Umetna inteligenca 2021-2022

## Seminarska naloga 1

**Klasifikacijski problem**

## Aljaž Hribar 30 November 2021

**Ocenjevanje in konstrukcija atributov**

najprej faktoriziramo vse atribute, ki niso zvezni

ucna$regija<-as.factor(ucna$regija) testna$regija<-as.factor(testna$regija) ucna$namembnost<-as.factor(ucna$namembnost) testna$namembnost<-as.factor(testna$namembnost) ucna$oblacnost<-as.factor(ucna$oblacnost) testna$oblacnost<-as.factor(testna$oblacnost) summary(ucna)

## datum regija stavba

## Length:24125 vzhodna:11315 Min. : 1.00 ## Class :character zahodna:12810 1st Qu.: 39.00 ## Mode :character Median : 79.00

## Mean : 87.49

## 3rd Qu.:135.00

## Max. :193.00

## namembnost povrsina leto\_izgradnje temp\_zraka ## izobrazevalna :13301 Min. : 329.3 Min. :1903 Min. :-7.20 ## javno\_storitvena : 2979 1st Qu.: 4106.6 1st Qu.:1950 1st Qu.:10.00 ## kulturno\_razvedrilna: 3263 Median : 6763.3 Median :1970 Median :20.00

## poslovna : 3057 Mean :10958.1 Mean :1970 Mean :19.15 ## stanovanjska : 1525 3rd Qu.:14409.3 3rd Qu.:2000 3rd Qu.:28.30 ## Max. :79000.4 Max. :2017 Max. :41.70

## temp\_rosisca oblacnost padavine pritisk smer\_vetra ## Min. :-19.400 0:3090 Min. :-1.0000 Min. : 997.2 Min. : 0.0

## 1st Qu.: -2.800 2:8390 1st Qu.: 0.0000 1st Qu.:1011.9 1st Qu.: 70.0

## Median : 2.800 4:4514 Median : 0.0000 Median :1015.9 Median :140.0

## Mean : 3.816 6:5126 Mean : 0.3113 Mean :1017.1 Mean :156.6 ## 3rd Qu.: 11.100 8:2950 3rd Qu.: 0.0000 3rd Qu.:1021.8 3rd Qu.:250.0 ## Max. : 25.000 9: 55 Max. :56.0000 Max. :1040.9 Max. :360.0

## hitrost\_vetra poraba

## Min. : 0.000 Min. : 0.00 ## 1st Qu.: 2.100 1st Qu.: 53.48

## Median : 3.600 Median : 112.90 ## Mean : 3.756 Mean : 224.55 ## 3rd Qu.: 5.100 3rd Qu.: 215.41

## Max. :12.400 Max. :2756.54

opazimo da lahko je smer vetra podana koz zvezni podatek ampak bi nam bila bolj uporabna kot diskretni zato jo faktoriziramo

table(ucna$smer\_vetra) ##

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ## brezveterje | jug | jugo\_vzhod | jugo\_zahod | sever | severo\_vzhod |
| ## 2482 | 4210 | 3325 | 1044 | 1740 | 2670 |
| ## severo\_zahod | vzhod | zahod |  |  |  |
| ## 2831 | 3799 | 2024 |  |  |  |

iz atributa “datum” lahko generiramo nov atribut “season”, ki nam pove letni čas meritve in atribut “vikend” ki nam pove ali je na ta datum bil vikend ali delovni teden

table(ucna$season)

##

## Fall Spring Summer Winter ## 6741 4235 5140 8009

table(ucna$vikend)

##

## FALSE TRUE ## 17184 6941

prav tako lahko iz atributa “poraba” izvlečemo atributa “dosedanja\_povpreča” in “dosedanja\_skupna” ki nam povesta kolikšna je povprečna in skupna poraba stavbe do vključno trenutnega datuma meritve

summary(ucna$dosedanja\_povprecna)

## Min. 1st Qu. Median Mean 3rd Qu. Max. ## 3.616 60.733 118.774 228.265 204.453 2196.688

summary(ucna$dosedanja\_skupna)

## Min. 1st Qu. Median Mean 3rd Qu. Max. ## 3.7 4531.1 12638.2 28874.0 30728.0 424149.1

sedaj lahko izločimo atribut stavba saj je za klasifikacijo odvečen atribut, ki bi samo kvaril modele

ucna$stavba<-NULL testna$stavba<-NULL

z attrEval() funkcijo ocenimo atribute,

library(CORElearn)

