



DETEKCIJA STADIJUMA CIROZE JETRE UZ POMOĆ MAŠINSKOG UČENJA

MILOŠ ČUTURIĆ – SV11/2020

LUKA ĐORĐEVIĆ – SV14/2020

MARKO JANOŠEVIĆ - SV46/2020

UNIVERZITET U NOVOM SADU, FAKULTET TEHNIČKIH NAUKA

PROBLEM I MOTIVACIJA

- Ozbiljna hronična bolest
- Pravovremena dijagnoza
- Brza i precizna predikcija uz pomoć mašinskog učenja

PROBLEM I MOTIVACIJA

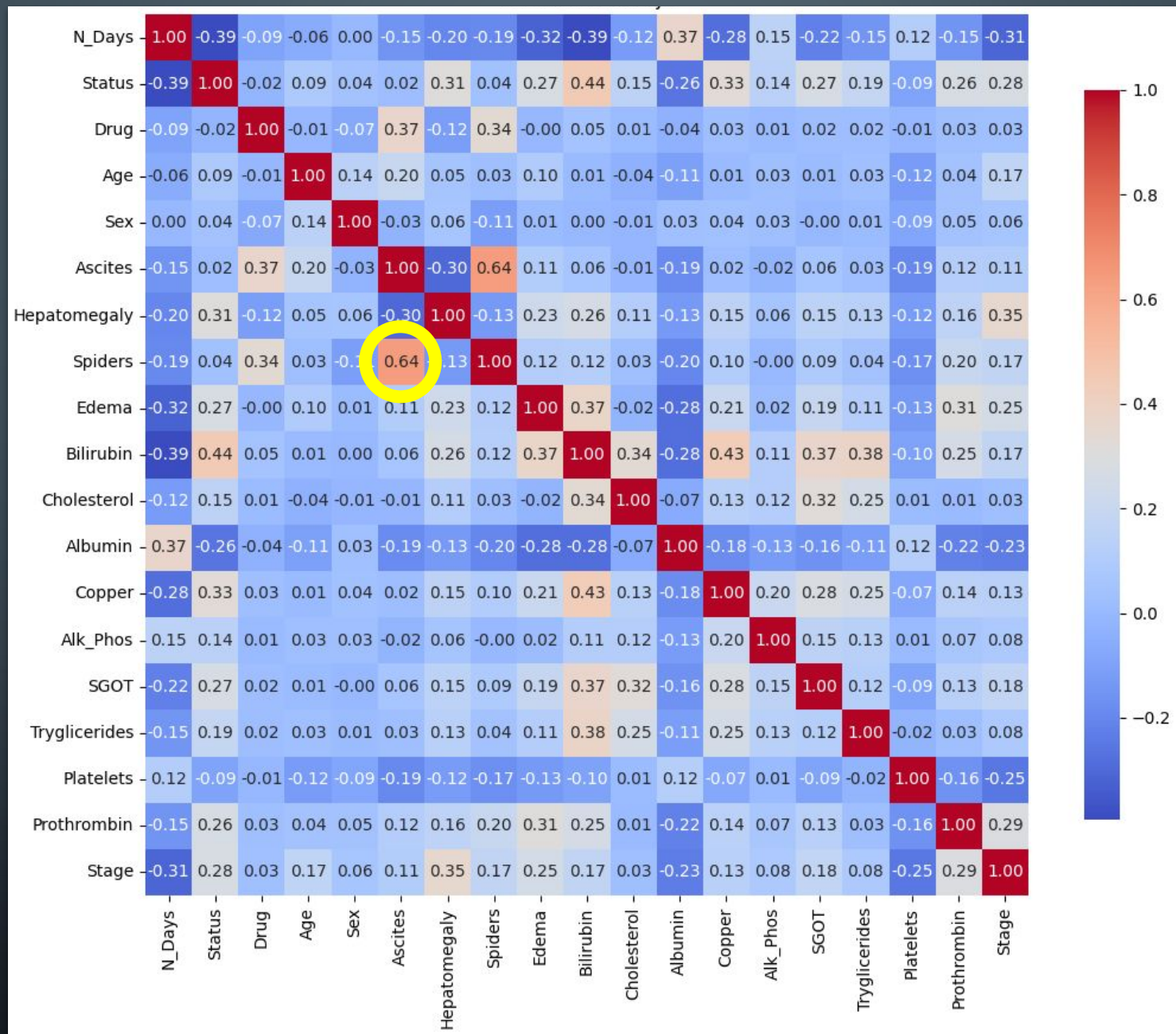
- Stadijumi ciroze jetre: 1, 2 i 3
- Podaci preuzeti sa Kaggle platforme
- Sadrži 19 obeležja i preko 10.000 redova podataka
- Informacije o demografskim podacima pacijenata, kliničkim karakteristikama i laboratorijskim rezultatima.

