



UVOD

TEKMOVANJE RIS – 2021

ANDREJ STUDEN

1. VSEBINA TEKMOVANJA

Izziv:

Pandemija koronavirusne bolezni COVID-19 ni ogrozila le **svetovnega zdravja**, temveč je močno obremenila celotni zdravstveni sistem. Natančnejše napovedi obsežnosti COVID-19 okužbe v pljučih bi lahko pomagale pri **odločitvah o hospitalizaciji in zdravljenju**. Ali lahko sestaviš algoritem, ki bo iz CT slik prsnega koša čim natančneje **napovedal obsežnost COVID-19 okužbe v pljučih?**

2. POTEK TEKMOVANJA

Naloge:

- Objavljene bodo **5.4.** na spletni strani
- Rok za oddajo **19.4.**
- Način oddaje: preko povezave na spletni strani
- Lahko oddate več verzij, upoštevati bomo zadnjo.

Drugi krog:

- Tekmovanje za vikend, **8.5.**
- Obvestilo o udeležbi v drugem krogu: **22. 4.**
- Razglasitev končnih rezultatov:
- **14.5.2021, ob 15h**

Nagradni sklad **1500 Eur.**

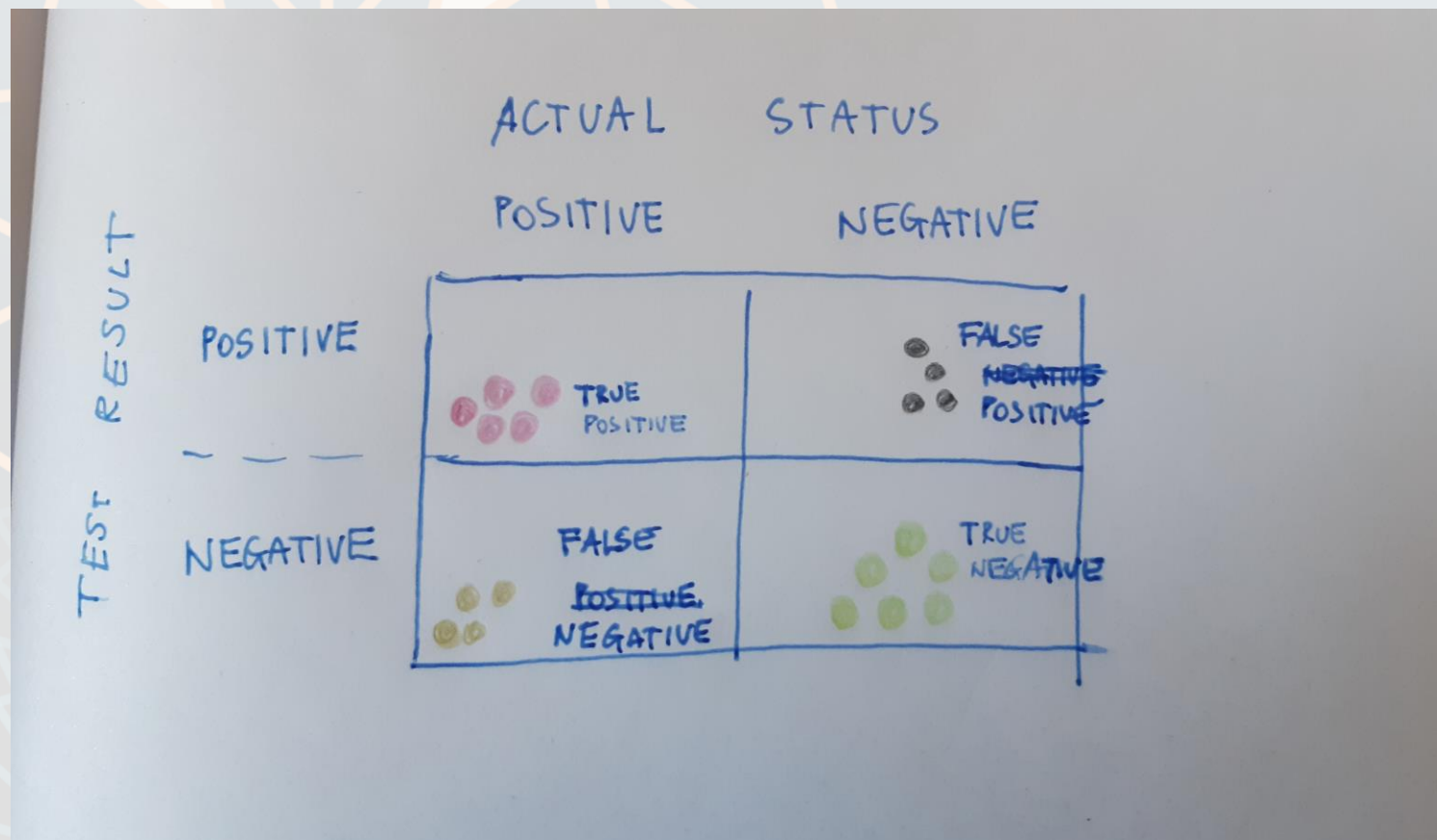
E-pošta za komentarje, težave: ris@dmfa.si

