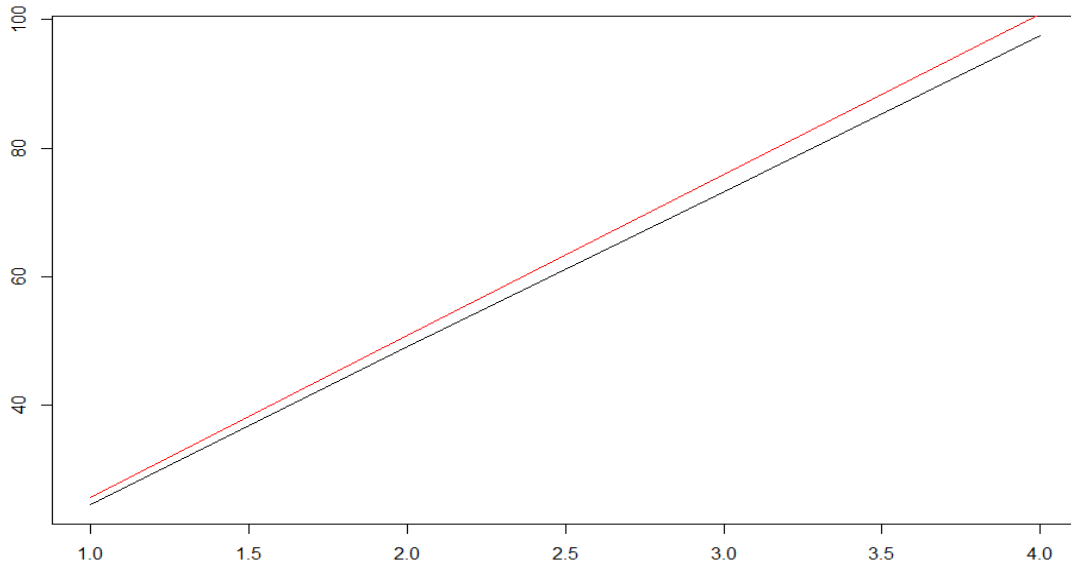


# Vizualizacija podatkov

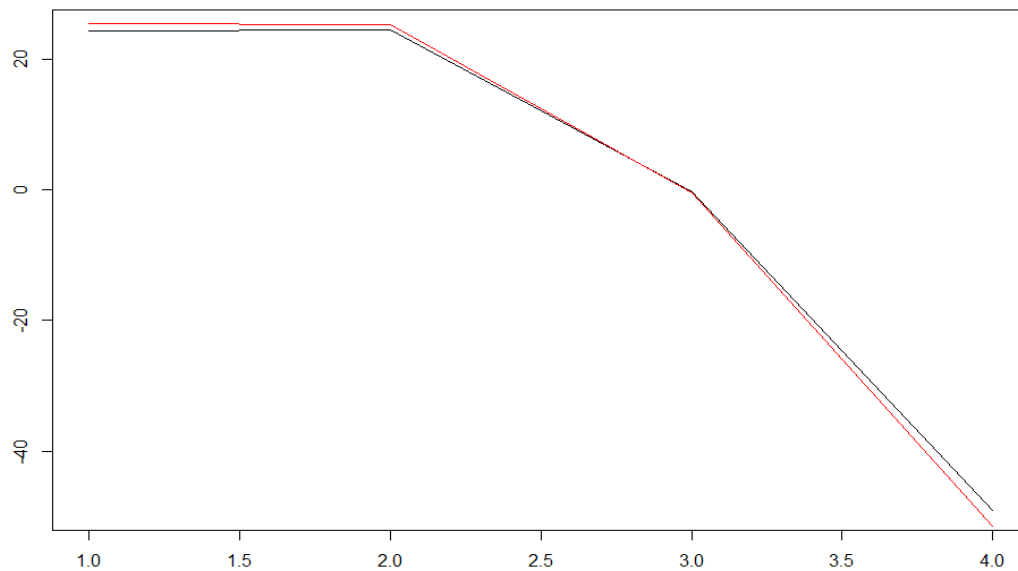
## Povprečne točke domače/gostujoče ekipe

Izračun povprečja točk po čertinah in izris.