PRETPROCESIRANJE - ENKODIRANJE I NORMALIZACIJA

- Većina kategoričkih obeležja su binarna
- Dimenzionalnost dataset-a velika
- Label encoding
- Z - Score normalizacija

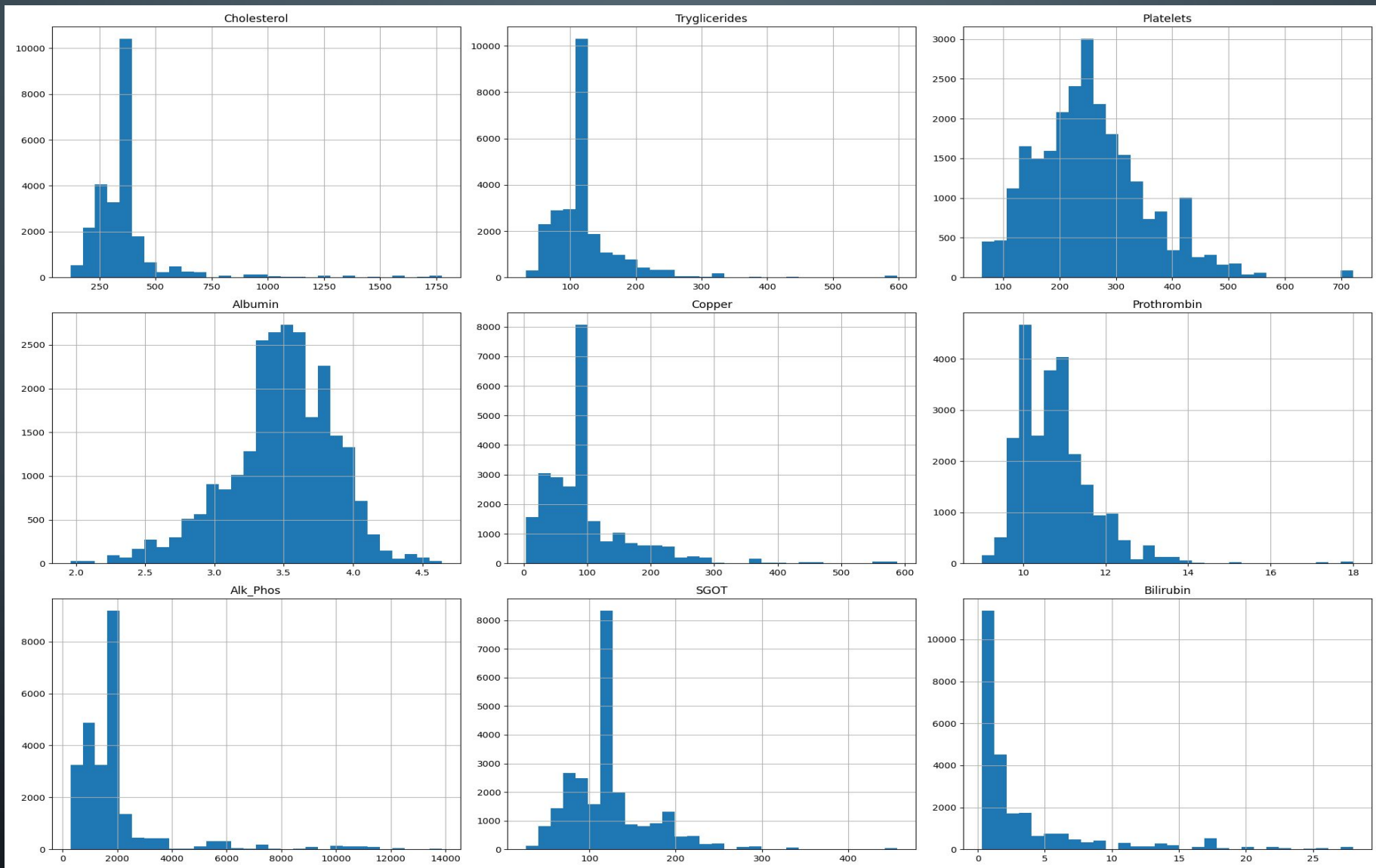
PRETPROCESIRANJE - REDUKCIJA DIMENZIONALNOSTI

- Pokušaj korišćenja PCA algoritma
- Bez učinka



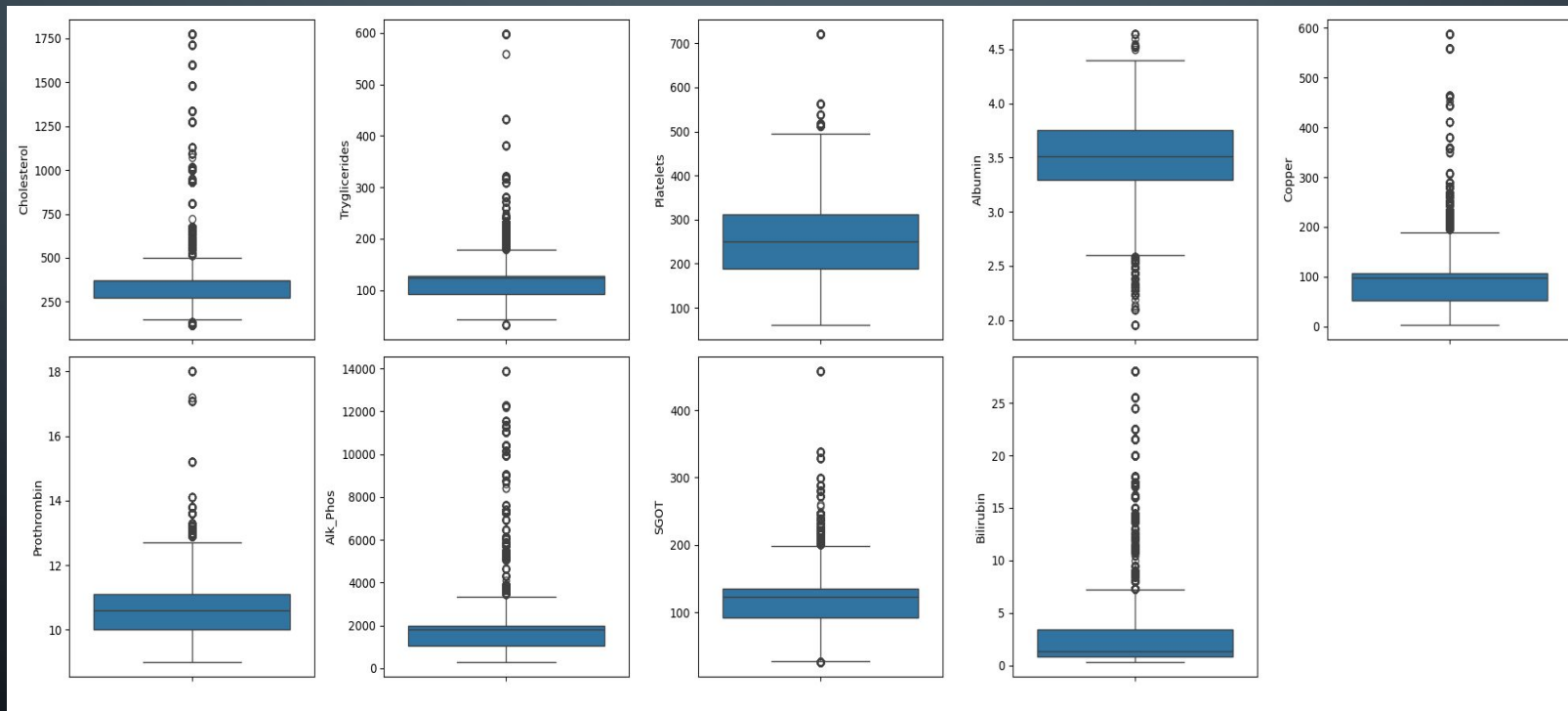
PRETPROCESIRANJE - OUTLIER-I

- Prilikom analize distribucija obeležja, primećeno je da u skupu podataka postoje potencijalni outlier-i



PRETPROCESIRANJE - OUTLIER-I

- Box-plot obeležja potvrđuje postojanje outlier-a:



TRENIRANJE MODELA

- Podjela podataka na trening i test skup u odnosu 80:20
- Testirana 4 modela:
 - a. Random Forest
 - b. Bagging Classifier
 - c. Logistička regresija
 - d. Gradient Boosting

TRENIRANJE MODELA

- 5-fold cross-validacija
- Optimizacija hiperparametara
- GridSearchCV

EVALUACIJA I REZULTATI

- Klase skupa podataka su izbalansirane

Stadijum	Broj uzoraka
1	8265
2	8441
3	8294

EVALUACIJA I REZULTATI

- Micro f1 mera, tačnost, preciznost, odziv

Model	Accuracy	Precision	Recall	Micro f1
Random forest	0.9552	0.9552	0.9552	0.9552
Bagging	0.9588	0.9588	0.9588	0.9588
Boosting	0.9632	0.9632	0.9632	0.9632
Logistic regression	0.6002	0.6002	0.6002	0.6002

The background is a dark blue gradient. In the corners, there are decorative white line art elements resembling circuit boards or neural networks, with lines and small circles connecting them.

HVALA NA PAŽNJI!