## Warning: package 'CORElearn' was built under R version 4.0.5

sort(attrEval(namembnost~., ucna,"Relief"),decreasing =TRUE)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ## | dosedanja\_povprecna | povrsina | poraba | dosedanja\_skupna |
| ## | 0.061888110 | 0.035969465 | 0.035938575 | 0.035583211 |
| ## | leto\_izgradnje | regija | padavine | season |
| ## | 0.016055510 | 0.000000000 | -0.000613520 | -0.002901554 |
| ## | vikend | temp\_zraka | temp\_rosisca | pritisk |
| ## | -0.005015544 | -0.007857875 | -0.011347804 | -0.013097497 |
| ## | oblacnost | smer\_vetra | hitrost\_vetra | datum |
| ## | -0.013471503 | -0.015046632 | -0.015718982 | -0.057616580 |

sort(attrEval(namembnost~., ucna,"ReliefFequalK"),decreasing =TRUE)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ## | leto\_izgradnje | povrsina | regija | dosedanja\_povprecna |
| ## | 0.242451468 | 0.166456481 | 0.139201471 | 0.136221147 |
| ## | poraba | dosedanja\_skupna | datum | temp\_zraka |
| ## | 0.115882015 | 0.105300619 | 0.102695830 | 0.082560008 |
| ## | smer\_vetra | pritisk | temp\_rosisca | hitrost\_vetra |
| ## | 0.067548928 | 0.062092321 | 0.047547866 | 0.040365334 |
| ## | oblacnost | season | vikend | padavine |
| ## | 0.038386700 | 0.027151592 | 0.007127530 | 0.004231328 |
| sort(attrEval(namembnost~., ucna,"ReliefFexpRank"),decreasing =TRUE) | | | | |
| ## | leto\_izgradnje | povrsina | regija | datum |
| ## | 0.260707133 | 0.166830040 | 0.164866503 | 0.119054407 |
| ## | dosedanja\_povprecna | poraba | temp\_zraka | dosedanja\_skupna |
| ## | 0.115823728 | 0.095175641 | 0.094856759 | 0.085446344 |
| ## | smer\_vetra | pritisk | temp\_rosisca | oblacnost |
| ## | 0.079984195 | 0.074110398 | 0.056642911 | 0.048480491 |
| ## | hitrost\_vetra | season | vikend | padavine |
| ## | 0.046881204 | 0.033750883 | 0.008384743 | 0.004979457 |
| iz avaluacije atributov opazimo da imajo atributi ki opisujejo vremenske razmere in datum meritev le teh | | | | |

zelo majheno povezavo z namembnostjo stavbe zato jih lahko izločimo

ucna$temp\_zraka<-NULL ucna$pritisk<-NULL ucna$temp\_rosisca<-NULL ucna$padavine<-NULL ucna$hitrost\_vetra<-NULL ucna$smer\_vetra<-NULL ucna$oblacnost<-NULL ucna$datum<-NULL ucna$season<-NULL ucna$vikend<-NULL testna$temp\_zraka<-NULL testna$pritisk<-NULL testna$temp\_rosisca<-NULL testna$padavine<-NULL testna$hitrost\_vetra<-NULL testna$smer\_vetra<-NULL testna$oblacnost<-NULL testna$datum<-NULL testna$season<-NULL testna$vikend<-NULL ucna$dosedanja\_skupna<-NULL

testna$dosedanja\_skupna<-NULL

# gradnja modelov

### odločitveno drevo

najprej sem zgradil model z vsemi atributi

rpart.plot(dt)

