

**Uvod.** Pri tej nalogi smo analizirali glasovanja za Pesem Evrovizije. Predstavniki posameznih držav naj bi glasovali pristransko in pri tem favorizirali nastopajoče iz bližnjih ali sorodnih držav. Cilj naloge je bil preveriti ali to drži.

**Podatki.** Podatki o glasovanjih za Pesem Evrovizije so bili podani v obliki csv datoteke. V datoteki je bilo zapisanih okol 33 tisoč vrstic. Vsaka vrstica je vsebovala leto glasovanja, tip glasovanja, katera država je glasovala, za katero državo je glasovala ter koliko točk ji je podelila. Točke posameznega glasovanja so se gibale med 1 in 12. Ker je bil razpon let med 1975 do 2019, so se nekatere države pojavile večkrat kot druge.

**Luščenje profilov glasovanj.** Datoteko s podatki sem prebral dvakrat. Ko sem prvič šel čez vse vrstice, sem določil vse attribute oz. stolpce. Vsak stolpec je predstavljal svoje leto glasovanja ter državo za katero je bilo glasovano. Tako sem dobil čez tisoč različnih atributov, po katerih sem profile kasneje primerjal. Ko sem drugič bral datoteko, sem za vsako državo naredil slovar, ki je vseboval prej omenjene attribute, ter vnesel vrednosti le teh. Tako pripravljen slovar, sem nato uporabil pri gručenju.

**Gručenje.** Gručenje je postopek razvrščanja profilov, v našem primeru profilov glasovanj, v skupine podobnih profilov. Sestavljen je iz: računanja razdalje, ki predstavlja podobnost dveh profilov, računanja razdalje med gručami profilov, iskanjem najkrajših razdalj med gručami profilov ter združevanje le teh.

**Računanje razdalje med profili.** Računanje razdalje med profili mi je predstavljalo največjo oviro pri gručenju. Razdaljo sem računal po evklidskem izreku. Problem se je pojavil, ko profili niso imeli vseh atributov. To sem rešil tako, da po takih atributih preprosto nisem primerjal in na koncu razdaljo primerno normaliziral. Če se je zgodilo da profila nista imela nobenih skupnih atributov, sem vrnil vrednost -1, da sem pri računanju razdalje med gručami vedel da le te razdalje ne smem upoštevati.

```

# racunanje razdalje med profili
# r1 ter r2 sta imeni drzav
def row_distance(self, r1, r2):

    row1 = self.data[r1]
    row2 = self.data[r2]

    sum_of_attributes = 0
    attributes_counter = 0
    for i in range(0, len(row1)):
        val1 = row1[i]
        val2 = row2[i]
        # tu preverimo ce atribut obstaja pri obeh profilih
        if val1 is not None and val2 is not None:
            sum_of_attributes += pow(val1 - val2, 2)
            attributes_counter += 1

    if attributes_counter == 0:
        return -1
    # normalizacija
    return math.sqrt((sum_of_attributes / attributes_counter) * len(row1))


```

**Računanje razdalje med gruči.** Razdaljo med gruči sem računal z uporabo metode "Average Linkage", ki izračuna povprečno razdaljo med dvema gruči. To naredi tako, da izračuna razdalje med vsemi profili v obeh gručah in jih povpreči. Kot že prej napisano, če je bila razdalja med profiloma -1, je pri povprečenju nisem upošteval.

**Iskanje najpodobnejših gruči** Iskanje najkrajše razdalje med dvema gruči je bilo tedaj preprosto. Izračunati sem mogel zgolj razdalje med vsemi možnimi kombinacijami gruči ter ugotoviti katera je najkrajša.

**Združevanje v gruče.** Gruče je bilo tedaj enostavno zgraditi. Potrebna je bila zgolj ena while zanka, ki je poiskala najpodobnejše gruče in jih združevala dokler ni ostala ena sama gruča.

**Dendrogram.** Naloga je zahtevala tudi izris dendrograma. Na dendrogramu lahko vidimo razdalje med posameznimi profili in sami poizkusimo določiti število gruči.

 slika-dendrogram.png

Slika 1: Dendrogram.

**Rezultati.** Sam sem določil 11 gruč oz. skupin, ki vsebujejo države, ki so podobno glasovale skozi leta 1975 do 2019. V tabeli 1 lahko za vsako skupino vidimo 3 najbolj preferirane države ter 3 države katerim je namenila najmanj točk.

**Skupina 1.** San Marino

**Skupina 2.** Australia

**Skupina 3.** North Macedonia

**Skupina 4.** Italy, Morocco

**Skupina 5.** Serbia, Luxembourg, Slovenia, Montenegro, Croatia, Bosnia Herzegovina, F.Y.R. Macedonia, Serbia & Montenegro

**Skupina 6.** Poland, Estonia, Lithuania, Latvia, Ireland, United Kingdom, Finland, Denmark, Norway, Sweden, Iceland

**Skupina 7.** Spain, Portugal, Andorra, Romania, Israel, Moldova

**Skupina 8.** Armenia, Bulgaria, Greece, Cyprus

**Skupina 9.** Azerbaijan, Turkey, Czech Republic, Georgia, Ukraine, Russia, Belarus

**Skupina 10.** Yugoslavia, Monaco, Hungary, France, Belgium, The Netherlands, Switzerland, Germany, Austria

**Skupina 11.** Albania, Malta, Slovakia

Tabela 1: Gruče ter njihove značilnosti.

skupina	preferirane države	ne glasujejo
Skupina 1	Italy, Greece, Russia	Serbia & Montenegro, Montenegro, San Marino.
Skupina 2	Sweden, Belgium, Bulgaria	Montenegro, San Marino, Australia.
Skupina 3	Albania, Italy, The Netherlands	Georgia, Montenegro, North Macedonia.
Skupina 4	France, United Kingdom, Spain	Serbia & Montenegro, Montenegro, San Marino.
Skupina 5	Serbia, Bosnia & Herzegovina, Croatia	Luxembourg, Morocco, Andorra.
Skupina 6	Sweden, Norway, Denmark	San Marino, Morocco, Andorra
Skupina 7	Sweden, Italy, Germany	Morocco, Andorra, Montenegro
Skupina 8	Greece, Russia, Cyprus	Slovakia, Morocco, Andorra
Skupina 9	Russia, Ukraine, Azerbaijan	Morocco, Slovakia, Andorra
Skupina 10	Sweden, United Kingdom, Israel	Slovakia, Andorra, Montenegro
Skupina 11	Italy, Greece, Sweden	Monaco, Morocco, Andorra

**Izjava o izdelavi domače naloge.** Domačo nalogo in pripadajoče programe sem izdelal sam.