3. OSNOVE STATISTIKE – BINARNA KLASIFIKACIJA

Postopek ločevanja na dve skupini na podlagi (numeričnega) testa.

Zlati standard je presoja, ki jo štejejo za neoporečno, in ji pripišemo veljavnost dejanskega stanja.

Njen nasproti preverjamo lastno napoved.



3. OSNOVE STATISTIKE – FPR, TPR

Uspešno ločevanje daje samo prave rezultate; vsi pozitivni so prepoznani kot taki, negativni izločeni.

Definiramo:

True positive rate, TPR

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN})$$

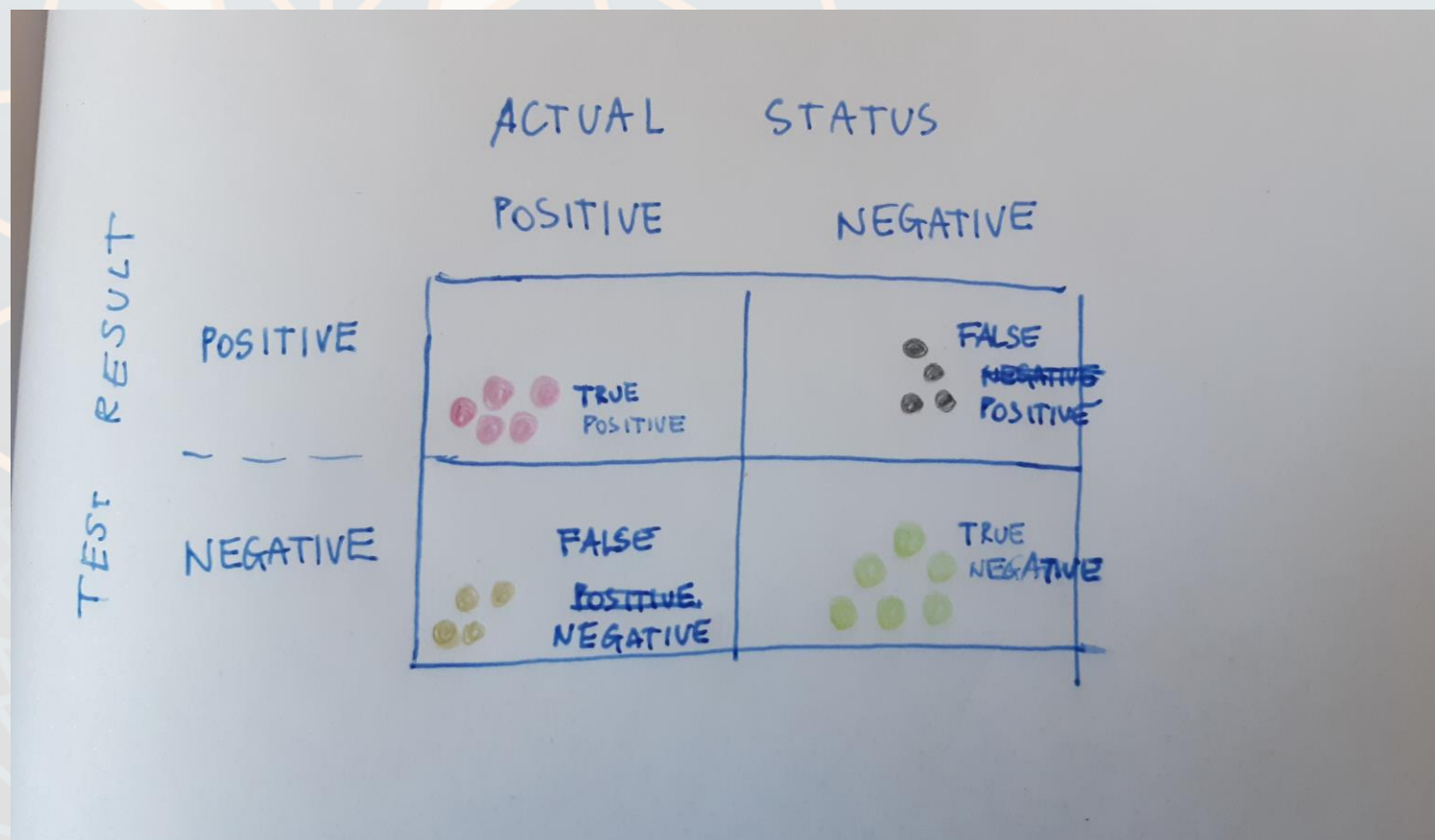
False positive rate, FPR

$$\text{FPR} = \text{FP} / (\text{TN} + \text{FP})$$

Idealno:

$$\text{TPR} = 1$$

$$\text{FPR} = 0$$



3. OSNOVE STATISTIKE – SENZITIVNOST, SPECIFIČNOST

V zdravstvu se uporabljata izraza senzitivnost in specifičnost:

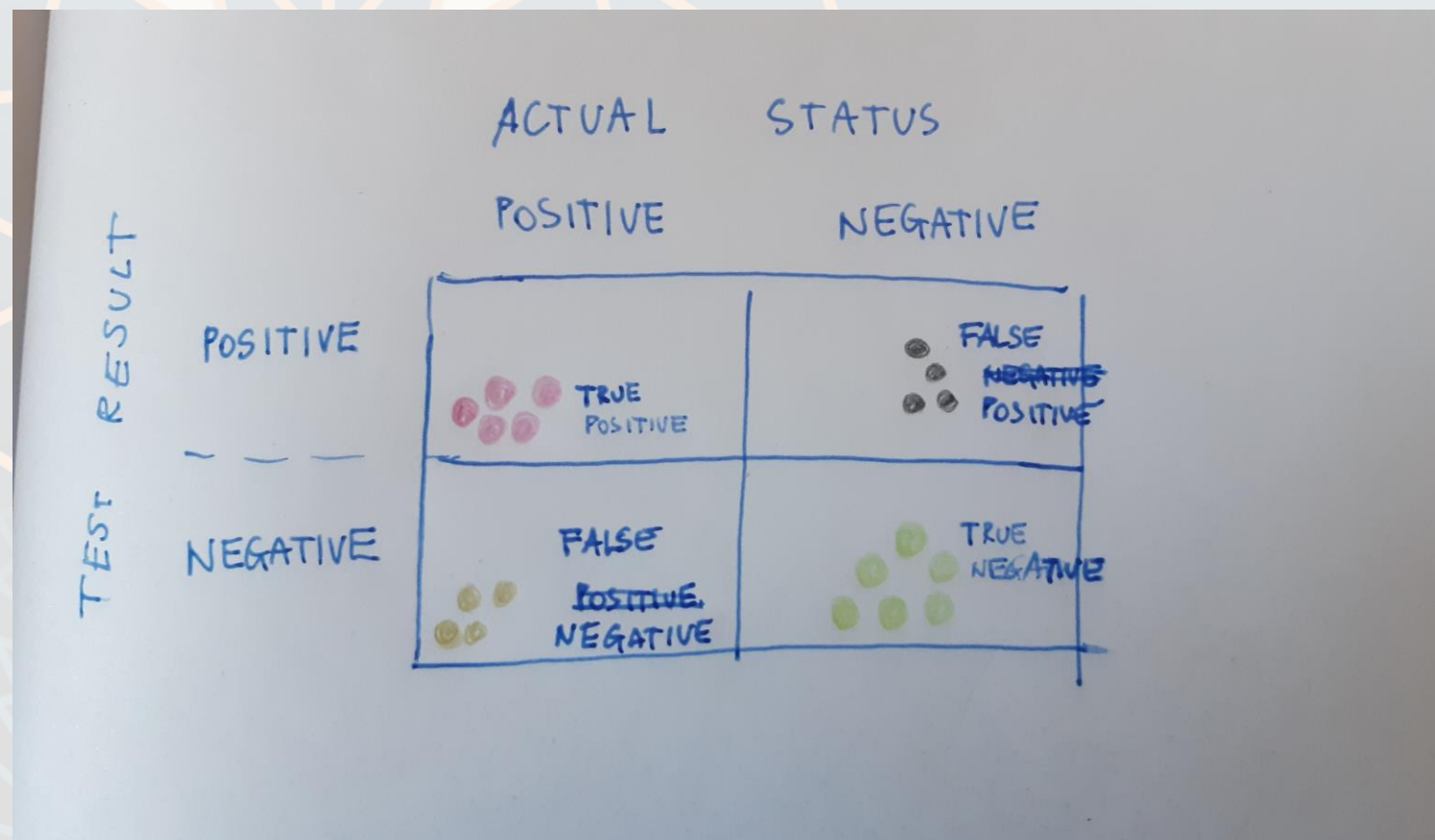
Specifičnost: $1 - \text{FPR}$

Senzitivnost: TPR

Idealno:

Specifičnost: 1 (nobenega zdravega v vzorcu bolnih)

Senzitivnost: 1 (nobenega bolnega test ne spregleda)



3. OSNOVE STATISTIKE – SENZITIVNOST, SPECIFIČNOST

Dva scenarija uporabe testov za rak dojke:

I

Presejanje
(mamogram)

Senzitivnost ~ 1
Specifičnost ~ 0.93

Diagnostika
(biopsija)

