



# **Klasifikacija karcinoma dojke strojnim učenjem**

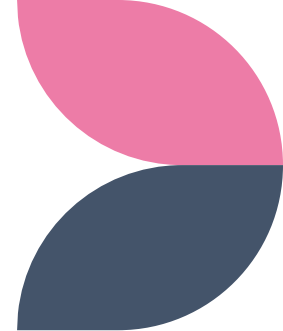
Fran Hrabar, Leon Novački, Matej Čubek, Fani Sentinella-Jerbić



# Karcinom dojke



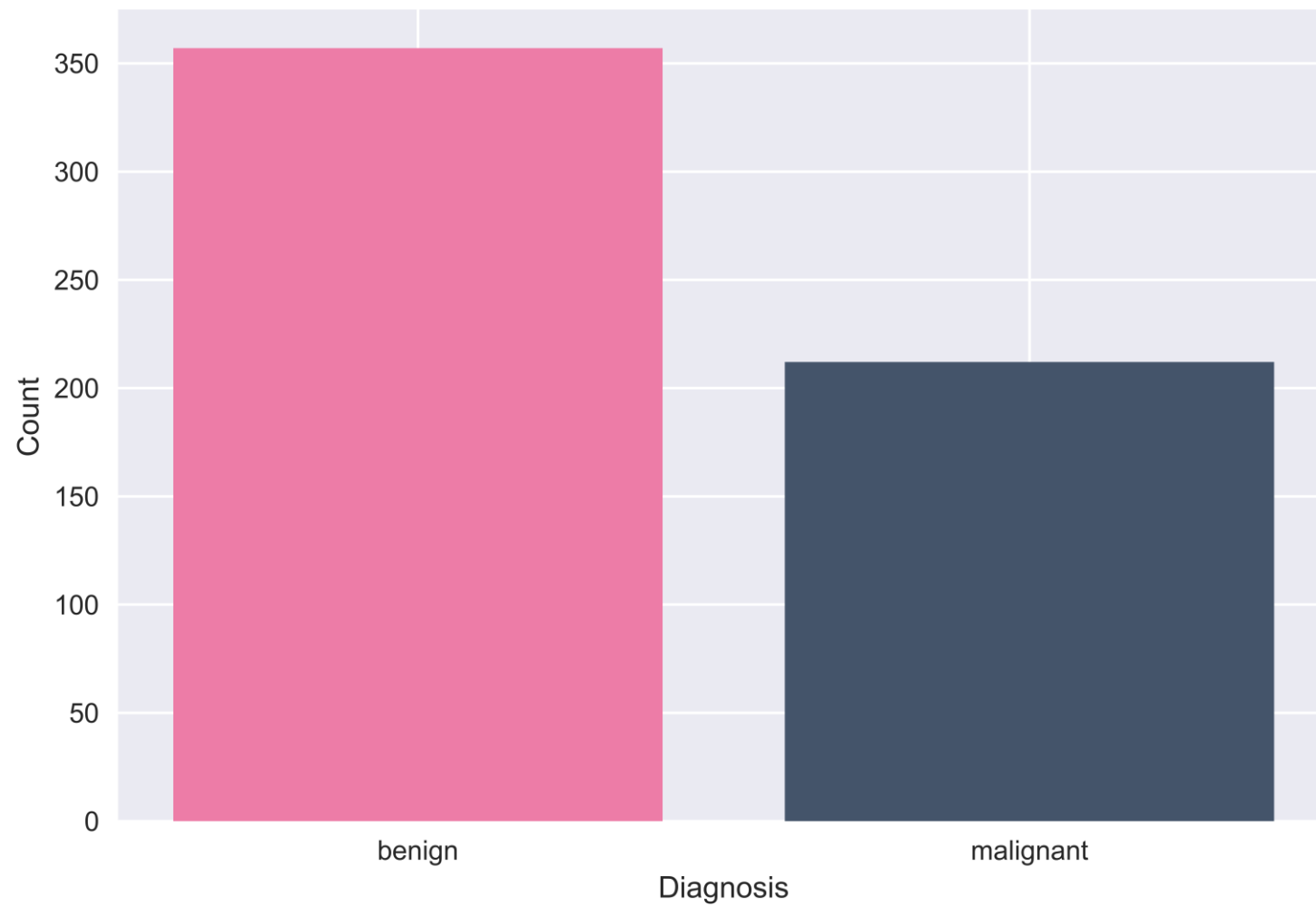