## Povprečne točke po čertinah domače/gostujoče ekipe

Izračun povprečja točk doseženih v posamezni četrtni

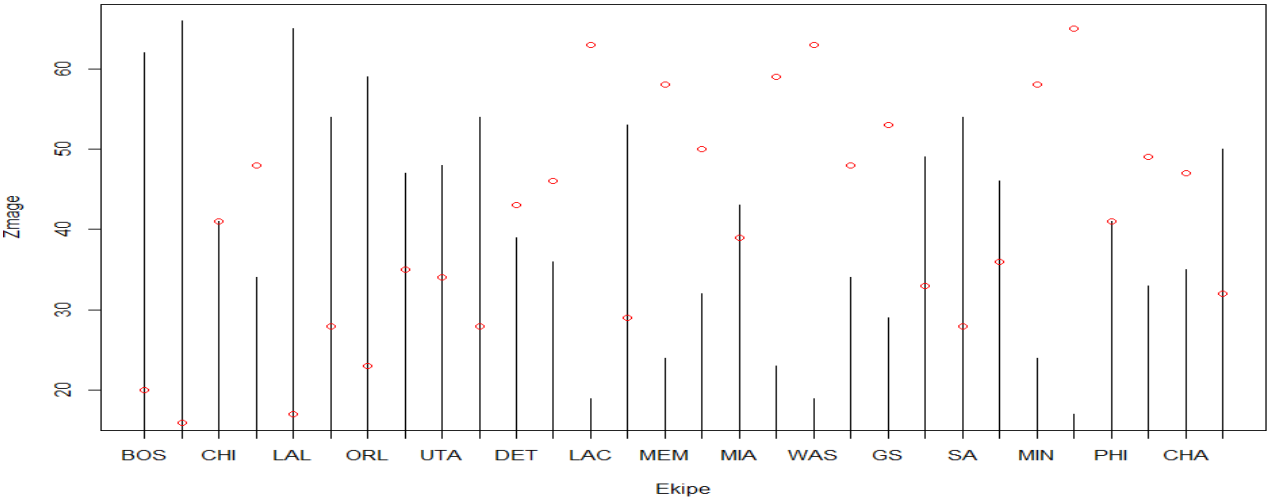


## Zmage in porazi posameznih ekip

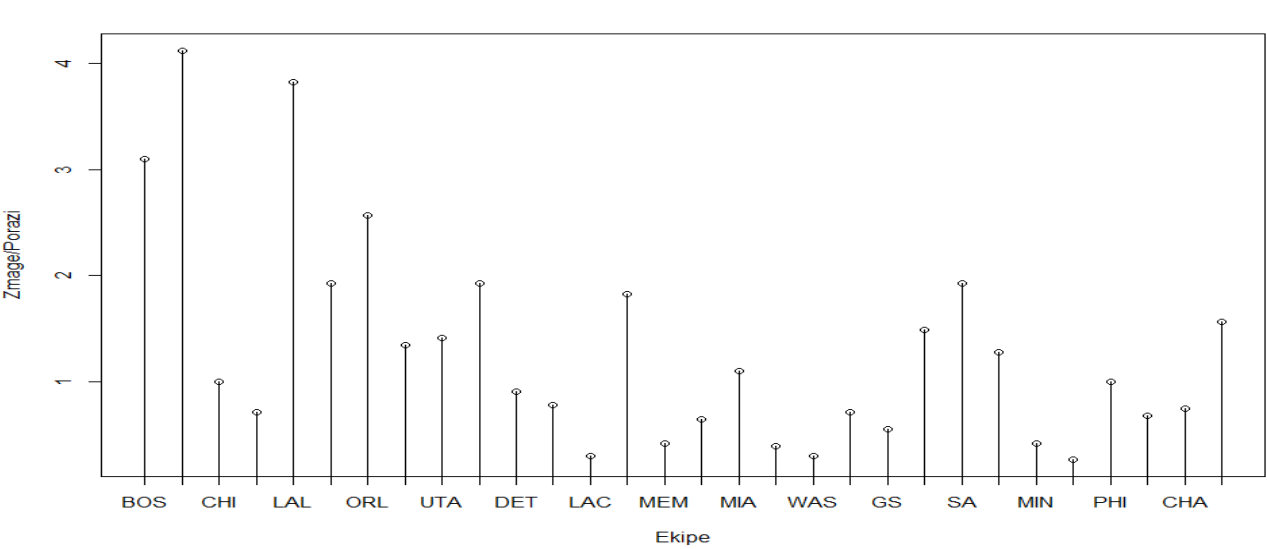
Izracun število zmag in porazov posameznih ekip.

[1] "Ekipa : BOS Zmag: 62 Porazov: 20"  
[1] "Ekipa : CLE Zmag: 66 Porazov: 16"  
[1] "Ekipa : CHI Zmag: 41 Porazov: 41"  
[1] "Ekipa : MIL Zmag: 34 Porazov: 48"  
[1] "Ekipa : LAL Zmag: 65 Porazov: 17"  
[1] "Ekipa : POR Zmag: 54 Porazov: 28"  
[1] "Ekipa : ORL Zmag: 59 Porazov: 23"  
[1] "Ekipa : ATL Zmag: 47 Porazov: 35"  
[1] "Ekipa : UTA Zmag: 48 Porazov: 34"  
[1] "Ekipa : DEN Zmag: 54 Porazov: 28"  
[1] "Ekipa : DET Zmag: 39 Porazov: 43"  
[1] "Ekipa : IND Zmag: 36 Porazov: 46"  
[1] "Ekipa : LAC Zmag: 19 Porazov: 63"  
[1] "Ekipa : HOU Zmag: 53 Porazov: 29"  
[1] "Ekipa : MEM Zmag: 24 Porazov: 58"  
[1] "Ekipa : NY Zmag: 32 Porazov: 50"  
[1] "Ekipa : MIA Zmag: 43 Porazov: 39"  
[1] "Ekipa : OKC Zmag: 23 Porazov: 59"  
[1] "Ekipa : WAS Zmag: 19 Porazov: 63"  
[1] "Ekipa : NJ Zmag: 34 Porazov: 48"  
[1] "Ekipa : GS Zmag: 29 Porazov: 53"  
[1] "Ekipa : NO Zmag: 49 Porazov: 33"  
[1] "Ekipa : SA Zmag: 54 Porazov: 28"  
[1] "Ekipa : PHO Zmag: 46 Porazov: 36"  
[1] "Ekipa : MIN Zmag: 24 Porazov: 58"  
[1] "Ekipa : SAC Zmag: 17 Porazov: 65"  
[1] "Ekipa : PHI Zmag: 41 Porazov: 41"  
[1] "Ekipa : TOR Zmag: 33 Porazov: 49"  
[1] "Ekipa : CHA Zmag: 35 Porazov: 47"  
[1] "Ekipa : DAL Zmag: 50 Porazov: 32"