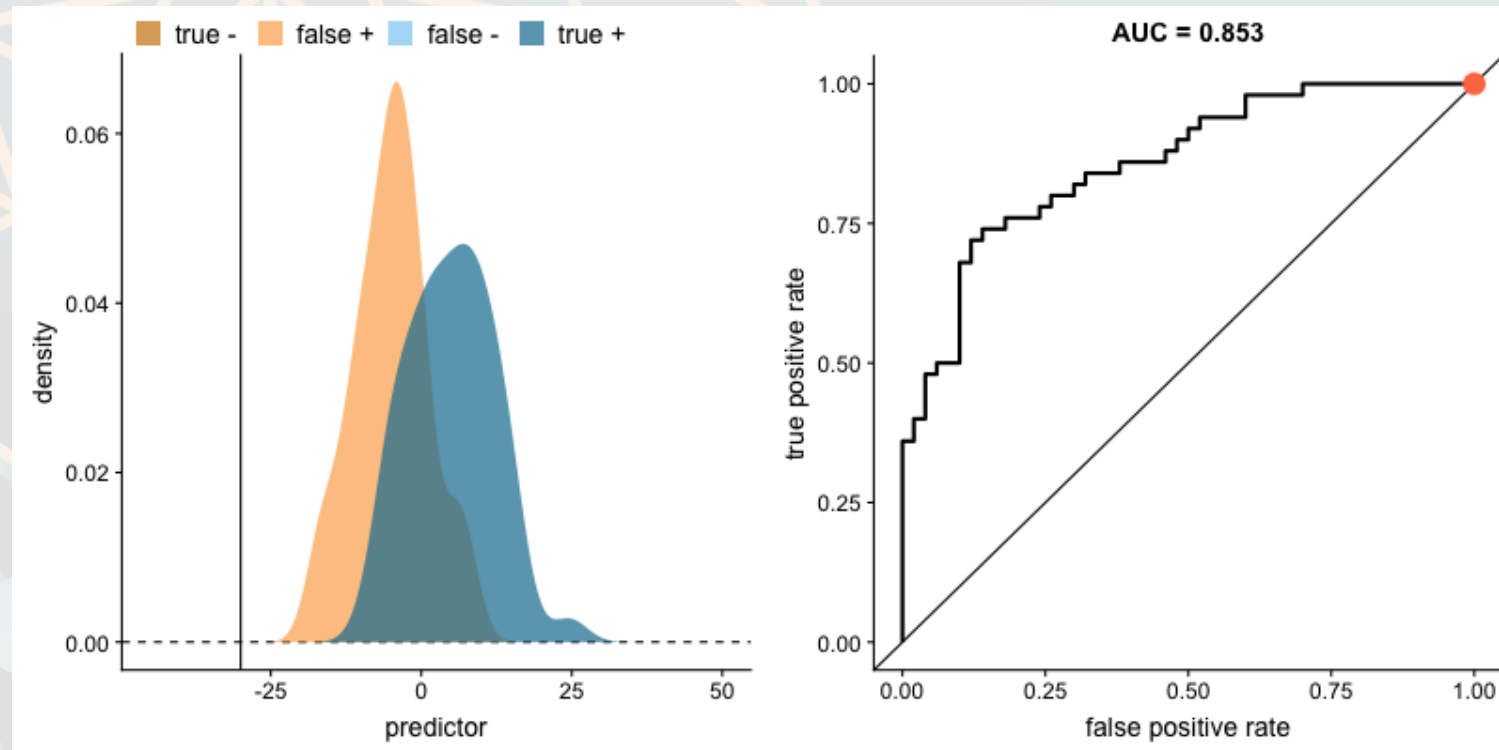
Senzitivnost ~ 0.8
Specifičnost ~ 1

3. STATISTIKA KLASIFIKACIJE - ROC

ROC, receiver-operator curve je krivulja, ki povezuje FPR (vodoravna os) in TPR (navpična os).

Navadno pri odločanju uporabljamo zvezno spremenljivko, npr. temperaturo, na podlagi katere ločujemo med bolnimi in obolelimi.

Če se porazdelitvi populacije po tej spremenljivki prekrivata, bo od meje odvisno tudi razmerje med FPR in TPR, ki ga narišemo na grafu.

