

**Uvod.** Pri prvi domači nalogi smo analizirali rezultate glasovanja Za Pesem Evrovizije in poskusili države hierarhično razvrstiti v smiselne skupine. Pri tem smo uporabili programski jezik Python.

**Podatki.** Algoritem smo testirali na podatkih o glasovanju v finalnem delu Evrovizije od leta 1975 do leta 2019 (podatki so dostopni na URL <https://data.world/datagraver/eurovision-song-contest-scores-1975-2019/>). Zapisani so v formatu .csv, njihova struktura pa je vidna v tabeli 1. Profile smo izluščili iz podatkov za leta od 2000 do 2010, saj le ta predstavljajo dovolj veliko množico, da opazimo preference v glasovanju, hkrati pa ni prevelika, da bi se značilnosti skupin skozi leta povprečile in izgubile. Profil vsake države sestoji iz točk, ki jih ja ta država v določenem letu podarila drugi državi. Upoštevali smo le točke pridobljene z žirijo, ker se televoting pojavi šele po letu 2016. Primer profila je viden v tabeli 2.

Tabela 1: Struktura podatkov, ki smo jih analizirali. Zbirka podatko vsebuje 33376 takšnih vnosov.

leto	tip glasovanja (J = žirija, T = televoting)	od države	za državo	točke
1975	J	Belgium	Finland	0

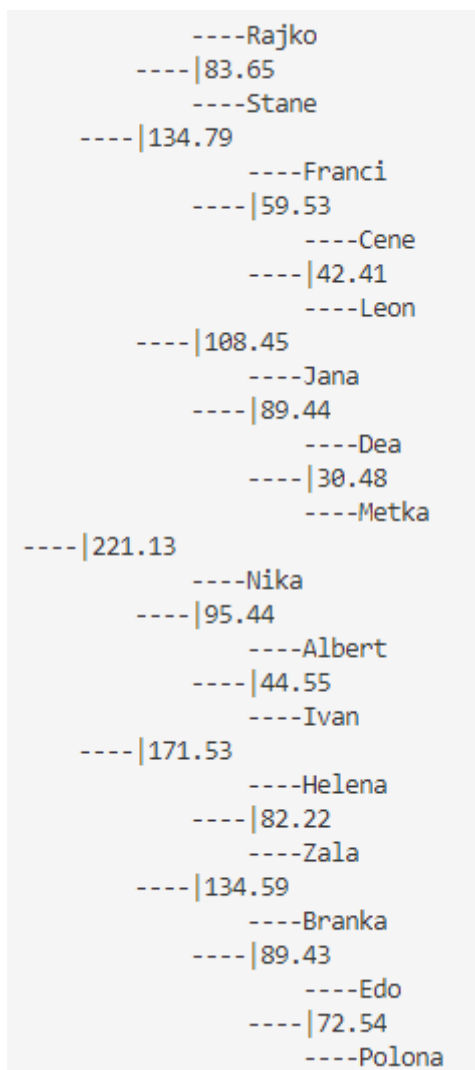
Tabela 2: Del profila za državo Avstrija. Profil ima toliko dimenzij kot je nastopov v izbranem časovnem obdobju (za obdobje 2000 - 2019 ima 268 dimenzij).

država	2000Belgium	2000Croatia	2000Cyprus	2000Denmark	...
Avstrija	0	6	0	10	...

**Računanje razdalj.** Razdalje med posameznimi profili smo računali z evklidsko razdaljo, za računanje razdalj med posameznimi skupinami pa se je najbolje obnesla maksimalna razdalja (*angl. Complete linkage*). Potrebno je omeniti, da smo evklidsko razdaljo utežili s številom let, ko sta nastopali obe državi. Ker vse države ne nastopijo vsako leto, nam ta pristop prepreči situacije, ko dve oddaljeni državi ocenimo kot bližnji, ker sta nastopali manjkrat kot druge. V kodi je utežitev zakomentirana, da lahko algoritem opravi avtomatske teste, ki ne obdelujejo Evrovizije. Neznane vrednosti profila (pojavi se v letih, ko država ni nastopala, v profilu pa ima atribut za države, ki pa so nastopale v teh letih) označimo z *NaN*, ter jih nato pri računanju razdalj profilov in skupin ignoriramo.