Vizualizacija v enakem zaporedju, imen se ne vidi vseh.  
Črte so zmage, pike porazi.

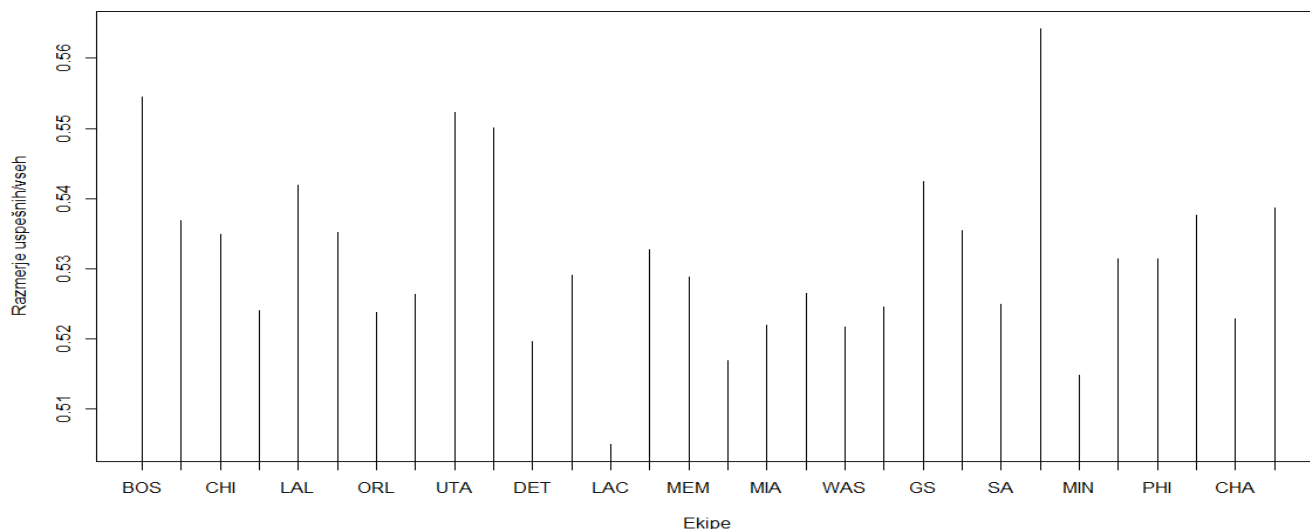


Razmerje zmag in porazov posameznih ekip



## Uspešnost metov proti košu posameznih ekip

Izračun razmerja vseh metov in uspešnih metov posameznih ekip.



