

Univerza v Ljubljani

Fakulteta za Računalništvo in Informatiko

Večna pot 113, Ljubljana

1. Seminarska raziskovalna naloga 2018-2019 za predmet

UMETNA INTELIGENCA

prof. Igor Kononenko
as. Petar Vračar

Avtorja: Simon Klemenčič, 63170144 in Nejc Kozjek 63170157

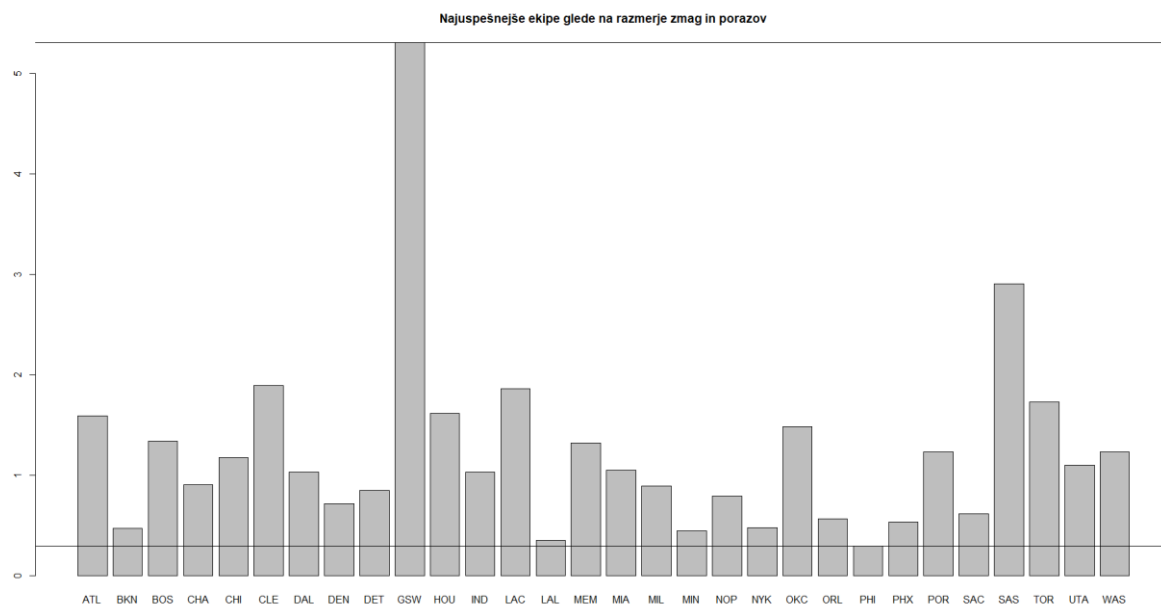
Izbira atributov

Na začetku sva pripravila attribute s katerimi sva pripravila novo množico. Attribute sva pripravila tako, da sva se s zanko sprehodila čez vse tekme. Vsak atribut je izračunan le iz predhodnih tekem. Pripravila sva sledeče attribute.

- povprečno število točk na tekmo
- povprečje zadetih metov za tri točke
- povprečje zadetih metov za dve točki
- povprečje zadetih prostih metov
- povprečje skokov v napadu
- povprečje skokov v obrambi
- povprečje asistenc
- povprečje ukradenih žog
- povprečje izgubljenih žog
- povprečje blokad
- povprečje osebnih napak
- Forma ekipe(število zmag zadnjih pet tekem)
- Razmerje zmage porazi

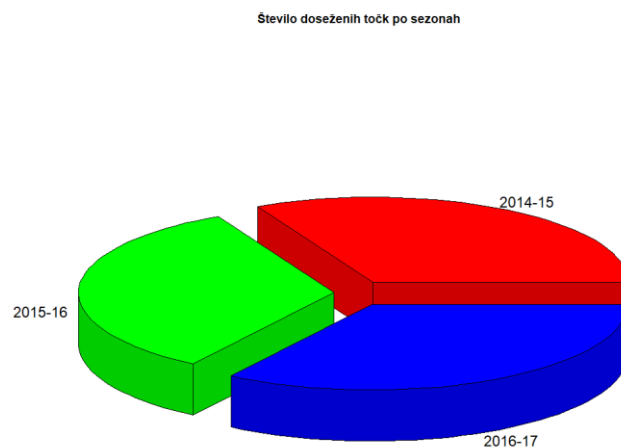
Vizualizacija podatkov

Prvi graf prikazuje najuspešnejše ekipe glede na razmerje zmag in porazov. Najboljša ekipa je Golden State Warriors.