**Dendrogram.** Rezultate hierarhičnega razvrščevanja smo predstavili z dendrogramom, s katerega lahko odčitamo kako so združene skupine in pri kateri razdalji. V okviru naloge smo

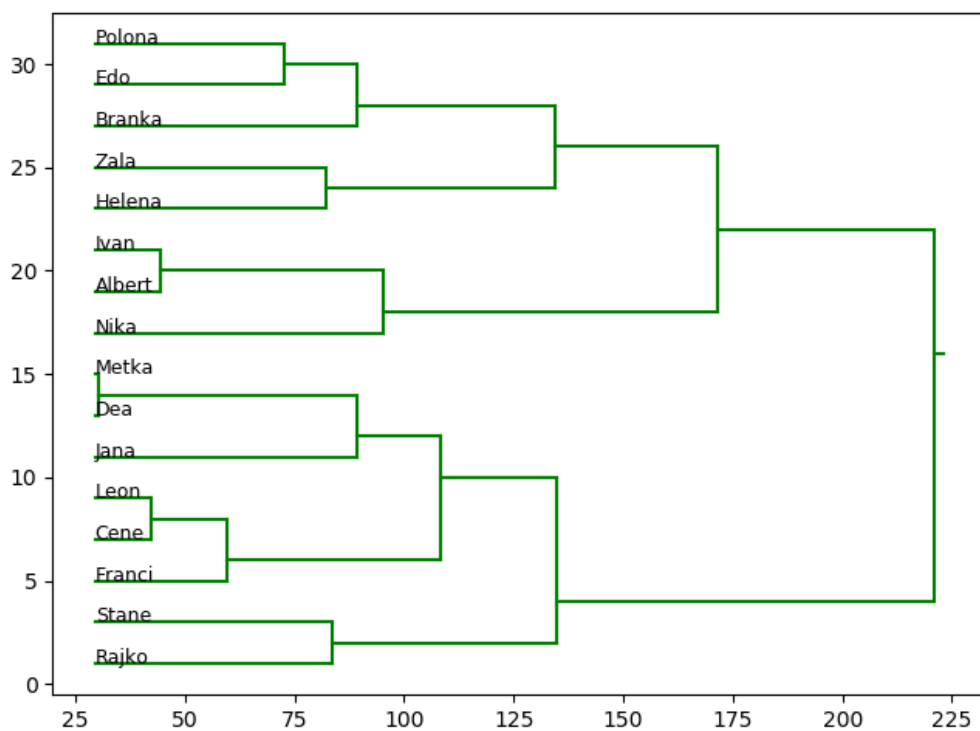
razvili izris preprostega tekstovnega dendrograma (viden na sliki 1) in izris grafičnega dendrograma (viden na sliki 2) s knjižnjico matplotlib.



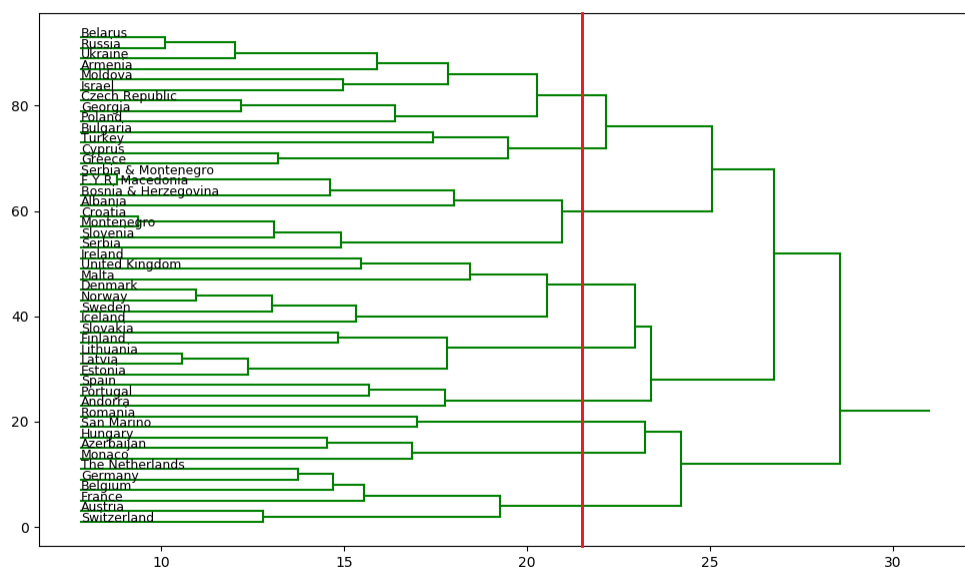
Slika 1: Izris tekstovnega dendrograma na krajšem testnem primeru dijakov.