- [1] "Ekipa : BOS Vsi meti: 8408 Uspešni meti: 4662 Razmerje: 0.554471931493815"
- [1] "Ekipa : CLE Vsi meti: 8466 Uspešni meti: 4545 Razmerje: 0.536853295535082"
- [1] "Ekipa : CHI Vsi meti: 8898 Uspešni meti: 4759 Razmerje: 0.534839289728029"
- [1] "Ekipa : MIL Vsi meti: 8822 Uspešni meti: 4622 Razmerje: 0.523917479029698"
- [1] "Ekipa : LAL Vsi meti: 9068 Uspešni meti: 4914 Razmerje: 0.541905602117336"
- [1] "Ekipa : POR Vsi meti: 8482 Uspešni meti: 4539 Razmerje: 0.535133223296392"
- [1] "Ekipa : ORL Vsi meti: 8669 Uspešni meti: 4540 Razmerje: 0.523705156304072"
- [1] "Ekipa : ATL Vsi meti: 8536 Uspešni meti: 4493 Razmerje: 0.526358950328022"
- [1] "Ekipa : UTA Vsi meti: 8975 Uspešni meti: 4957 Razmerje: 0.552311977715877"
- [1] "Ekipa : DEN Vsi meti: 8997 Uspešni meti: 4949 Razmerje: 0.550072246304324"
- [1] "Ekipa : DET Vsi meti: 8408 Uspešni meti: 4369 Razmerje: 0.519624167459562"
- [1] "Ekipa : IND Vsi meti: 8970 Uspešni meti: 4745 Razmerje: 0.528985507246377"
- [1] "Ekipa : LAC Vsi meti: 8535 Uspešni meti: 4309 Razmerje: 0.504862331575864"
- [1] "Ekipa : HOU Vsi meti: 8429 Uspešni meti: 4490 Razmerje: 0.532684778740064"
- [1] "Ekipa : MEM Vsi meti: 8388 Uspešni meti: 4435 Razmerje: 0.528731521220792"
- [1] "Ekipa : NY Vsi meti: 8991 Uspešni meti: 4647 Razmerje: 0.51685018351685"
- [1] "Ekipa : MIA Vsi meti: 8517 Uspešni meti: 4445 Razmerje: 0.521897381707174"
- [1] "Ekipa : OKC Vsi meti: 8785 Uspešni meti: 4625 Razmerje: 0.526465566306204"
- [1] "Ekipa : WAS Vsi meti: 8603 Uspešni meti: 4488 Razmerje: 0.52167848424968"
- [1] "Ekipa : NJ Vsi meti: 8516 Uspešni meti: 4467 Razmerje: 0.524542038515735"
- [1] "Ekipa : GS Vsi meti: 9447 Uspešni meti: 5124 Razmerje: 0.542394410924103"
- [1] "Ekipa : NO Vsi meti: 8199 Uspešni meti: 4390 Razmerje: 0.535431150140261"
- [1] "Ekipa : SA Vsi meti: 8175 Uspešni meti: 4291 Razmerje: 0.524892966360856"
- [1] "Ekipa : PHO Vsi meti: 8946 Uspešni meti: 5048 Razmerje: 0.564274536105522"
- [1] "Ekipa : MIN Vsi meti: 8723 Uspešni meti: 4490 Razmerje: 0.514731170468875"
- [1] "Ekipa : SAC Vsi meti: 8792 Uspešni meti: 4672 Razmerje: 0.531392174704277"
- [1] "Ekipa : PHI Vsi meti: 8748 Uspešni meti: 4649 Razmerje: 0.531435756744399"
- [1] "Ekipa : TOR Vsi meti: 8534 Uspešni meti: 4588 Razmerje: 0.537614248886806"
- [1] "Ekipa : CHA Vsi meti: 8264 Uspešni meti: 4321 Razmerje: 0.522870280735721"
- [1] "Ekipa : DAL Vsi meti: 8622 Uspešni meti: 4644 Razmerje: 0.538622129436326"

# Klasifikacija

Učil sem se na učni množici nba0809. Podatke sem predelal tako da so opisi ekip z podatki ki so znani pred začetkom posameznih tekem.

Podatke sem še dodatno obdelal tako da namesto dveh stolpcev(vseh košev in zadetih košev) sem združil v enega samega z razmerjem the vrednosti. Spremenil sem tudi vrednosti atributov o skokih. Tako da sem izračunal razmerje skokov glede na vse skoke pod enim košem.

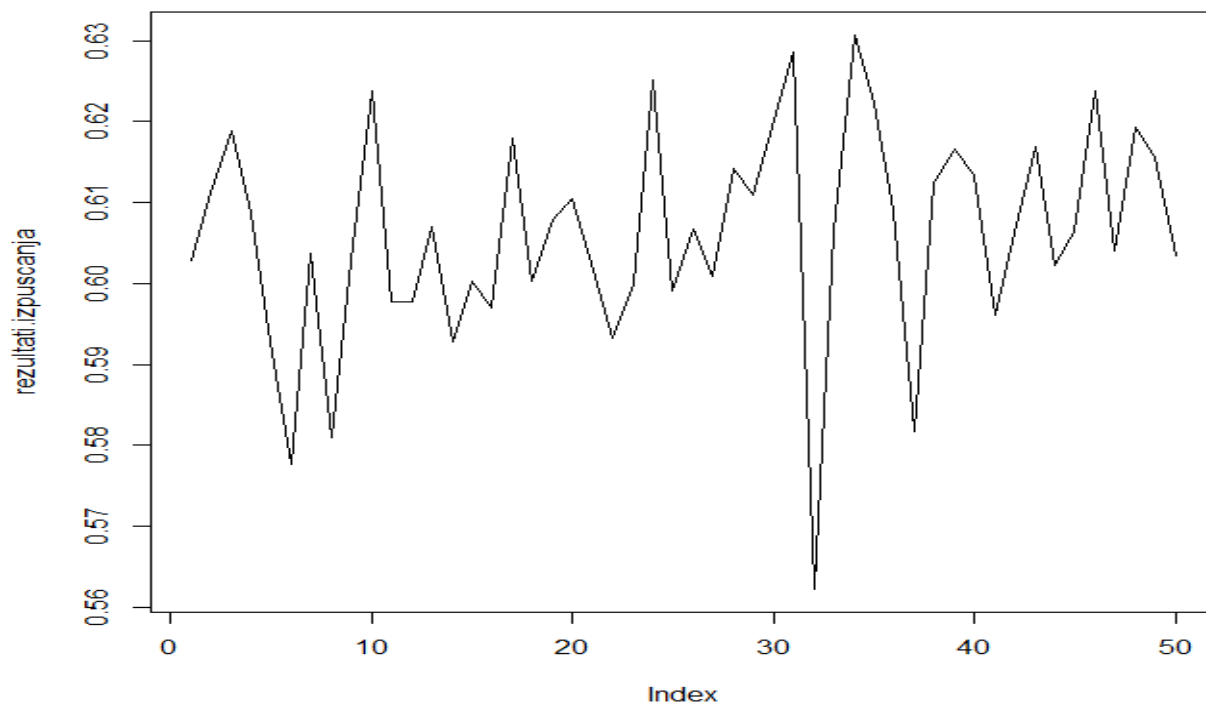
Prvo sem poskušal z **10-kratnim prečnim preverjanjem** z modelom drevesa.

```
res <- errorest(WINNER~., data=train, model = mymodel.coremodel, predict = mypredict.coremodel,
target.model = "tree")
uspesnost = 1-res$error
```

Brez obdelave podatkov je uspešnost okoli 0.6033613 z obdelavo se ne spremeni.