## Warning: labs do not fit even at cex 0.15, there may be some overplotting



izobrazevalna

.55 .12 .14 .13 .06

100%

*yes* **povrsina >= 3495** *no*

kulturno\_razvedrilna

.07 .38 .40 .14 .00

18%

**leto\_izgradnje < 1981**

izobrazevalna

.66 .07 .08 .12 .08

82%

**leto\_izgradnje < 1972** izobrazevalna

.46 .08 .17 .22 .07

37%

**leto\_izgradnje >= 1991**

kulturno\_razvedrilna

.00 .10 .75 .15 .00

8%

**dosedanja\_povprecna >= 8.7**

izobrazevalna

.81 .06 .00 .05 .08

45%

**povrsina < 29e+3** javno\_storitvena

.00 .67 .00 .33 .00

3%

**povrsina >= 34e+3**

poslovna

.13 .07 .39 .42 .00

13%

**leto\_izgradnje < 1981**

javno\_storitvena

.14 .64 .09 .14 .00

9%

**regija = vzhodna** izobrazevalna

.50 .00 .00 .50 .00

3%

**povrsina < 1914**

izobrazevalna

.86 .02 .00 .03 .09

43%

**regija = vzhodna** izobrazevalna

.76 .00 .00 .06 .18

21%

**leto\_izgradnje >= 1964**

izobrazevalna

.64 .09 .05 .11 .11

24%

**povrsina >= 16e+3** izobrazevalna

.40 .15 .09 .18 .18

14%

**leto\_izgradnje >= 2014**

javno\_storitvena

.00 .87 .13 .00 .00

7%

poslovna **dosedanja\_p**k**o**u**v**lt**p**u**r**r**e**n**c**o**n**\_**a**ra**<**zv**7**e**7**drilna

.00 .00 .23 .77 .00 .00 .50 .50 .00 .00

kulturno\_razvedrilna

.00 .12 .88 .00 .00

7%

**leto\_izgradnje >= 2010**

6%

**leto\_izgradnje >= 1988**

2%

**povrsina >= 2733**

izobrazevalna

.96 .04 .00 .00 .00

21%

**povrsina >=**izo**4**b**2**r**9**a**6**zevalna

.67 .33 .00 .00 .00

3%

**povrsina < 4135**

izobrazevalna

.56 .00 .00 .11 .33

11%

**dosedanja\_povprecna >= 164**

stanovanjska

.11 .26 .16 .16 .32

8%

**povrsina < 12e+3**

kulturno\_razvedrilna

.22 .11 .50 .17 .00

8%

**regija = vz**k**h**u**o**lt**d**u**n**rn**a**o\_razvedrilna

.00 .00 .75 .25 .00

5%

**leto\_izgradnje >= 1974**

stanovanjska

.41 .00 .00 .15 .45

9%

**povrsina < 4561**

izobrazevalna

.79 .00 .00 .21 .00

6%

**leto\_izgradnje**iz**>**o**=**br**2**a**0**ze**1**v**5**alna

.50 .00 .00 .50 .00

3%

**povrsina >= 10e+3**

izobrazevalna

.67 .33 .00 .00 .00

3%

**povrsina >= 5275**

kulturno\_razvedrilna

.00 .00 .60 .40 .00

2%

**povrsina < 46e+3**

stanovanjska

.16 .00 .00 .21 .63

6%

**leto\_izgradnje < 1941**

stanovanjska

.13 .31 .19 .00 .37

7%

**leto\_izgradnje >**s**=**ta**2**n**0**o**1**v**2**anjska

.16 .38 .00 .00 .46

5%

**povrsina >= 9050**

stanovanjska

.20 .00 .00 .00 .80

5%

**dosedanja\_povprecna >= 128**

stanovanjska

.00 .46 .00 .00 .54

5%

**povrsina < 4820**

stanovanjska

.00 .25 .00 .00 .75

3%

**povrsina >= 7153**

stanovanjska

.43 .00 .00 .00 .57

2%

**povrsina < 9409**

izobrazevalna izobrazevalna izobrazevalna izobrazevalna stanovanjska

poslovna

izobrazevalna poslovna

izobrazevalna javno\_storitvena poslovna javno\_storitvena

poslovna

poslovna javno\_storitvena kulturno\_razvedrilna

poslovna kulturno\_razvedrilna

1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00

2% 10% 3% 1% 3% 1% 3% 1% 1% 1% 1% 1% 1% 1% 5% 1% 1% 6%

izobrazevalna javno\_storitvena izobrazevalna poslovna stanovanjska javno\_storitvena izobrazevalna izobrazevalna kulturno\_razvedrilna javno\_storitvena stanovanjska izobrazevalna kulturno\_razvedrilna kulturno\_razvedrilna poslovna javno\_storitvena izobrazevalna javno\_storitvena poslovna