# Tijek rada



# Analiza skupa podataka

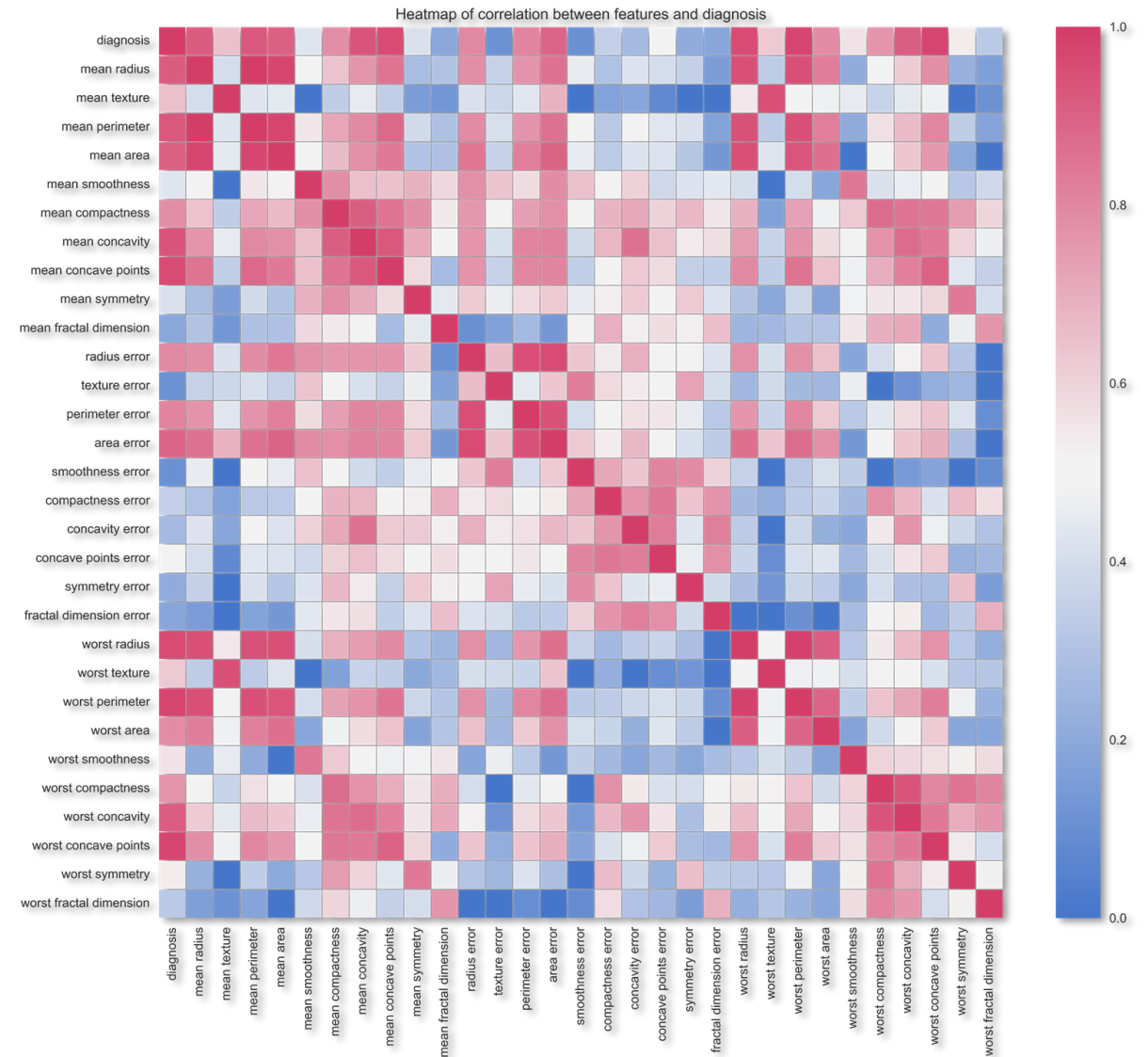
- Sveučilišne bolnice u Wisconsinu
- Stanice tkiva dojke prikupljene s pacijenata biopsijom
- Analiza mikroskopom
- Izmjerene statističke značajke jezgara stanica
  - $\bar{x}$ , SE, max