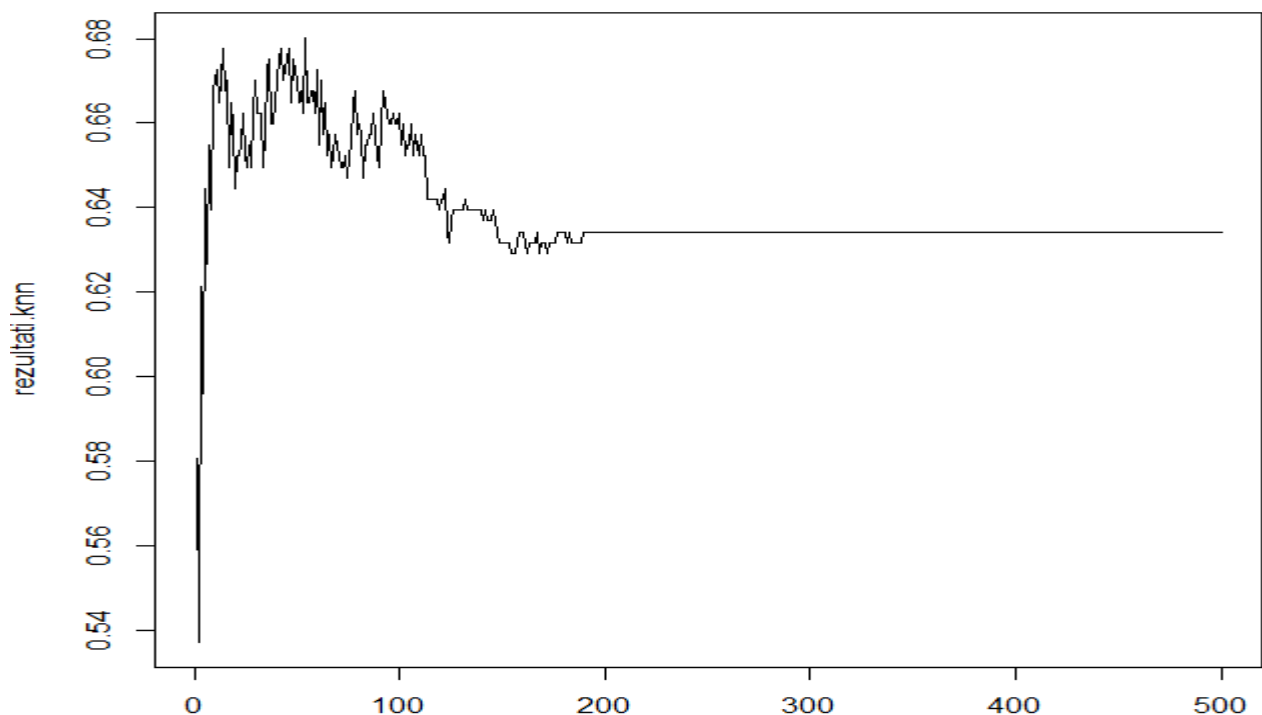
Kar je tudi precej slabo saj ima večinski razred malo nižjo verjetnost.

Ob večkratnem zagonu izračuna rezultat ni bil vedno isti.