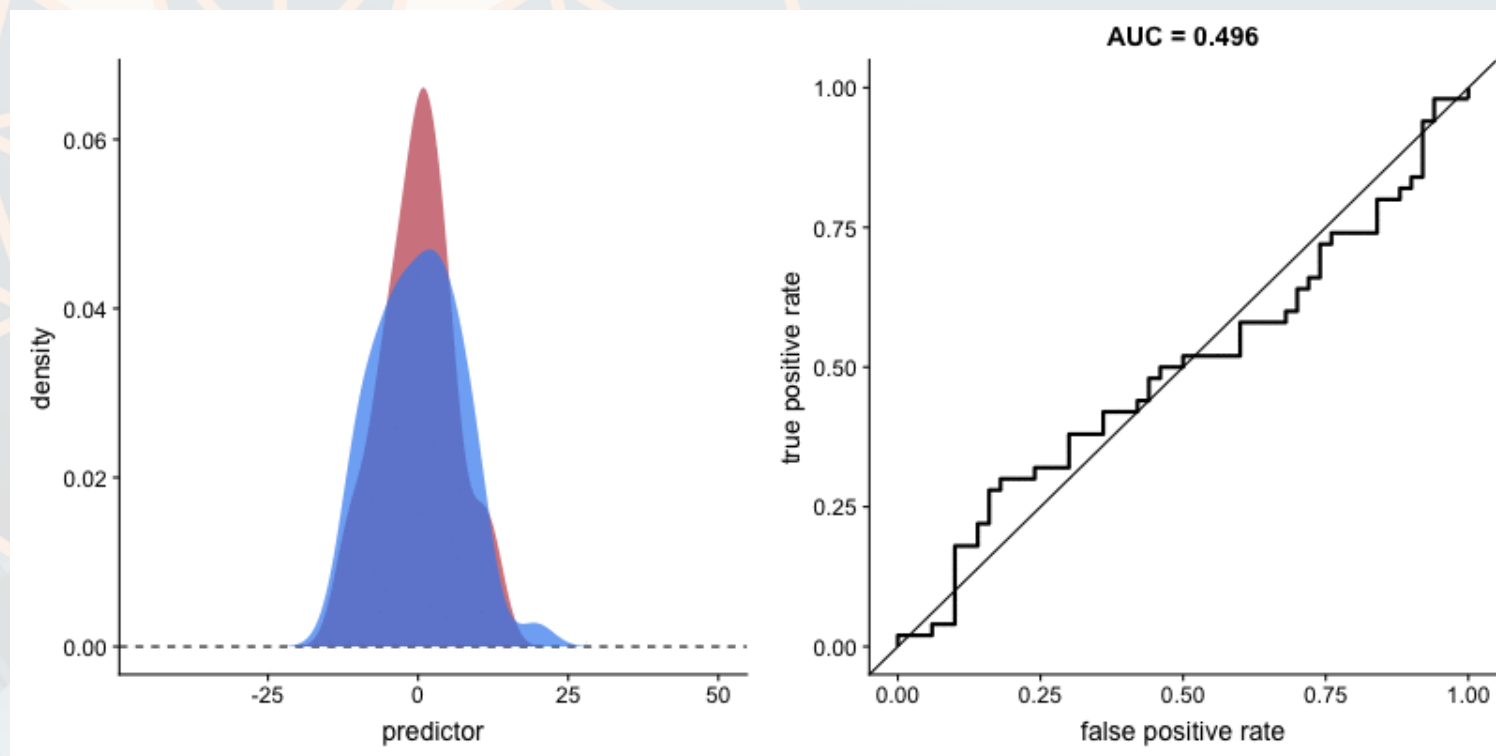


3. STATISTIKA KLASIFIKACIJE - ROC

Oblika ROC je povezana z porazdelitvijo skupin po izbrani napovedni spremenljivki.

Za spremenljivko, ki dobro ločuje med skupinami, skoči TPR takoj na 1 in tam tudi ostane.

Mera, ki primerja ROC je tako integral, area under curve of AUC.



3. STATISTIKA KLASIFIKACIJE

Tako je bila AUC izbrana kot metrika za tekmovanje RIS.

Pozor - točnost, definirana kot delež pravilno napovedanih izidov:

$$A = (TP + TN) / N$$

kjer je N število testiranih, ni dobra ocena, kar nanjo vpliva delež pozitivnih v populaciji. Trivialna negativna napoved je namreč pravilna v deležu, enakem številu negativnih v populaciji, ki je lahko zelo velik za redke bolezni.