Broj pojedine vrste karcinoma u korištenom skupu podataka



# Odabir značajki

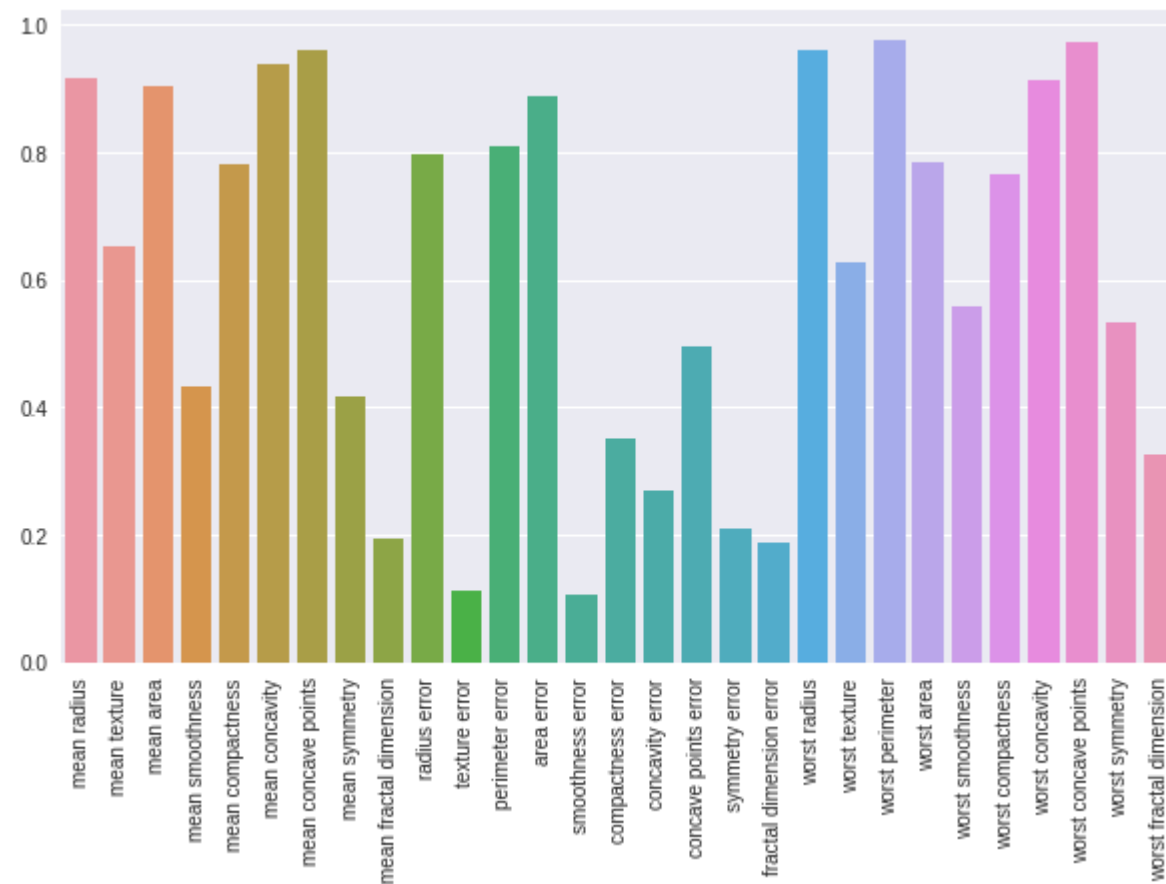
- Poželjne značajke koje najviše predviđaju dijagnozu
- $\phi$ k matrica korelacije
- Uklonjene visoko korelirane značajke



# Odabir značajki

- Podskup preostalih značajki
- Korelacija značajki prema dijagnozi
- *SequentialFeatureSelector* uz logističku regresiju kao procjenitelj

Korelacija pojedine značajke s dijagnozom



# Korišteni algoritmi strojnog učenja

## Iz izvornog članka

- Logistička regresija
- Stroj potpornih vektora
- Algoritam k-najbližih susjeda
- Naivni Bayes
- Stablo odluke
- Algoritam slučajnih šuma