1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00

19% 1% 3% 1% 1% 2% 10% 1% 1% 1% 3% 2% 4% 1% 3% 1% 1% 1% 1%

CA(observed,predicted) ## [1] 0.5093227

predMat<-predict(dt, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 0.9812315

nato pa še poiskusil maksimalizirati točnost z fukcijo wrapper(), ki je vrnila da so najboljšo točnost imeli atributi povrsina in leto\_izgradnje z pričakovano napako 0.007543904

rpart.plot(dt) ## Warning: labs do not fit even at cex 0.15, there may be some overplotting



izobrazevalna

.55 .12 .14 .13 .06

100%

*yes* **povrsina >= 3495** *no*

kulturno\_razvedrilna

.07 .38 .40 .14 .00

18%

**leto\_izgradnje < 1981**

izobrazevalna

.66 .07 .08 .12 .08

82%

**leto\_izgradnje < 1972** izobrazevalna

.46 .08 .17 .22 .07

37%

**leto\_izgradnje >= 1991**

kulturno\_razvedrilna

.00 .10 .75 .15 .00

8%

**dosedanja\_povprecna >= 8.7**

izobrazevalna

.81 .06 .00 .05 .08

45%

**povrsina < 29e+3** javno\_storitvena

.00 .67 .00 .33 .00

3%

**povrsina >= 34e+3**

poslovna

.13 .07 .39 .42 .00

13%

**leto\_izgradnje < 1981**

javno\_storitvena

.14 .64 .09 .14 .00

9%

**regija = vzhodna** izobrazevalna

.50 .00 .00 .50 .00

3%

**povrsina < 1914**

izobrazevalna

.86 .02 .00 .03 .09

43%

**regija = vzhodna** izobrazevalna

.76 .00 .00 .06 .18

21%

**leto\_izgradnje >= 1964**

izobrazevalna

.64 .09 .05 .11 .11

24%

**povrsina >= 16e+3** izobrazevalna

.40 .15 .09 .18 .18

14%

**leto\_izgradnje >= 2014**

javno\_storitvena

.00 .87 .13 .00 .00

7%

poslovna **dosedanja\_p**k**o**u**v**lt**p**u**r**r**e**n**c**o**n**\_**a**ra**<**zv**7**e**7**drilna

.00 .00 .23 .77 .00 .00 .50 .50 .00 .00

kulturno\_razvedrilna

.00 .12 .88 .00 .00

7%

**leto\_izgradnje >= 2010**

6%

**leto\_izgradnje >= 1988**

2%

**povrsina >= 2733**

izobrazevalna

.96 .04 .00 .00 .00

21%

**povrsina >=**izo**4**b**2**r**9**a**6**zevalna

.67 .33 .00 .00 .00

3%

**povrsina < 4135**

izobrazevalna

.56 .00 .00 .11 .33

11%

**dosedanja\_povprecna >= 164**

stanovanjska

.11 .26 .16 .16 .32

8%

**povrsina < 12e+3**

kulturno\_razvedrilna

.22 .11 .50 .17 .00

8%

**regija = vz**k**h**u**o**lt**d**u**n**rn**a**o\_razvedrilna

.00 .00 .75 .25 .00

5%

**leto\_izgradnje >= 1974**

stanovanjska

.41 .00 .00 .15 .45

9%

**povrsina < 4561**

izobrazevalna

.79 .00 .00 .21 .00

6%

**leto\_izgradnje**iz**>**o**=**br**2**a**0**ze**1**v**5**alna

.50 .00 .00 .50 .00

3%

**povrsina >= 10e+3**

izobrazevalna

.67 .33 .00 .00 .00

3%

**povrsina >= 5275**

kulturno\_razvedrilna

.00 .00 .60 .40 .00

2%

**povrsina < 46e+3**

stanovanjska

.16 .00 .00 .21 .63

6%

**leto\_izgradnje < 1941**

stanovanjska

.13 .31 .19 .00 .37

7%

**leto\_izgradnje >**s**=**ta**2**n**0**o**1**v**2**anjska

.16 .38 .00 .00 .46

5%

**povrsina >= 9050**

stanovanjska

.20 .00 .00 .00 .80

5%

**dosedanja\_povprecna >= 128**

stanovanjska

.00 .46 .00 .00 .54

5%

**povrsina < 4820**

stanovanjska

.00 .25 .00 .00 .75

3%

**povrsina >= 7153**

stanovanjska

.43 .00 .00 .00 .57

2%

**povrsina < 9409**

izobrazevalna izobrazevalna izobrazevalna izobrazevalna stanovanjska poslovna izobrazevalna poslovna izobrazevalna javno\_storitvena poslovna javno\_storitvena poslovna poslovna javno\_storitvena kulturno\_razvedrilna poslovna kulturno\_razvedrilna 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00

