

SEMINARSKA NALOGA

Umetna inteligenca

Tilen Volk 63140336, David Borštner 63170059

Ljubljana 26.11.2018

1. Predstavitev podatkov

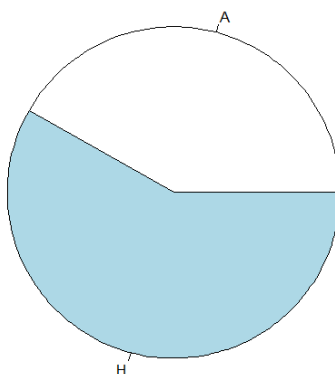
Za predstavitev podatkov je najprej potrebno podatke pridobiti. Podatke sva pridobila tako, da sva shranila vse vrednosti (PTS, 2PM, 2PA, 3PM, 3PA, FTM, FTA, ORB, DRB, AST, STL, TOV, BLK, PF, FG, FGM) v vektorje s katerimi sva si pomagala ter pridobila povprečja kot so: povprečni prosti meti, povprečen met za 2, povprečen met za 3 ter povprečen met z igre. Poleg teh podatkov pa tudi nekaj napredne statistike: število posesti, napadalna učinkovitost, učinkovitost meta iz igre, izgubljene žoge glede na posest, verjetnost skoka v napadu ter verjetnost izvajanja prostih metov.

- Število posesti (pove koliko posesti ima ekipa na 48 min),
uporablja formulo: meti iz igre - skoki v obrambi + izgubljene žoge + $(0.4 * \text{prosti meti})$
- Napadalna učinkovitost (število točk, ki jih ekipa dobi v 100 posestih),
uporablja formulo: $(\text{Točke} * 100) / \text{posesti}$
- Učinkovitost meta iz igre (izenači met za 3 in 2),
uporablja formulo: $(\text{Zadeti meti iz igre} + (0.5 * \text{zadeti meti za 3})) / \text{meti iz igre}$
- Verjetnost izgubljene žoge (koliko žog ekipa izgubi glede na število posesti),
uporablja formulo: $\text{Izgubljene žoge} / \text{posesti}$
- Verjetnost skoka v napadu,
uporablja formulo: $\text{Skoki v napadu} / (\text{Skoki v napadu} + \text{skoki v obrambi nasprotnika})$
- Verjetnost izvajanja prostih metov
uporablja formulo: $\text{Prosti meti} / \text{meti iz igre}$

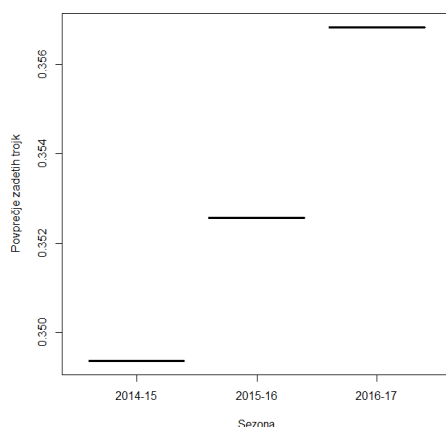
Poleg podatkov z vsake posamezne tekme pa sva še preštela vse osnovne statistike za posamezne ekipe za podane 3 sezone z uporabo for zank.

Prvi graf, ki sva ga naredila predstavlja primerjavo zmag v gosteh z zmagami doma. Na grafu je očitno, da več ekip zmaguje doma kot pa v gosteh. Graf je bil narejen z uporabo funkcije pie, ter podatki pridobljeni z primerjanjem števila točk, ki jih je dosegla posamezna ekipa.