**Skupine in njihove preferenčne izbire.** Po hierarhičnem razvrščanju smo skupine določili tako, da smo dendrogram "prerezali" na razdalji 21,5, kar lahko vidimo na sliki 3. Takšna razvrstitev držav v skupine je smiselna saj dobimo 9 skupin, ki vsebujejo države, ki se nam v splošnem zdijo sorodne (na primer skupina nordijskih, balkanskih in vzhodno evropskih držav). Pri iskanju primerne razdalje smo upoštevali, da je v njeni bližini čim manj združitvev oz. da so združitve čimbolj oddaljene zaradi morebitnih šumov.

Tabela 3 prikazuje skupine in države za katere posamezne skupine največkrat in najmanjkrat glasujejo. Te preferenčne države smo določili, tako da smo za vsako državo seštelili vse točke, ki jih je dobila od članic skupine in delili s seštevkom nastopov vseh članic. Tako smo dobili uteženo povprečje točk, ki jih dobi posamezna država od izbrane skupine.



Slika 2: Izris grafičnega dendrograma na krajšem testnem primeru dijakov.



Slika 3: Dendrogram hierarhičnega razvrščanja držav Evrovizije v skupine. Rdeča črta predstavlja razdaljo 21,5 pri kateri smo "razrezali" dendrogram in določili skupine.

skupina	članice	5 najbolj preferiranih držav	5 najmanj preferiranih držav
1	Switzerland	Turkey	Belarus
	Austria	Greece	Moldova
	France	Armenia	Georgia
	Belgium	Germany	F.Y.R. Macedonia
	Germany	Bosnia and Herzegovina	Lithuania
	The Netherlands		
2	Monaco	Ukraine	Belgium
	Azerbaijan	Turkey	Finland
	Hungary	Greece	The Netherlands
		Israel	Austria
		Russia	Poland
3	Romania	Greece	The Netherlands
	San Marino	Turkey	Austria
		Moldova	Lithuania
		Russia	Slovenia
		Israel	Georgia

4	Andorra	Romania	F.Y.R. Macedonia
	Portugal	Spain	Albania
	Spain	Ukraine	Belarus
		Germany	Georgia
		Greece	Croatia
5	Estonia	Russia	F.Y.R. Macedonia
	Latvia	Sweden	The Netherlands
	Lithuania	Norway	Albania
	Finland	Ukraine	Bulgaria
	Slovakia	Latvia	Portugal
6	Iceland	Sweden	F.Y.R. Macedonia
	Sweden	Denmark	Armenia
	Norway	Norway	Bulgaria
	Denmark	Greece	Belarus
	Malta	Turkey	Croatia
	United Kingdom		
	Ireland		
7	Serbia	Bosnia and Hercegovina	Poland
	Slovenia	Greece	Portugal
	Montenegro	Turkey	The Netherlands
	Croatia	Croatia	Switzerland
	Albania	F.Y.R. Macedonia	Austria
	Bosnia and Hercegovina		
	Serbia and Montenegro		
	F.Y.R. Macedonia		
8	Greece	Greece	Portugal
	Cyprus	Russia	Slovenia
	Bulgaria	Armenia	The Netherlands
	Turkey	Romania	Switzerland
		Ukraine	Lithuania

9	Poland	Russia	Austria
	Georgia	Ukraine	Albania
	Czech Republic	Armenia	Poland
	Israel	Romania	Ireland
	Moldova	Norway	Portugal
	Armenia		
	Ukraine		
	Russia		
	Belarus		

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