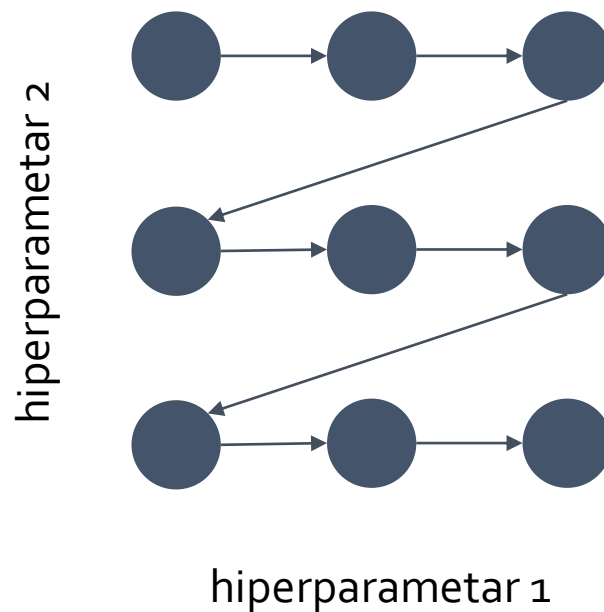
## Dodatni

- XGBoost
- Algoritam dodatnih stabala
- Višeslojni perceptron

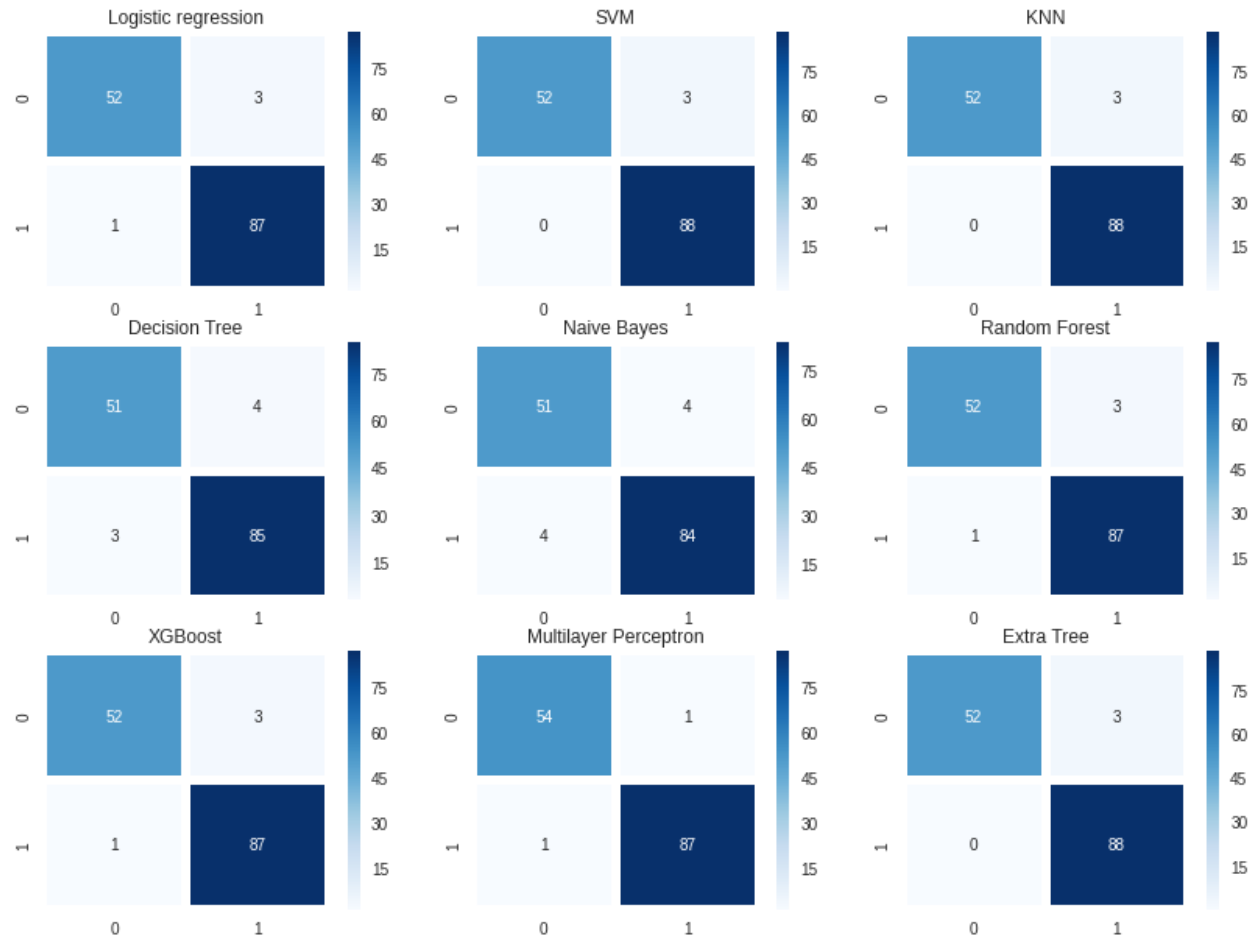


# Optimizacija hiperparametara

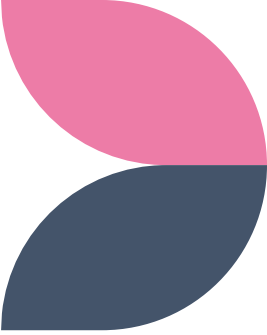
- Korišteno pretraživanje po rešetci s tipičnim vrijednostima



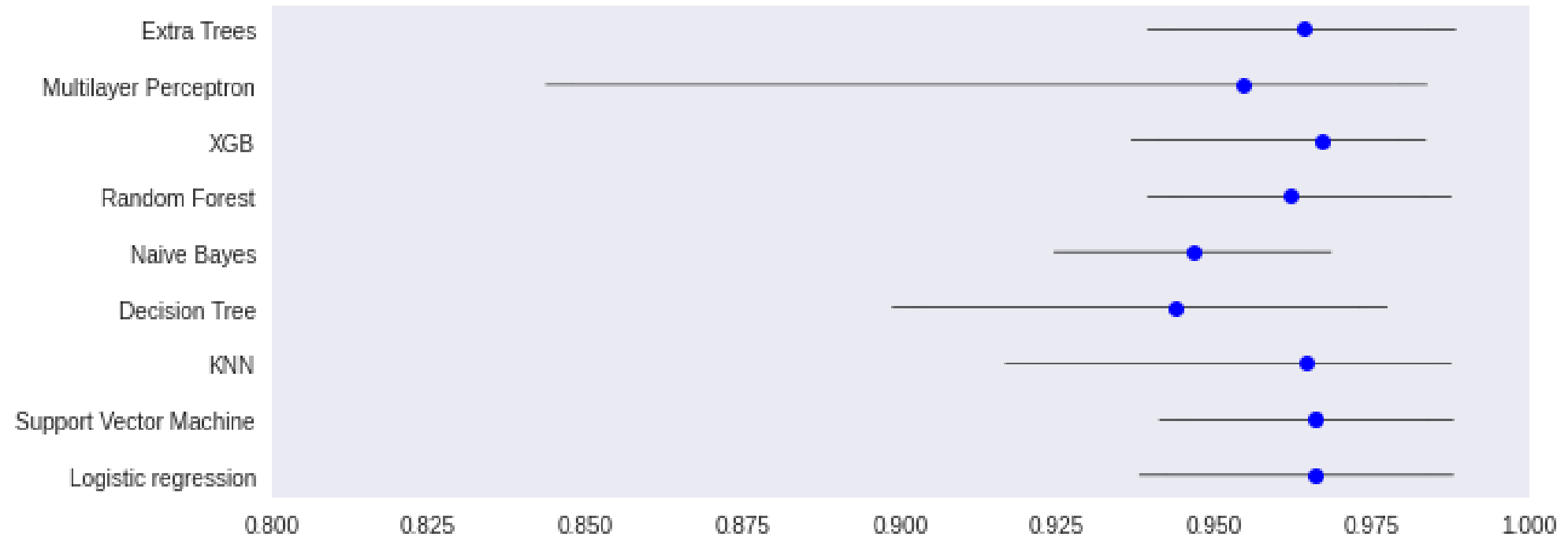
# Analiza rezultata



# Analiza rezultata



Usporedba dobivenih F1 vrijednosti za različite algoritme



# Sažetak napravljenog

- Poboljšan odabir značajki korištenjem *SequentialFeatureSelector*
- Dodane metrike uspješnosti algoritama F1, Recall, AUC
- Optimizirani hiperparametari pretraživanjem po rešetci
- Dodani novi algoritmi: Algoritam dodatnih stabla, XGBoost i Višeslojni perceptron

