



Mašinsko učenje 2018

Zadatak 3



Sadržaj

- Zadatak 2 - Rekapitulacija
- Zadatak 3



Zadatak 2 - Rekapitulacija



Zadatak 2 - Rekapitulacija

- Procenat uspešnosti: **83%** (24/29)
- Najbolji rezultati po terminima:
 - Ponedeljak:
 - **code10** (RMSE: 23527)
 - Utorak 1:
 - **blueseven** (RMSE: 23796.1)
 - Utorak 2:
 - **tartufi** (RMSE: 23714.76)
 - Sreda:
 - **tim9** (RMSE: 23274.49)
- Maksimalno preklapanje po MOSS-u: **~30%**



Zadatak 3



Zadatak 3

- Klasifikacija:
 - Na osnovu više atributa, klasifikovati geografske regione u dve klase (**NoYes**):
 - 0 - nije bilo zabeleženih slučajeva Denga groznice
 - 1 - bilo je zabeleženih slučajeva Denga groznice.
 - Zadatak je uspešno urađen ukoliko se na kompletnom testnom skupu podataka dobije tačnost (accuracy) veća od 0.73 (73%).
 - Zadatak se rešava upotrebom SVM klasifikatora.
 - Rok: **22.04.2018. u 19:59h.**
 - Dostupne biblioteke: **NumPy, SciPy, Pandas i Scikit-learn.**



Zadatak 3

- Atributi:
 - **humid** - prosečna gustina vodene pare u periodu od 30 godina
 - **temp** - prosečna temperatura u periodu od 30 godina
 - **trees** - procenat regiona pokriven drvećem (šumom)
 - **Xmin** - najmanja geografska dužina regiona
 - **Xmax** - najveća geografska dužina regiona
 - **Ymin** - najmanja geografska širina regiona
 - **Ymax** - najveća geografska širina regiona.



Zadatak 3

- Trening skup podataka sadrži nedostajuće vrednosti (označene sa 'NA').
- Rad sa nedostajućim vrednostima:
 - Uklanjanje torki koje sadrže nedostajuće vrednosti
 - Zamena nedostajućih vrednosti sa srednjom vrednošću tog atributa.



Zadatak 3

- scikit-learn biblioteka:
 - [Instalacija](#)
 - [Docs.](#)
- Izdvojeno:
 - [Selekcija modela](#)
 - [Selekcija obeležja](#)
 - [Metrike](#)
 - [Linearni modeli](#)
 - [Support Vector Machines.](#)