



Mašinsko učenje 2018

Zadatak 4



Sadržaj

- Zadatak 3 - Rekapitulacija
- Zadatak 4



Zadatak 3 - Rekapitulacija



Zadatak 3 - Rekapitulacija

- Procenat uspešnosti: **86%** (25/29)
- Najbolji rezultati po terminima:
 - Ponedeljak:
 - **code10** (Acc: 94.05%)
 - Utorak 1:
 - **tim1** (Acc: 95.7%)
 - Utorak 2:
 - **smd** (Acc: 95.17%)
 - Sreda:
 - **tim9** (Acc: 96.285%)
- Maksimalno preklapanje po MOSS-u: **~25%**



Zadatak 4



Zadatak 4

- Klasifikacija - Model ansambla:
 - Dostupan je deo policijskih izveštaja o saobraćajnim nesrećama u SAD u periodu od 1997-2002. Na osnovu podataka vezanih za vozače i suvozače vozila koja su učestvovala u nesrećama izvršiti procenu brzine vozila u trenutku sudara (**dvc**at):
 - 1-9km/h
 - 10-24
 - 25-39
 - 40-54
 - 55+



Zadatak 4

- Klasifikacija - Model ansambla:
 - Zadatak je uspešno urađen ukoliko se na kompletnom testnom skupu podataka dobije mikro f1 mera (eng. *micro f1 score*) > 0.52 .
 - Zadatak se rešava upotrebom ansambla klasifikatora.
 - Rok: **04.05.2018. u 23:59h.** (nema vežbi sledeće nedelje)
 - Dostupne biblioteke: **NumPy, SciPy, Pandas i Scikit-learn.**
 - Trening skup podataka sadrži nedostajuće vrednosti (prazne ćelije).



Zadatak 4

- Atributi:
 - **weight** - procenjena masa učesnika udesa (sadrži neprecizne procene)
 - **dead** - da li je učesnik preživeo udes ili ne:
 - **alive** - preživeo
 - **dead** - nastradao
 - **airbag** - da li je učesnik imao airbag:
 - **none** - ne
 - **airbag** - da
 - **seatbelt** - da li je učesnik bio vezan:
 - **none** - ne
 - **belted** - da



Zadatak 4

- Atributi:
 - **frontal** - frontalni udar ili ne:
 - 0 = ne
 - 1 = frontalni udar
 - **sex** - pol:
 - f - ženski
 - m - muški
 - **ageOFocc** - starost učesnika
 - **yearacc** - godina kada se dogodila saobraćajna nesreća
 - **yearVeh** - godina proizvodnje vozila



Zadatak 4

- Atributi:
 - **abcat** - da li se aktivirao airbag:
 - **unavail** - vozilo nije imalo airbag za tog učesnika
 - **nodeploy** - airbag se nije aktivirao
 - **deploy** - airbag se aktivirao
 - **occRole** - tip učesnika:
 - **driver** - vozač
 - **pass** - suvozač
 - **deploy** - označava da li se airbag aktivirao:
 - **0** = airbag nije dostupan za tog učesnika ili se nije aktivirao
 - **1** = airbag se aktivirao



Zadatak 4

- Atributi:
 - **injSeverity** - stepen povreda:
 - 0 = bez povreda
 - 1 = lakše telesne povrede
 - 2 = teže telesne povrede, bez invaliditeta
 - 3 = teže telesne povrede, sa invaliditetom
 - 4 = smrt
 - 5 = nepoznato
 - 6 = teške telesne povrede sa smrtnim ishodom (smrt nastupila kasnije).



Zadatok 4

- Ansambli klasifikatora:
 - Bagging
 - Boosting
 - Stacking
 - [scikit-learn](#)

Zadatak 4

- Matrica konfuzije (eng. *confusion matrix*):

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Type I error
(false positive)



Type II error
(false negative)

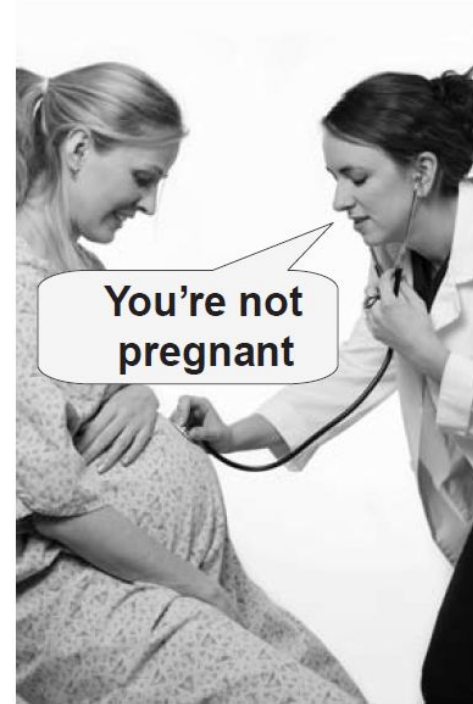


Figure 3.1 Type I and Type II errors



Zadatak 4

- **Precision** - procenat relevantnih (tačnih) među prediktovanim - $P = TP / (TP + FP)$
- **Recall** - procenat relevantnih (tačnih) koje su prediktovane - $R = TP / (TP + FN)$
- **F1 score (*F-measure*)** - harmonijska sredina **P** i **R** - $F1 = 2 * P * R / (P + R)$
- **Micro F1 score** - računa globalne TP, FN i FP:
 - `sklearn.metrics.f1_score(y_true, y_pred, average='micro')`
- Prilikom treninga, od pomoći može biti i [classification report](#).