni, da je bilo nekaj nezanemljivih atributov kot so ime avtorja, naslov pesmi, jezik petja in nato še nekaj podatkov o avtorju in državi gostiteljici - bolj natančno atributi od 3. do 16. stolpca.

1.1 Interpretacija in čiščenje podatkov.

Ugotoviti smo morali pristranskost glasovanja držav, zato sem seštel vse glasove skozi leta in sestavil slovar oblike {'država': ['1.0', '2.0', ... '0.0']} in ga nato transponiral. Manjša težava so bili le pomanjkljivi podatki zaradi domnevno nesodelovanja. Manjkajoča polja sem nadomestil z '0.0', saj skoraj ne vpliva na končen rezultat.

2 Implementacija algoritma.

Skupine sem združeval na podlagi Evklidske razdalje, skupke skupin pa po razdalji med povprečjem vseh primerov (angl. avg. linkage), ker sem domneval, da je najenostavnejša za implementirati v Pythonu.

3 Rezultati.

Zaradi velikosti dendrograma sem ga vključil v datoteko Drevo.txt.

3.1 Skupine.

Skupina 01: Croatia, Slovenia, Bosnia and Herzegovina, Macedonia, Montenegro

Skupina 02: Greece, Russia, Poland, Andorra, Portugal, Armenia, Belarus, Hungary, Czech Republic, Moldova, Serbia, Monaco, Azerbaijan, San Marino, Slovakia, Romania, Albania, Bulgaria, Spain, Cyprus, Israel **Skupina 03:** Estonia, Latvia, Lithuania, Ireland, Malta, Iceland, Norway, Denmark, Finland

Skupina 04: France, Germany, Belgium, Netherlands

Skupina	Preferirana država	Nepreferirana država
Skupina 01:	Czech Republic	Azerbaijan
Skupina 02:	Bolgaria	Bosnia and Herzegovina
Skupina 03:	Cyprus	Croatia
Skupina 04:	Hungary	Albania

3.2 Razlaga.

Skupine sem izbral na podlagi dendrograma, kot je profesor povedal na predavanjih sem uporabil navidezno črto, ki je razdelila dendrogram na 5 skupin. Preferirane države sem izbral po metodi $pref_nepref(hc)$.