2% 10% 3% 1% 3% 1% 3% 1% 1% 1% 1% 1% 1% 1% 5% 1% 1% 6%

izobrazevalna javno\_storitvena izobrazevalna poslovna stanovanjska javno\_storitvena izobrazevalna izobrazevalna kulturno\_razvedrilna javno\_storitvena stanovanjska izobrazevalna kulturno\_razvedrilna kulturno\_razvedrilna poslovna javno\_storitvena izobrazevalna javno\_storitvena poslovna

1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00

19% 1% 3% 1% 1% 2% 10% 1% 1% 1% 3% 2% 4% 1% 3% 1% 1% 1% 1%

CA(observed,predicted) ## [1] 0.5005853

predMat<-predict(dt, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 0.9812315

ko sem pognal wraper() za minimizacijo “brier score” sem dobil podobne rezultate saj mi je funcija vrnila “best model: estimated error = 0.01390878 , selected feature subset = namembnost ~ povrsina + leto\_izgradnje”

rpart.plot(dt) ## Warning: labs do not fit even at cex 0.15, there may be some overplotting



izobrazevalna

.55 .12 .14 .13 .06

100%

*yes* **povrsina >= 3495** *no*

kulturno\_razvedrilna

.07 .38 .40 .14 .00

18%

**leto\_izgradnje < 1981**

izobrazevalna

.66 .07 .08 .12 .08

82%

**leto\_izgradnje < 1972** izobrazevalna

.46 .08 .17 .22 .07

37%

**leto\_izgradnje >= 1991**

kulturno\_razvedrilna

.00 .10 .75 .15 .00

8%

**leto\_izgradnje >= 2010**

izobrazevalna

.81 .06 .00 .05 .08

45%

**povrsina < 29e+3** javno\_storitvena

.00 .67 .00 .33 .00

3%

**povrsina >= 34e+3**

poslovna

.13 .07 .39 .42 .00

13%

**leto\_izgradnje < 1981**

javno\_storitvena

.14 .64 .09 .14 .00

9%

**povrsina <** j**9**a**0**vn**2**o\_storitvena

.00 .74 .11 .16 .00

8%

**povrsina < 1962**

kulturno\_razvedrilna

.00 .00 .83 .17 .00

8%

**povrsina >= 737**

izobrazevalna

.86 .02 .00 .03 .09

43%

**leto\_izgradnje < 1950** izobrazevalna

.81 .04 .00 .00 .16

24%

**leto\_izgradnje >= 1964**

izobrazevalna

.64 .09 .05 .11 .11

24%

**povrsina >= 16e+3** izobrazevalna

.40 .15 .09 .18 .18

14%

**leto\_izgradnje >= 2014**

poslovna

.00 .00 .23 .77 .00

6%

**leto\_izgradnje >= 1988**

poslovna

.00 .29 .29 .43 .00

3%

**leto\_izgradnje >= 1959**

poslovna

.00 .00 .40 .60 .00

2%

**leto\_izgradnje >= 2000**

izobrazevalna

.93 .00 .00 .07 .00

19%

**povrsina >**iz**=**o**7**b**3**ra**2**z**0**evalna

.83 .00 .00 .17 .00

8%

**povrsina < 6679**

izobrazevalna

.65 .07 .00 .00 .29

13%

**leto\_izgradnje < 1962**

stanovanjska

.11 .26 .16 .16 .32

8%

**povrsina < 12e+3**

kulturno\_razvedrilna

.22 .11 .50 .17 .00

8%

**leto\_izgradnje >= 1974**

izobrazevalna

.79 .00 .00 .21 .00

6%

**leto\_izgradnj**i**e**zo**>**b**=**ra**2**z**0**e**1**v**5**alna

.50 .00 .00 .50 .00

3%

**povrsina >= 10e+3**

kulturno\_razvedrilna

.27 .13 .60 .00 .00

6%

**leto\_izgradnje >=**ku**1**lt**9**u**7**rn**7**o\_razvedrilna

.18 .00 .82 .00 .00

5%

**povrsina < 12e+3**

kulturno\_razvedrilna

.00 .00 .60 .40 .00

2%

**povrsina < 46e+3**

javno\_storitvena

.00 .50 .50 .00 .00

2%

**povrsina >= 2733**

izobrazevalna

.71 .07 .00 .00 .21

12%

**povrsina >= 8**i**2**z**5**o**0**brazevalna

.47 .13 .00 .00 .40

6%

**leto\_izgradnje >= 1958**

stanovanjska

.13 .31 .19 .00 .37

7%

**leto\_izgradnje >**st**=**an**2**o**0**v**1**a**2**njska

.16 .38 .00 .00 .46

