

# Предикција рака дојке

Даница Газдић SV12/2020  
Милош Обрадовић SV55/2020

## Циљ пројекта

Досадашња истраживања у медицини показују велику потребу за применом експертских система и метода у дијагностици болести. Циљ пројекта је дијагностиковати тумор на основу познатих обележја која представљају физичке и морфолошке карактеристике.

## Подаци

Коришћен је скуп података који је прикупљен у америчким здравственим установама у Висконсину. Садржи 11 колона:

- |                        |                          |
|------------------------|--------------------------|
| • Идентификација       | Величине имају           |
| • Дебљина груднице     | вредности у распону 1-10 |
| • Величина ћелија      | Последња колона има      |
| • Облик ћелија         | вредности 2(бенигни) и   |
| • Маргинална адхезија  | 4(малигни). Подела 70:30 |
| • Величина епителне ћ. |                          |

## Коришћене методологије

За израду пројекта коришћен је python 3.9. Коришћена је библиотека pandas за припрему података, а за моделе машинског учења користили смо библиотеку scikit-learn, из које смо користили и могућност оцене прецизности, и сопствену имплементацију неуронске мреже.

У препроцесирању смо дефинисали две технике рада са подацима који су имали непотпуне вредности: одбацити неисправне и вршити израчунавање непознате вредности на основу сличних случајева података.

Укупно је коришћено 5 алгоритама, и то:

1. Decision Tree
2. Random Forest
3. K Neighbors
4. SciKit NN
5. FTN NN

Прецизност смо рачунали као бр. тачних/бр. укупних.

## Евалуација решења



Као што можемо приметити, сопствена имплементација неуронске мреже се показала најпрецизније за постављене вредности броја епоха и величине беча. Такође се да приметити да је прецизност свих модела преко 90%, што је више него прихватљиво.

## Даљи рад

Сакупити више података и поново покренути тестове, истражити напредније техника руковања недостајућим подацима.