

# Glasovanje za Pesem Evrovizije

Karmen Gostiša (63130057)

15. oktober 2017

## 1 Podatki

Analizirala sem podatke glasovanj za tekmovanje Pesem Evrovizije, podane v dveh CSV datotekah. Vrstice v vsaki datoteki predstavljajo posamezno državo, podatke o izvajalcu in pesmi ter ocene, ki jih je prejela od drugih držav. Postopek generiranja profila glasovanja za posamezno državo lahko opišem z naslednjimi koraki:

**Korak 1** V orodju Excel sem iz obeh datotek odstranila stolpce (attribute), ki jih za algoritem hierarhičnega razvrščanja ne potrebujem. Pustila sem le leto, državo tekmovalko in ocene, ki jih je prejela od vseh držav.

**Korak 2** Iz imen držav sem pobrisala odvečne presledke s pomočjo orodja Notepad++ z ukazom nadomesti vse (angl. *replace all*).

**Korak 3** V programski kodi sem podatke iz obeh datotek naložila v podatkovni strukturi `DataFrame` knjižnice `pandas`.

**Korak 4** Ker vsaka datoteka vsebuje rezultate posamezne države iz več let, sem posamezne vrstice združila po imenih države in za vrednost v stolpcu izračunala povprečje točk: `data.groupby("Country").mean()`.

**Korak 5** Matriki sem transponirala in združila v eno, tako da sem eno matriko priključila drugi z metodo `pandas.concat`. Profil (vektor) glasovanja za posamezno državo je torej sestavljen iz dveh delov - glasovanja te države v polfinalnih in finalnih delih.

## 2 Računanje razdalj

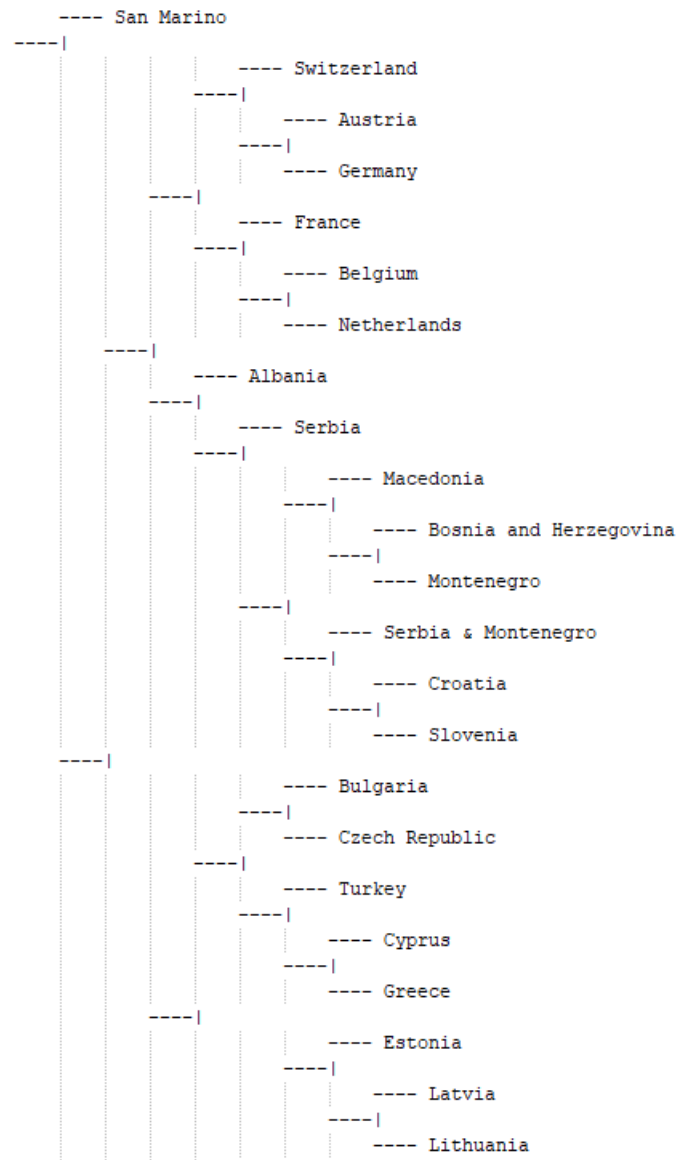
Med profiloma držav  $a$  in  $b$  sem izračunala evklidsko razdaljo (1) in jo normalizirala s številom uporabljenih parov  $p$ . Neznanih ali neštevilskih vrednosti (`NaN`) torej nisem upoštevala in jih izpustila iz izračuna razdalje.

