# Poročilo o analizi Bayesove mreže za napoved DEATH\_EVENT

## 1. Uvod

V tej analizi smo razvili napovedni model na osnovi Bayesove mreže z uporabo podatkov o bolnikih s srčnim popuščanjem. Glavni cilj je bil napovedati verjetnost smrtnosti (DEATH\_EVENT) na osnovi ključnih spremenljivk: age\_group, time\_group, serum\_creatinine\_group in ejection\_fraction. Uporabljeni so bili diskretizirani podatki, pridobljeni iz datoteke heart\_failure\_discretized.csv.

## 2. Struktura mreže

Mreža je bila zgrajena ročno, z neposredno povezavo vseh izbranih atributov proti ciljni spremenljivki DEATH\_EVENT. To omogoča neposredno razumevanje vpliva posameznih spremenljivk na izid.  
  
Ciljna spremenljivka: DEATH\_EVENT  
Vhodne spremenljivke:  
- age\_group  
- time\_group  
- serum\_creatinine\_group  
- ejection\_fraction

## 3. Rezultati klasifikacije

Točnost modela: 0.916 (274 pravilnih napovedi od 299 primerov)

Natančnost po razredih:

- DEATH\_EVENT = No: 0.951 (193/203)

- DEATH\_EVENT = Yes: 0.844 (81/96)

Matrika zmede (Confusion Matrix):

|  |  |  |
| --- | --- | --- |
|  | Napovedano: No | Napovedano: Yes |
| Dejanski: No | 193 | 10 |
| Dejanski: Yes | 15 | 81 |

Komentar: Model zelo dobro loči med živimi in umrlimi pacienti, še posebej pri pravilni identifikaciji tistih, ki niso umrli (TN = 193). Občasno pa napačno klasificira preživele kot umrle in obratno, kar je ključno za zdravstvene aplikacije.

## 4. ROC krivulja

AUC (Area Under Curve): 0.967  
  
To je odličen rezultat, ki nakazuje, da ima model zelo dobro ločevalno sposobnost med pozitivnimi in negativnimi primeri. Krivulja je blizu zgornjega levega kota, kar pomeni visoko občutljivost in specifičnost.

## 5. Kalibracija

Kalibracijska krivulja za DEATH\_EVENT = No kaže na odstopanje od popolne kalibriranosti (H-L = 37.5), kar pomeni, da je napovedana verjetnost nekoliko precenjena oz. podcenjena v nekaterih razredih. Priporoča se izboljšava z uporabo večje količine podatkov ali preurejanjem razredov.

## 6. Tornado diagram (občutljivost)

Za DEATH\_EVENT = Yes:

- Največ vpliva ima time\_group = srednje in kratko, sledijo serum\_creatinine\_group = visoka in age\_group = mlajši.

Za DEATH\_EVENT = No:

- Prav tako prevladuje time\_group, s tem da imata srednje in kratko najčinkovitejši vpliv.

Komentar: Spremenljivka time\_group ima največji vpliv na izid, kar je skladno s preteklo interpretacijo WoE in IV analize. Sledi serumski kreatinin, ki je znan kazalnik odpovedi ledvic in težjega srčnega stanja.

## 7. Zaključek

Model Bayesove mreže za napoved smrtnosti pacientov s srčnim popuščanjem kaže zelo dobre napovedne sposobnosti (AUC = 0.967, točnost 91,6%). Glavni dejavniki tveganja so: krajše trajanje spremljanja (time\_group), visoka vrednost seruma kreatinina, in nizka iztisna frakcija.  
  
Za izboljšanje lahko priporočam:  
- Še eno testiranje z validacijskim podatkovnim nizom  
- Razmislek o širitvi mreže z dodatnimi atributi (npr. anaemia, diabetes tiste kar imajo boljša povezanost)