



Mašinsko učenje 2018

Zadatak 5



Sadržaj

- Zadatak 4 - Rekapitulacija
- Zadatak 5



Zadatak 4 - Rekapitulacija



Zadatak 4 - Rekapitulacija

- Procenat uspešnosti : **79%** (23/29)
- Najbolji rezultat po terminima:
 - Ponedeljak:
 - **code10** (0.566)
 - Utorak 1:
 - **tim2** (0.571)
 - Utorak 2:
 - **tartufi** (0.548)
 - Sreda:
 - **fleka_od_dzeka** (0.566)
- Maksimalno preklapanje po MOSS-u: **11%**



Zadatak 5



Zadatak 5

- Klasterovanje:
 - Klasterovati države na osnovu njihovih osobina u klastere koji predstavljaju kontinente (**region**):
 - Africa
 - Americas
 - Asia
 - Europe
 - Zadatak je uspešno urađen ukoliko se na kompletnom testnom skupu podataka dobije v mera (eng. v *measure score*) > 0.41 .
 - Zadatak se rešava upotrebom EM algoritma.
 - Rok **13.05.2018. u 19:59h.**
 - Dostupne biblioteke: **NumPy, SciPy, Pandas** i **scikit-learn**.
 - Trening skup podataka sadrži nedostajuće vrednosti (prazne ćelije).



Zadatak 5

- Atributi:
 - **income** - prihod po glavi stanovnika u \$
 - **infant** - smrtnost odojčadi na 1000 živorođenih
 - **oil** - da li je država izvoznik nafte:
 - **yes** - da
 - **no** - ne



Zadatak 5

- Algoritam
 - [EM](#)



Zadatak 5

- Evaluacija klasterovanja zasnovana na poznavanju *ground truth* labela klastera.
- Intuitivne metrike zasnovane na uslovnoj analizi entropije:
 - **homogenost** (*homogeneity*) - svaki klaster sadrži članove samo jedne grupe/klase
 - **potpunost** (*completeness*) - svi članovi iste grupe/klase su dodeljeni istom klasteru
- **v mera** predstavlja harmonijsku sredinu homogenosti i potpunosti:
 - `sklearn.metrics.v_measure_score(labels_true, labels_pred)`