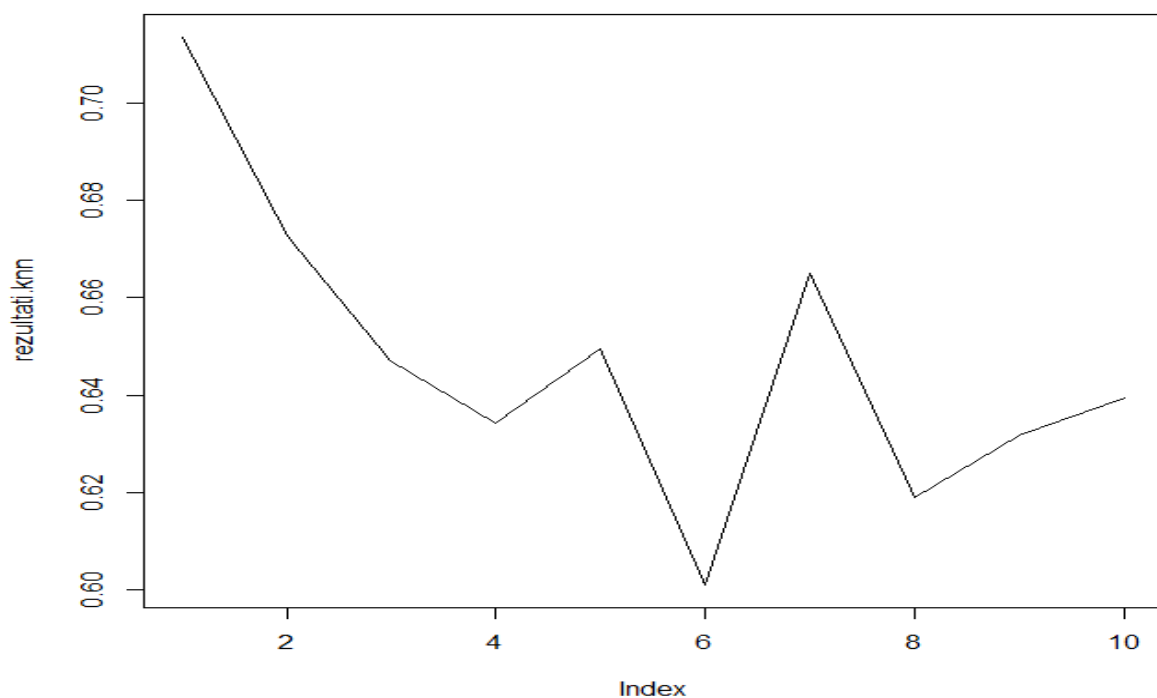
Za izgradnjo modela **k najbližjih sosedov** sem probal ugotoviti katera vrednost k je najboljša. Spodaj so vrednosti k, levo pa je točnost.

Slika pokaže da je to 54, takšen k sem tudi uporabil za izgradnjo modela.



Še za model **nevronske mreže** sem probal z večkratnim poganjanjem dobiti prave vrednosti. Spreminjal sem parametre : maksimalno št. iteracij, spreminjanje uteži in št. skritih nivojev.

Ker je kombinacij preveč sem ubral tako pot : št. skritih nivojev = Index, max. iteracij =  $1000 * \text{Index}$  uteži =  $0.00001 * (10 * \text{Index})$ . Najboljše je pri index = 1.



### Primerjava modelov klasifikacije in njihove ocene.

Večinski razred ima verjetnost : 0.6078925

Brez dodatne obdelave podatkov.

Model	CA	Sensitivity	Specificity	Brier Score
Decision Tree	0.66	0.32	0.85	0.47
Naïve Bayes	0.63	0.42	0.76	0.48
K-nn	0.65	0.2	0.92	0.43
Random Forest	0.67	0.34	0.86	0.42
SVM	0.68	0.36	0.87	
Neural Networks	0.69	0.64	0.72	

Po obdelavi podatkov:

Model	CA	Sensitivity	Specificity	Brier Score
Decision Tree	0.67	0.39	0.83	0.47
Naïve Bayes	0.66	0.2	0.93	0.42
K-nn	0.66	0.11	0.98	0.43
Random Forest	0.67	0.33	0.86	0.41
SVM	0.68	0.32	0.89	
Neural Networks	0.62	0.82	0.5	

Ocenjevanje atributov:

Z pomožno funkcijo wrapper sem pridobil sledeče attribute:

**Odločitveno drevo:**  $\text{WINNER} \sim \text{AWAY\_DRB} + \text{HOME\_TO} + \text{AWAY\_FTA} + \text{AWAY\_2PM} + \text{AWAY\_FTM} + \text{AWAY\_TO} + \text{AWAY\_3PA} + \text{HOME\_FTA} + \text{HOME\_FTM} + \text{HOME\_DRB}$

**Naivni Bayes:**  $\text{WINNER} \sim \text{AWAY\_DRB} + \text{AWAY\_3PA} + \text{AWAY\_FTA} + \text{HOME\_3PM} + \text{AWAY\_ORB} + \text{HOME\_FTA} + \text{HOME\_TO} + \text{AWAY\_TO} + \text{HOME\_FTM} + \text{HOME\_DRB} + \text{HOME\_2PA} + \text{HOME\_2PM}$