# Zaključak

Algoritmi strojnog učenja pokazuju **dobar potencijal za klasifikaciju raka dojke** i nadamo se da bi daljnja poboljšanja mogla povećati njihovu sposobnost točne klasifikacije.

# Hvala na pažnji.



**Matej Čubek**  
matej.cubek@fer.hr



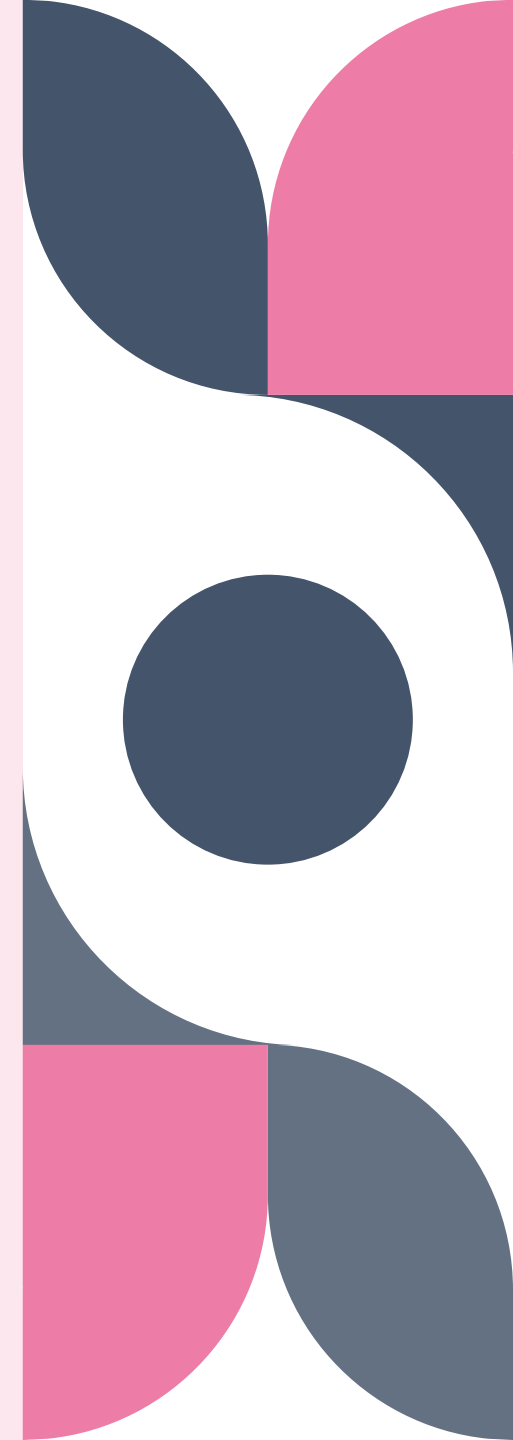
**Fran Hrabar**  
fran.hrabar@fer.hr



**Fani Sentinella-Jerbić**  
fani.sentinella-jerbic@fer.hr



**Leon Novački**  
leon.novacki@fer.hr



# Izvori

- S. Ara, A. Das and A. Dey: *Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms*, 2021 International Conference on Artificial Intelligence (ICAI), 2021, pp. 97-101, doi: 10.1109/ICAI52203.2021.9445249
- W.N. Street, W.H. Wolberg and O.L. Mangasarian. *Nuclear feature extraction for breast tumor diagnosis*. IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology, volume 1905, pages 861-870, San Jose, CA, 1993.
- Scikit-learn Documentation 1.0.2 <https://scikit-learn.org/>
- Seaborn Documentation 0.11.2 <https://seaborn.pydata.org/>