

Uvod. Cilj naloge je, da s primerjavo podatkov iz preteklih Evrovizijskih glasovanj in metodo hierarhičnega razvrščanja ugotovimo, ali države na Evroviziji res glasujejo pristrasno glede na svojo geografsko lego. Cela programska koda naloge se nahaja v datoteki "naloga1.py".

Podatki. V tej nalogi sem za bazo podatkov vzel podatke iz preteklih Evrovizijskih glasovanj, ki so dostopni na portalu Kaggle, hkrati pa priloženi v .zip datoteki (eurovision-final.csv). Navedena datoteka vsebuje veliko podatkov, ki v naši raziskavi ne koristijo preveč (kot na primer naslov pesmi) in jih zato nisem uporabil. Uporabil pa sem podatke od 16. do 63. stolpca, ki nam povejo koliko točk je katera država namenila komu na Evroviziji (profil glasovanja je torej vektor točk, ki jih je država podelila). Vrednosti segajo od 0 do 12 točk, če država tisto leto na Evroviziji ni nastopila, ali pa v primeru, ko bi morala glasovati sama zase, je podatek neznan. Delež neznanih podatkov na državo precej varira, saj se nekatere države Evrovizije redno udeležujejo, nekatere pa ne. Pričakovan rezultat je, da bodo države svojim geografskim sosedam namenile več točk kot tistim, ki so od njih bolj oddaljene.

Računanje razdalj. Pred razlago računanja razdalj med profili in skupinami, moram najprej omeniti kako sem ravnal z neznanimi vrednostmi. V primeru, da država neko leto na Evroviziji ni nastopila, sem namesto vrednosti od 0 do 12 podatke za tisto leto nadomestil z "None". V primeru, da je država tisto leto na Evroviziji nastopila in je razlog, da podatka ni, ker država nesme glasovati sama zase, sem namesto "None" to vrednost nadomestil z 12. Razlog za tem je, da če bi država lahko glasovala sama zase bi si gotovo podelila maksimalno število točk, ki je v našem primeru 12.

Razdalje med posameznimi profili sem računal s funkcijo "row_distance", ki po spodaj navedeni formuli izračuna razdaljo med dvema vektorjema.

```
dist = math.sqrt(sum([(a - b) ** 2 for a, b in zip(x, y)
                      if not a is None and not b is None]))
```

If stavek, ki se nahaja v drugi vrstici zgornje kode, se navezuje na neznane vrednosti (None). V primeru, da je bila ena od vrednosti neznana, jih med sabo nisem primerjal saj to nebi bilo reprezentativno glede na druga leta. V primeru da državi nobeno leto nista skupaj nastopili na Evroviziji (kar po moji metodi računanja pomeni, da nobene njune vrednosti iz vektorja nisem primerjal), sem skupno razdaljo med njima nastavljal na 55. Razlog za to je, da o državah na podlagi danih podatkov ne vemo nič, 55 pa je povprečna razdalja med pari vseh držav.

```
if dist>0:
    return dist
else:
    return float(55) #povprecje
```

Razdalje med posameznimi skupinami pa sem izračunal tako, da sem primerjal najbolj oddaljena profila v skupini:

```
max=0
for a1 in arr1:
```

```

for a2 in arr2:
    nov = self.row_distance(a1, a2)
    if nov>max:
        max=nov
return max

```

Dendrogram. Izris dendrograma je na voljo v prilogi 1 kot tudi v tekstovni datoteki (dendrogram.txt).

Skupine in njihove preferenčne izbire. Glede na rezultate razvrščanja sem se odločil profile (države) razdeliti v 7 skupin. Lahko bi jih razbil tudi na več skupin in bi bile še bolj specifične (severna skupina razpade na noridjsko in pribaltsko skupino), vendar je razdalja med skupinama precej majhna zato sem se odločil ustaviti pri sedmih. Skupine sem torej določil na podlagi razdalj med njimi s funkcijo `cluster_distance`. Vidimo, da so skupine dejansko precej logično izoblikovane in bi lahko trdili, da obstaja neka korelacija med geografsko lego in glasovanjem na Evroviziji. Imena skupin in njihovih profilov so:

Balkan. Albania, Macedonia, Serbia & Montenegro, Croatia, Montenegro, Serbia, Slovenia, Austria, Czech Republic, Bulgaria, Bosnia and Herzegovina, Switzerland