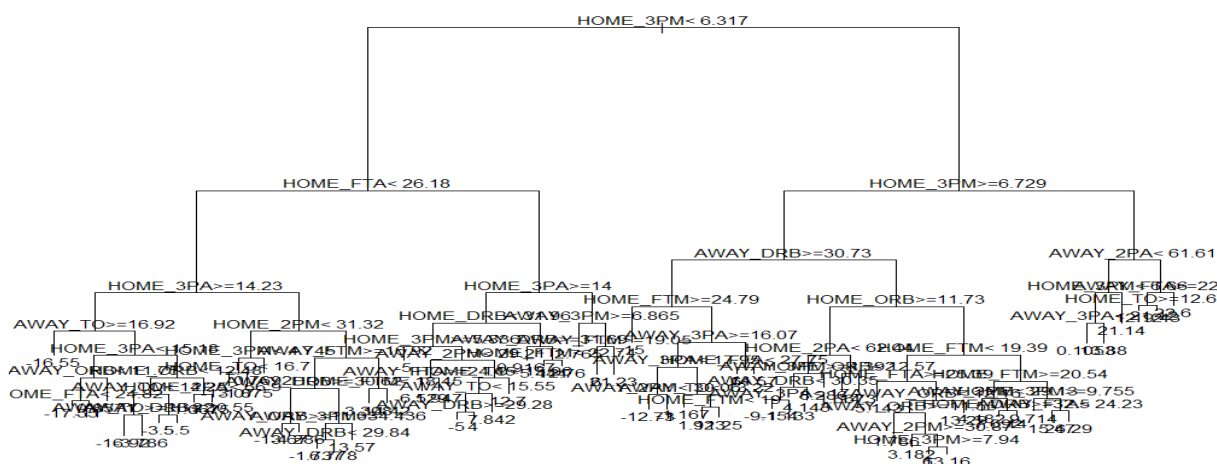


# Regresija

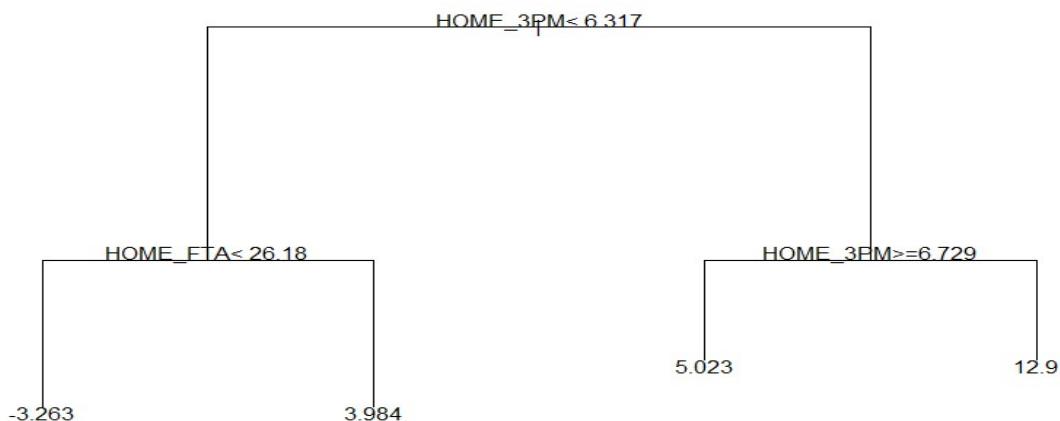
### Rezanje regresijskega drevesa:

Za regresijo sem učno množico dodatno predelal tako da razred ni več kdo je zmagovalec ampak razlika med točkami domačih/gostujočih.

### Grajenje drevesa brez dodatnih parametrov:

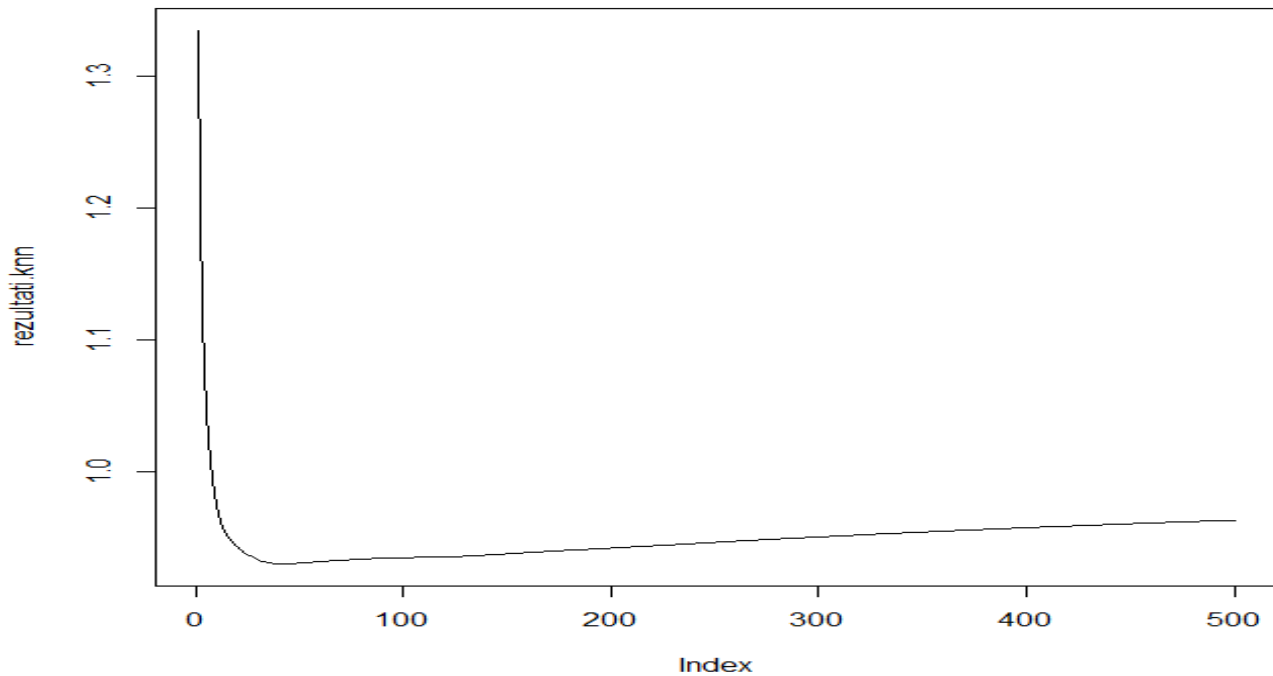


Če pogledamo ocenjene napake za različne vrednosti parametra  $\rho$  vidimo da ima najmanjšo napako pri vrednosti 0.0176958. Če s to vrednostjo porežemo drevo dobimo:

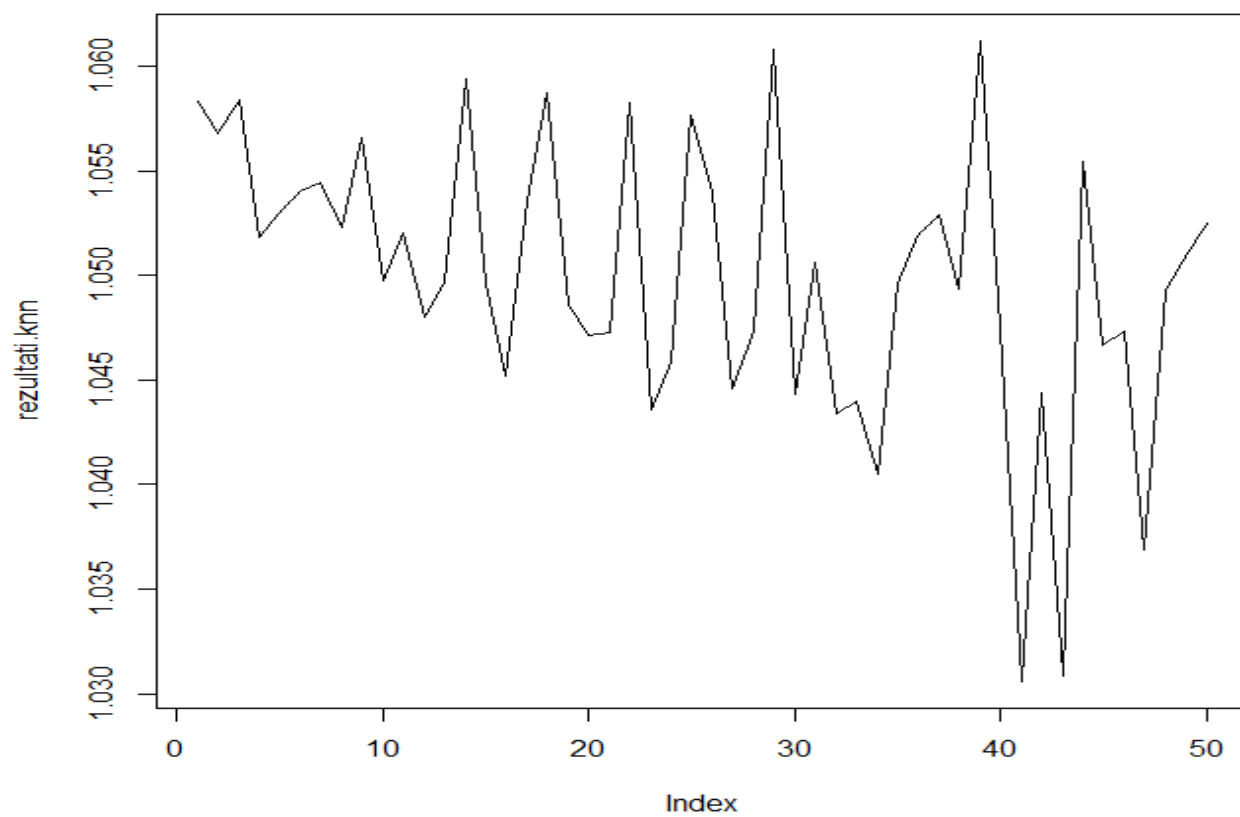


Poleg tega da je porezano drevo bolj preprosto tudi relativna napaka je nižja.  $1.18 \rightarrow 0.97$

Za model **k-najbližjih sosedov** pri regresiji sem prav tako poskušal ugotoviti optimalni k. To je 42.



Enako za **nevronske mreže**. Ampak rezultati so slabi, napaka je vedno večja od 1. Pri zelo kompleksni nevronske mreži je verjetno že prilagajanje. Zato sem izbral standardne nastavitve.



Ocenjevanje atributov:

Z pomožno funkcijo wrapperReg sem pridobil sledeče attribute:

**Regresijsko drevo:** WINNER ~ HOME\_2PA z napako 12.97533

## Primerjava modelov regresije in njihove ocene.

Povprečna vrednost je : 3.291352

Brez dodatne obdelave podatkov.

Model	RMAE	RMSE
Linear model	0.9	0.85
Regression Tree	0.97	0.97
Random Forest	0.91	0.89
SVM	0.89	0.85
K-NN	0.94	0.92
Neural Networks	1.06	1.07

Po obdelavi podatkov:

Model	RMAE	RMSE
Linear model	0.93	0.89
Regression Tree	0.96	0.93
Random Forest	0.92	0.89
SVM	0.92	0.89
K-NN	0.93	0.9
Neural Networks	1.05	1.06