Primerjava zmage domači/gostje

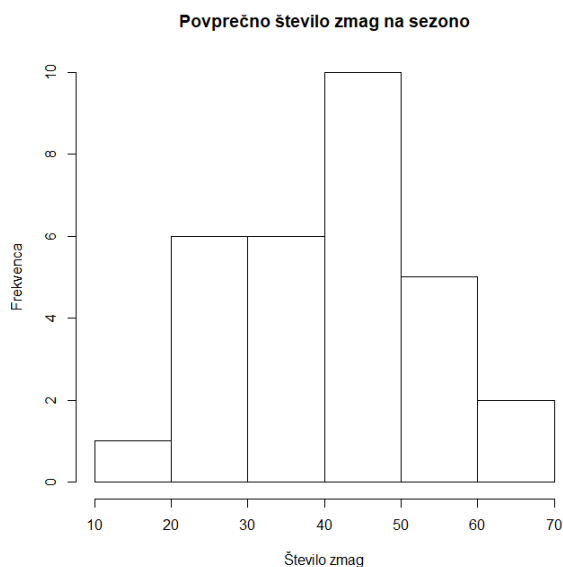


Drugi graf prikazuje odstotek meta za tri točke skozi vse 3 podane sezone. Vse sezone so precej izenačene glede meta, a odstotek meta za 3 vseeno narašča. Graf prikazuje predvsem konsistentnost

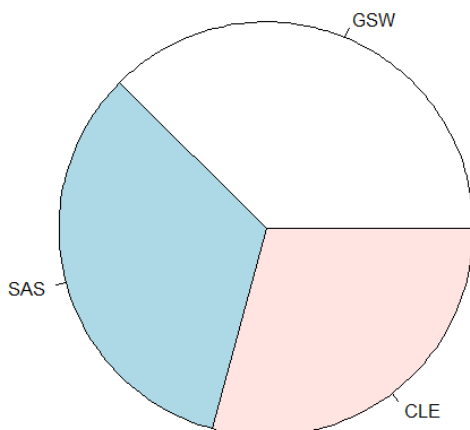


igralcev ter kako se igra v NBA razvija.

Tretji graf prikazuje povprečno število zmag, ki so jih ekipe dosegle na sezono. Dvema ekipama je uspelo doseči med 60 in 70 zmag, ena ekipa je na dnu izstopala z 10 do 20 zmagami, najbolj pogosto pa imajo ekipe med 40 in 50 zmag. Graf je narejen z uporabo funkcije hist.



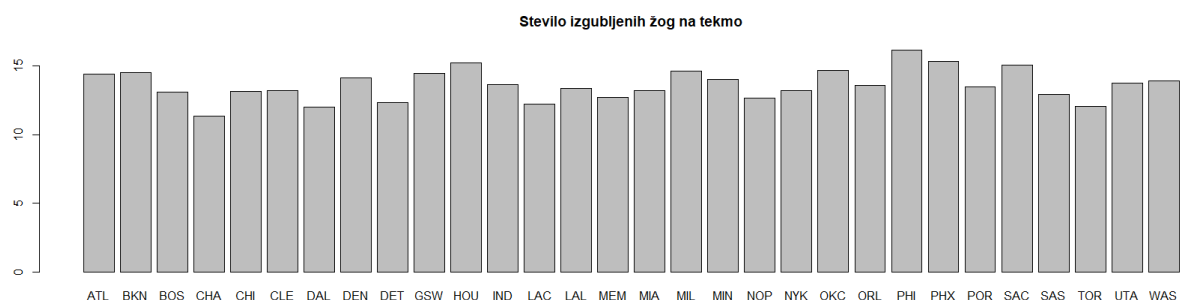
Četrty graf prikazuje ekipe z največ zmagami v teh treh sezonah in sicer GSW, SAS in CLE. Na grafu lahko vidimo, da so GSW imeli veliko zmag in tretje uvrščeni.



Peti graf prikazuje

število izgubljenih žog

na tekmo posamezne ekipe. Z grafa je razvidno, da so najmanj žog izgubljali igralci CHA, največ pa



igralci PHI.

2. Klasifikacija

Naloga za klasifikacijo je bila napovedati verjetnost zmage domačih. Za učenje in testiranje sva zgenerirala množico iz množice tekem. Zgenerirala sva jo tako, da sva za vsako tekmo, namesto podatkov o statistike trenutne tekme, sva izračunala povprečja statistike ekipe, ki je igrala doma in povprečja statistike ekipe, ki je igrala v gosteh. Povprečje pa sva izračunala iz tekem, ki so se že zgodile, do trenutne tekme. Na koncu sva še delila statistiko gostov z statistiko domačih, tako da sva dobila kot neko razmerje med gosti : domači. Naša na novo zgrajena množica je tudi vsebovala atribut KDO_ZMAGA, ki lahko vsebuje samo dve vrednosti (H/A) in atributa ST_ZMAG_DOMA in ST_ZMAG_AWAY, ki povesta koliko tekem je domača ekipa zmagala doma in koliko tekem so gosti zmagali v gosteh. Tako je naša množica vsebovala (KDO_ZMAGA, SAH_PTS, SAH_2PM, SAH_2PA, SAH_3PM,...). In na koncu odstranila tekme kjer ekipe, še nimajo statistike.