5%

**povrsina >= 9050**

izobrazevalna

.50 .50 .00 .00 .00

2%

**povrsina >= 5275**

stanovanjska

.27 .18 .00 .00 .54

5%

**povrsina < 4418**

stanovanjska

.00 .46 .00 .00 .54

5%

**povrsina < 4820**

kulturno\_razvedrilna

.40 .00 .60 .00 .00

2%

**povrsina >= 5941**

stanovanjska

.00 .25 .00 .00 .75

3%

**povrsina >= 7153**

izobrazevalna

.60 .40 .00 .00 .00

2%

**povrsina < 4190**

izobrazevalna izobrazevalna izobrazevalna javno\_storitvena stanovanjska poslovna izobrazevalna poslovna izobrazevalna javno\_storitvena poslovna javno\_storitvena kulturno\_razvedrilna poslovna poslovna izobrazevalna javno\_storitvena poslovna kulturno\_razvedrilna poslovna

1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00

6% 11% 2% 1% 1% 1% 3% 1% 1% 1% 1% 1% 1% 1% 1% 1% 1% 1% 5% 1%

izobrazevalna poslovna izobrazevalna izobrazevalna stanovanjska javno\_storitvena izobrazevalna izobrazevalna kulturno\_razvedrilna javno\_storitvena stanovanjska izobrazevalna izobrazevalna kulturno\_razvedrilna kulturno\_razvedrilna poslovna javno\_storitvena kulturno\_razvedrilna javno\_storitvena kulturno\_razvedrilna 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 .00 1.00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 .00 1.00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00 .00 1.00 .00 .00 .00 .00 .00 1.00 .00 .00

11% 1% 5% 1% 3% 2% 10% 1% 1% 1% 3% 1% 1% 3% 1% 3% 5% 1% 1% 1%

CA(observed,predicted) ## [1] 0.5005853

predMat<-predict(dt, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 0.9988294

### naivni bayes

tukaj sem uporabil isti postoke kot pri gradnji drevesa in dobil naslednje rezultate: z vsemi atributi

CA(observed, predicted)

## [1] 0.4540552

predMat<-predict(nb, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 0.7545656

wrapper z minimizacijo napake

CA(observed, predicted) ## [1] 0.4909699

predMat<-predict(nb, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 0.7259014

wrapper z minimizacijo brier

CA(observed, predicted) ## [1] 0.4909699

predMat<-predict(nb, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 0.7259014

### k-najbližjih sosedov

CA(observed, predicted) ## [1] 0.4678512

predMat<-predict(knn, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 1.000356

### naključno gozd

CA(observed, predicted) ## [1] 0.5141304

predMat<-predict(rf, testna,type ="prob") brier.score(obsMat, predMat)

## [1] 0.7388644

### SVM

brier.score(obsMat, predMat) ## [1] 0.6977981

CA(observed, predicted) ## [1] 0.5141304

### Umetne nevronske mreže

najprej je bilo potrebno normalizirati zvezne atribute v učni in testni množici nato pa sem dobil te rezultate

CA(observed, predicted) ## [1] 0.5119147

predMat<-predict(nn, testna\_scaled,type ="raw") brier.score(obsMat, predMat)

## [1] 0.8076981

### kobinirani modeli

za kombinirane modele sem se odločil uporabiti modele nevrenoske mreže, naibni bayes in SVM najprej sem poskusil z glasovanjem in dobil

CA(observed, predicted)

## [1] 0.5484532

nato z uteženim glasovanje

CA(observed, predicted)

## [1] 0.4796405

nazadnje pa še z boostingom

CA(observed, predicted) ## [1] 0.5048913