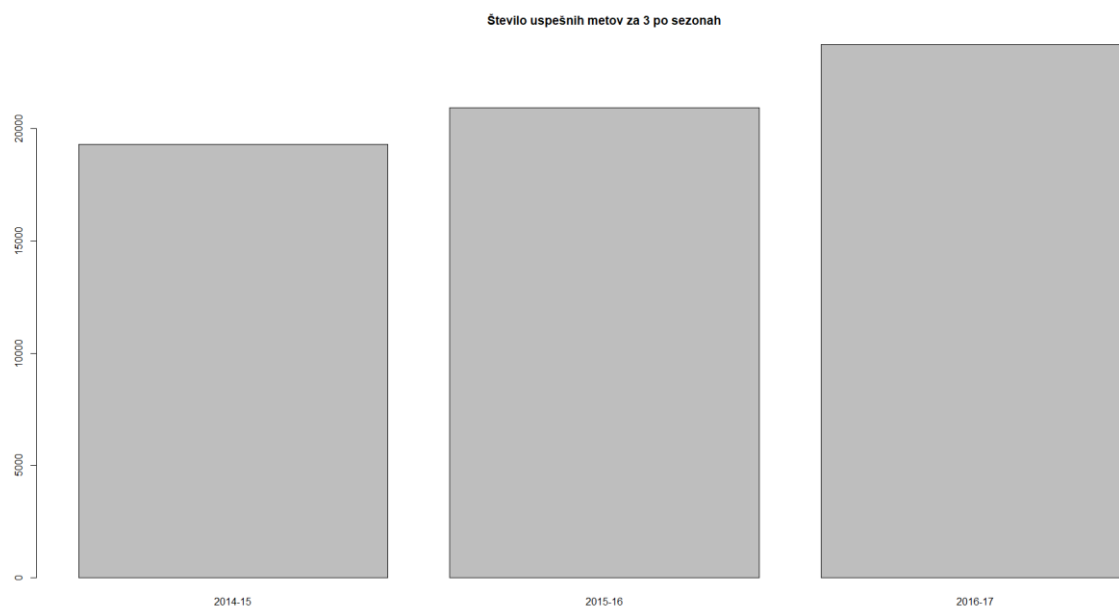
Slika 1: Razmerje zmag in porazov med ekipami

Drugi graf predstavlja število vseh doseženih točk po sezonah. Vidimo da je razmerje med sezonami skoraj isto.



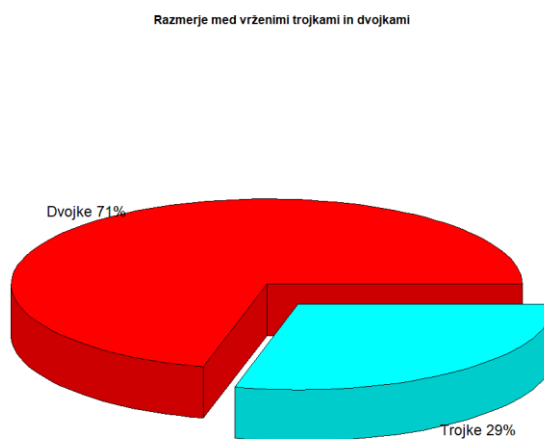
Slika 2: Število doseženih točk po sezonah

Naslednji graf predstavlja število zadetih trojk po sezonah. Zanimivo je, da je število zadetih trojk naraslo.



Slika 3: Število zadetih trojk po sezonah

Zadnji graf predstavlja razmerje med meti za dve ter tri točke.



Slika 4: Razmerje med meti za tri ter dve točki

Klasifikacija

Potrebno je bilo narediti model, ki je napovedoval verjetnost zmage domače ekipe. Za gradnjo modela sva tekme razdelila na tri dele. Prvih 2000 tekem sva uporabila za učno množico, naslednjih 900 tekem sva uporabila za testiranje kvalitete modela zadnjih 900 pa za evalvacijo modela. Preizkusila sva naslednje modele:

- Odločitveno drevo
- Naključni gozd
- K-najbližjih sosedov
- Nevronska mreža

Pri vsakem modelu sva se igrala z različnimi atributi, a sva ugotovila, da nama modeli delujejo najboljše z vsemi atributi.

Dobljena klasifikacijska točnost najboljših štirih je naslednja:

<i>Model</i>	<i>Prvi testni podatki</i>		<i>Končni testni podatki</i>	
	CA	Brierjeva mera	CA	Brierjeva mera
<i>Odločitveno drevo</i>	0,79	0,23	0,76	0,28
<i>Naključni gozd</i>	0,79	0,26	0,76	0,29
<i>K-najbližjih sosedov</i>	0,73	0,36	0,69	0,39
<i>Nevronska mreža</i>	0,72	0,38	0,71	0,35

Nastavitve:

- *Odločitveno drevo*: PRIVZETE
- *Naključni gozd*: ntree=81, proximity=T
- *K-najbližjih sosedov*: klnn = 36
- *Nevronska mreža*: size = 3 , decay = 0.0002, maxit = 10000

Regresija

Potrebno je bilo narediti model, ki napove razliko v točkah posamezne tekme. Tekme sva razdelila po istem postopku kot pri klasifikaciji. Prav tako so modeli z vsemi atributi delovali najbolje. Preizkusila sva naslednje modele:

- Regresijsko drevo
- Naključni gozd
- K-najbližjih sosedov
- Nevronska mreža

Najboljši modeli:

Model	Prvi Testni podatki				Končni testni podatki			
	MAE	RMAE	MSE	RMSE	MAE	RMAE	MSE	RMSE
Regresijsko drevo	6,47	0,57	71,08	0,361	6,16	0,57	62,75	0,354
Naključni gozd	6,5	0,57	71,36	0,36	6,46	0,60	66,66	0,37
K-najbližjih sosedov	6,92	0,61	90,02	0,45	6,82	0,63	84,51	0,47
Nevronska mreža	6,47	0,57	71,13	0,36	6,81	0,63	82,75	0,46

Nastavitve:

- Regresijsko drevo: PRIVZETE
- Naključni gozd: PRIVZETE
- K-najbližjih sosedov: k = 60
- Nevronska mreža: size = 5 , decay = 0.0001, maxit = 10000, linout = T