Za učenje sva uporabila: Odločitveno drevo z knjižnicama rpart in CORElearn, naivni Bayesov klasifikator z knjižnicami e1071 in CORElearn, K-najbližjih sosedov z knjižnico CORElearn in naključni gozd z knjižnicami randomForest in CORElearn. Velikost učne množice sva spreminjala glede na boljše rezultate, izbirala sva jo med 73% - 76% vseh podatkov, razen pri naključnem gozdu sva izbirala med 60% - 62%. Za vsak model, sva uporabila 3 različne podatke, da vidiva razlike rezultatov glede na podatke.

Rezultati za odločitveno drevo:

	večincki klasifikator * 100	klasifikacijska točnost * 100	Brierova vrednost * 100	napaka zgrešenih razvrstitev * 100
rpart, vsi podatki, 73,8%	57.79	60.71	47.38	37.42
CORElearn, vsi podatki, 73,8%	57.79	60.81	57.92	40.24
rpart, napredna stat. , 75,5%	57.61	60.55	47.67	37.92

CORElearn, napredna stat. 75,5%	57.61	59.88	57	40.02
rpart, St. zmag H/A, 75,3%	57.62	64.41	45.66	37.98
CORElearn, St. zmag H/A 75,3%	57.62	61.86	49.95	41.87

Rezultati za naivni Bayesov klasifikator:

	večincki klasifikator * 100	klasifikacijska točnost * 100	Brierova vrednost * 100	napaka zgrešenih razvrstitev * 100
e1071, vsi podatki, 75,5%	57.61	62.68	56.15	36.83
CORElearn, vsi podatki, 75,5%	57.61	61.34	58.29	41.33
e1071, napredna stat. , 74%	57.75	60.1	46.71	36.43
CORElearn, napredna stat. 74%	57.75	60.31	48.3	40.76
e1071, St. zmag H/A, 76%	57.64	64.19	46.81	36.84
CORElearn, St. zmag H/A 76%	57.64	62.94	46.69	40.84

Rezultati za K-najbližjih sosedov:

	večincki klasifikator * 100	klasifikacijska točnost * 100	Brierova vrednost * 100	napaka zgrešenih razvrstitev * 100
CORElearn, vsi podatki, 74,7%, k=15	57.71	64.82	46.43	38.58
CORElearn, napredna stat, 75%, K=5	57.63	62.32	51.35	38.9
CORElearn, St. zmag H/A , 74%, k=14	57.64	60.26	50.74	38.44

rezultati za naključni gozd:

	večincki klasifikator * 100	klasifikacijska točnost * 100	Brierova vrednost * 100	napaka zgrešenih razvrstitev * 100
randomForest, vsi podatki, 60%	57.85	64.05	44.26	36.8
CORElearn, vsi podatki, 60%	57.85	64.22	44.38	36.85
randomForest, napredna stat. , 62%	58.02	61,31	45.15	36.59
CORElearn, napredna stat. 62%	58.02	62.32	45.44	36.55
randomForest, St. zmag H/A, 61%	57.83	63.55	53.17	36.42
CORElearn, St. zmag H/A 61%	57.83	63.92	50.79	35.92

3. Regresija

Pri regresiji je bila naloga napovedati razliko med točkami domačih in gostov. Za učenje in testiranje sva uporabila množico, ki je vsebovala tako osnovno statistiko (AST, TO, REB, FA, ...) ter napredno statistiko (POS, eFG, OE, ...). Učna množica je vsebovala tekme z sezon 2014-15 in 2015-2016, testna pa tekme sezone 2016-17.

Za učenje sva uporabila: Linearno regresijo, regresijsko drevo, Randomforrest, svm ter CORElearn.

Rezultati:

	Mae	Rmae	Mse	Rmse
Linearna regresija	8.62	0.78	117.36	0.63
Regresijsko drevo	9.20	0.84	134.98	0.72
Random forrest	9.03	0.82	127.95	0.68
Svm	8.72	0.79	120.93	0.64
Core learn	9.39	0.85	140.45	0.75