Iberski polotok. Andorra, Portugal, Spain

Vzhod. Belarus, Georgia, Russia, Ukraine, Israel, Cyprus, Greece, Moldova, Romania, Armenia, Monaco, Poland, Azerbaijan, Hungary

Zahod. Belgium, San Marino, Netherlands, Germany, France

Sever. Denmark, Norway, Sweden, Finland, Iceland, Estonia, Latvia, Slovakia, Lithuania

Otoške. Ireland, United Kingdom, Malta

Turčija. Turkey

Tabela 1: Skupine in njihove preferirane in nepreferirane države.

ime skupine	preferirane države	nepreferirane države
Balkan	Montenegro, Serbia, Serbia & Montenegro	Estonia, Ireland, Spain, United Kingdom
Iberski polotok	Spain, Monaco, San Marino	Croatia, Sweden, Slovenia, Turkey
Vzhod	Azerbaijan, Czech Republic, Georgia	France, Sweden, Switzerland, Norway
Zahod	France, Slovakia, Czech Republic, Germany	Croatia, Malta, Estonia, Ireland, Sweden
Sever	Slovakia, Azerbaijan, Iceland	Turkey, Spain, Romania, France, Croatia
Otoške	Azerbaijan, Georgia, San Marino, Slovakia	Croatia, France, Spain, Turkey
Turčija	Turkey, Serbia & Montenegro, Georgia	Estonia, Ireland, Malta, Spain

Preferirane in nepreferirane države, ki jih lahko vidimo v tabeli, 1 sem določil z metodo iskanja najkrajših oziroma najdaljših povprečnih razdalj med skupino in vsemi državami. Tako sem v tabelo vpisal države, ki so najbolj izstopale.

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

```

      ---- Albania
      ----|
      ----|      ---- Macedonia
      ----|      ----|
      ----|      ----|      Serbia & Montenegro
      ----|
      ----|      ---- Croatia
      ----|      ----|
      ----|      ----|      Montenegro
      ----|
      ----|      ---- Serbia
      ----|      ----|
      ----|      ----|      Slovenia
      ----|
      ----|      ---- Austria
      ----|      ----|
      ----|      ----|      Czech Republic
      ----|      ----|
      ----|      ----|      Bulgaria
      ----|
      ----|      ----|      Bosnia and Herzegovina
      ----|
      ----|      ----|      Switzerland
      ----|
      ----|      ----|      Andorra
      ----|      ----|
      ----|      ----|      Portugal
      ----|
      ----|      ----|      Spain
      ----|
      ----|      ----|      Armenia
      ----|      ----|
      ----|      ----|      Monaco
      ----|      ----|
      ----|      ----|      Poland
      ----|
      ----|      ----|      Azerbaijan
      ----|      ----|
      ----|      ----|      Hungary
      ----|
      ----|      ----|      Belarus
      ----|      ----|
      ----|      ----|      Georgia
      ----|      ----|
      ----|      ----|      Russia
      ----|
      ----|      ----|      Ukraine
      ----|
      ----|      ----|      Israel
      ----|
      ----|      ----|      Cyprus
      ----|      ----|
      ----|      ----|      Greece
      ----|
      ----|      ----|      Moldova
      ----|      ----|
      ----|      ----|      Romania
      ----|
      ----|      ----|      Denmark
      ----|      ----|
      ----|      ----|      Norway
      ----|
      ----|      ----|      Sweden
      ----|
      ----|      ----|      Finland
      ----|      ----|
      ----|      ----|      Iceland
      ----|
      ----|      ----|      Estonia
      ----|
      ----|      ----|      Latvia
      ----|      ----|
      ----|      ----|      Slovakia
      ----|
      ----|      ----|      Lithuania
      ----|
      ----|      ----|      Ireland
      ----|      ----|
      ----|      ----|      United Kingdom
      ----|
      ----|      ----|      Malta
      ----|
      ----|      ----|      Belgium
      ----|      ----|
      ----|      ----|      San Marino
      ----|
      ----|      ----|      Netherlands
      ----|
      ----|      ----|      Germany
      ----|
      ----|      ----|      France
      ----|
      ----|      ----|      Turkey

```

Slika 1: Slika prikazuje tekstovni izris dendrograma