



Dodatak nadzirano učenje - evaluacija

Tablica zabune *

| | | Stvarno stanje <i>Actual value</i> | |
|--|------------------------------|--|--|
| | | Pozitivan <i>Positive</i> | Negativan <i>Negative</i> |
| Predviđeno stanje <i>Prediction outcome</i> | Pozitivan <i>Positive</i> | Točno pozitivan <i>True Positive = TP</i> | Lažno pozitivan <i>False Positive = FP</i> |
| | Negativan <i>Negative</i> | Lažno negativan <i>False Negative = FN</i> | Točno negativan <i>True Negative = TN</i> |

*(engl. *confusion matrix*),
- u statistici se još naziva i kontingencijska tablica
(engl. *contingency table*)

Tablica zabune

- Primjer: model koji dijagnosticira neku bolest.
- Evaluacija je izvršena na skupu od 150 primjera.

| | | Stvarno | |
|-------------------|---------------|----------|---------------|
| | | Bolestan | Nije bolestan |
| Predviđeno stanje | Bolestan | TP = 6 | FP = 12 |
| | Nije bolestan | FN = 2 | TN = 130 |

Osnovne evaluacijske mjere (1)

- Ukupan broj primjera = $TP + TN + FP + FN = 150$
- Broj točno klasificiranih primjera = $TP + TN = 136$
- Broj pozitivnih primjera = $TP + FN = 8$
- Broj negativnih primjera = $TN + FP = 142$
- **Točnost** (*accuracy*) je udio točno klasificiranih primjera u skupu svih primjera.
 - $Točnost = (TP + TN) / (TP + TN + FP + FN) = 90,7\%$

Osnovne evaluacijske mjere (2)

- **Preciznost** (*precision*) je udio točno klasificiranih primjera u skupu pozitivno klasificiranih primjera.
 - $\text{Preciznost} = \text{TP} / (\text{TP} + \text{FP}) = 33,3\%$
 - Još se koristi naziv *positive predictive value (PPV)*.
 - Vidimo da ovaj model nije precizan.
- **Odziv** (*recall*) je udio točno klasificiranih primjera u skupu svih pozitivnih primjera.
 - $\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) = 75,0\%$
 - Drugi nazivi: **osjetljivost** (*sensitivity*), *hit rate*, *true positive rate (TPR)*.
 - Želimo da odziv bude što veći (npr. želimo da otkrijemo što veći broj ljudi koji imaju neku bolest).



Osnovne evaluacijske mjere (3)

- **Specifičnost** (*specificity*) je udio točno klasificiranih primjera u skupu svih negativnih primjera.
 - Specifičnost = $TN / (TN + FP) = 91,6\%$
 - Još se koristi naziv *true negative rate (TNR)*.
 - Ako je specifičnost 100% onda su svi zdravi ljudi prepoznati kao zdravi. Manja specifičnost znači da dio zdravih ljudi dobiva krivu dijagnozu.
- Specifičnost se obično koristi u paru s osjetljivosti.
- Različite vrste mjera koristit će se u različitim područjima (domenama) ponajviše ovisno o uobičajenoj raspodijeli pozitivnih i negativnih primjera.

Osnovne evaluacijske mjere (4)

- Niti jedna od do sada navedenih mjera nije dovoljna sama za sebe.
- Primjer je model koji svaki primjer klasificira kao pozitivan kada se evaluira na skupu od 1000 primjera od kojih je 10 negativnih.
 - $TP = 990$, $TN = 0$, $FP = 10$, $FN = 0$
 - Točnost je 99%, odziv je 100%, a preciznost je 99%

- F mjera je harmonijska sredina preciznosti i odziva.
- U danom primjeru $F=46,1\%$

$$F = \frac{2 \cdot \text{preciznost} \cdot \text{odziv}}{\text{preciznost} + \text{odziv}}$$

- U općem slučaju međusobnu važnost preciznosti i odziva kontroliramo parametrom β (ako nam je važniji odziv koristit ćemo veću vrijednost parametra β).

$$F_{\beta} = \frac{(1 + \beta^2) \cdot \text{preciznost} \cdot \text{odziv}}{\beta^2 \cdot \text{preciznost} + \text{odziv}}$$