$$d_{evklidska}(a, b) = \sqrt{\frac{\sum_{i=1}^m (a_i - b_i)^2}{p}} \quad (1)$$

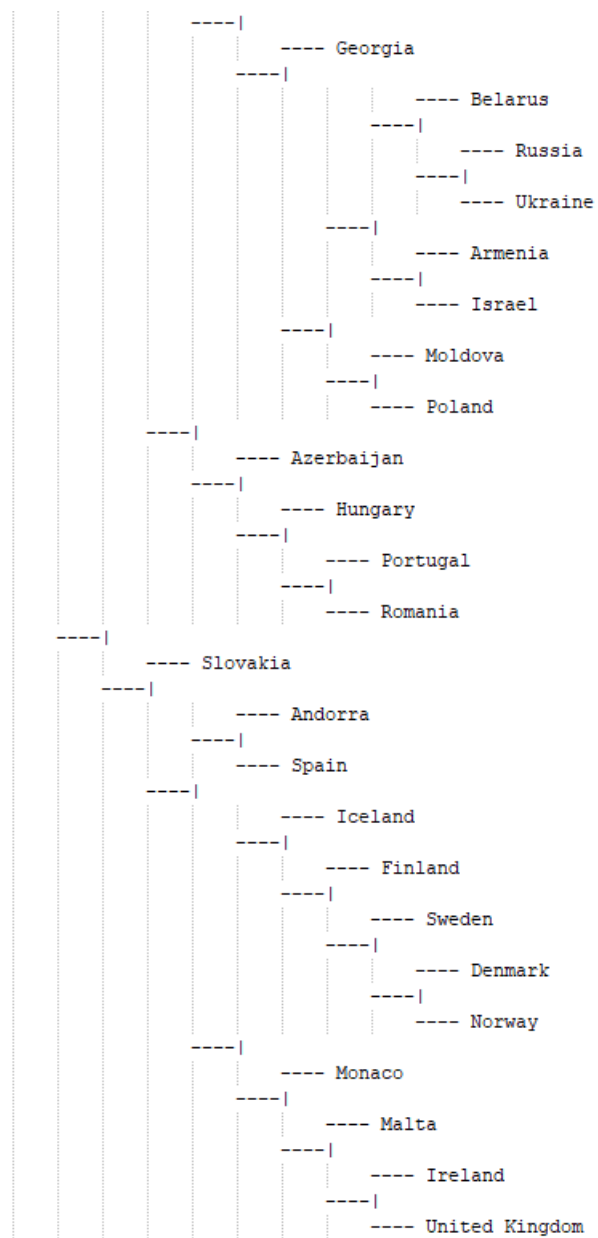
Razdaljo med skupinama držav sem izračunala tako, da sem za vse možne pare držav iz skupin izračunala evklidsko razdaljo (1) ter rezultat povprečila s številom parov.

### 3 Dendrogram

Dobljeni dendrogram prikazujeta sliki 1 in 2.



Slika 1: Zgornji del dendrograma.



Slika 2: Spodnji del dendrograma.

## 4 Skupine in njihove preferenčne izbire

Skupine držav	Preferirane države	Nepreferirane države
Švedska, Danska, Norveška, Finska, Islandija	Srbija in Črna gora, Bosna in Hercegovina, Švedska, Islandija	Monako, Češka, Črna gora, Avstrija
Rusija, Belorusija, Ukrajina, Armenija, Izrael	Rusija, Belorusija, Gruzija, Grčija	Monako, Češka, Avstrija, San Marino
Estonija, Latvija, Litva	Rusija, Gruzija, Azerbajdžan	Makedonija, Črna gora, Turčija, Slovaška, Avstrija
Srbija, Makedonija, Hrvaška, Slovenija, Bosna in Hercegovina, Črna gora	Albanija, Rusija, Slovenija, Grčija	Belgija, Poljska, Nizozemska, Ciper, Andora
Švica, Avstrija, Nemčija	Srbija, Srbija in Črna gora, Bosna in Hercegovina, Turčija	Andora, Belorusija, Belgija, Češka
Irska, Malta, Združeno kraljestvo	Grčija, Švedska, Romunija, Srbija in Črna gora	Avstrija, Hrvaška, Monako, San Marino
Turčija, Ciper, Grčija	Romunija, Srbija in Črna gora, Bolgarija, Armenija	Slovaška, Belgija, Češka, Avstrija

Tabela 1: Skupine držav in njihove preferenčne oz. nepreferenčne izbire.

Skupine držav sem določila s pomočjo dendrograma, tako da sem združila države zadnjih nekaj nivojev (v večini primerov 3 ali 4).

Za odkrivanje preferiranih in nepreferiranih držav posamezne skupine sem napisala funkcijo `find_pref_not_pref(group, countries, data)`, ki za dano skupino izračuna povprečje ocen za vsako državo v vektorju glasovanja. Seznam sem nato uredila naraščajoče po ocenah in v tabelo 1 zapisala prve oz. zadnje štiri države iz seznama.

## 5 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelala sama.