

Sveučilište Jurja Dobrile u Puli

Fakultet informatike u Puli



Projektni zadatak

Monte Carlo simulacija pobjednika skupina Eura 2020.

Izradio: Luka Đurašinović

JMBAG: 0069064282

Nositelj kolegija: Doc.dr.sc. Darko Etinger

Asistent: Robert Šajina, mag.inf

Pula, lipanj 2021.

Sadržaj:

1. Uvod	2
2. Opis prikupljenih podataka.....	3
3. Konceptualni model	4
4. Izrada računalnog modela i rezultati simulacije.....	4
5. Vizualizacija i tumačenje rezultata	6
6. Literatura	8

1. Uvod

Nogomet je svakako najpopularniji sport u cijelom svijetu. Gledanosti sezonskih finala klubskih natjecanja broje se u stotinama milijuna, a gledanost finala svjetskih reprezentativnih prvenstava najčešće se broji u milijardama.

Upravo takvo jedno veliko natjecanje u nogometu trenutno se odvija u cijeloj Europi, a to je europsko prvenstvo. Ono se odvija u 11 gradova u 11 zemalja diljem Europe, te u njemu sudjeluju 24 najbolje europske reprezentacije koje su plasman na prvenstvo ostvarile kroz prethodne kvalifikacije. Te 24 reprezentacije raspoređene su u 6 grupa sa po 4 reprezentacije u svakoj grupi. U ovom projektu pokušati ću simulirati najvjerojatnijeg pobjednika svake grupe bez pridavanja ikakve prednosti ijednoj reprezentaciji, te ću potom simulirati i potencijalne pobjednike grupa uz pridavanje realnih koeficijenta vjerojatnosti za osvajanje grupe svakoj pojedinoj reprezentaciji unutar iste.

Za potrebe simulacije koristiti će se Monte Carlo metoda koja je najprikladnija za ove vrste predviđanja ishoda, odnosno najčešće je korištena za simuliranje ishoda sportskih događaja ili bilo kakvih ishoda slučajnih događaja ili događaja različitih vjerojatnosti.

Najpoznatiji alati koji računaju Monte Carlo metodom su Geant (Gate), MCNP, Fluka i mnogi drugi.



Slika 1 - Službeni logo Eura 2020.

2. Opis prikupljenih podataka

S obzirom da je prvenstvo trenutno aktualna tema i odvija se upravo sada u realnom vremenu nikakav gotov dataset podataka nisam pronašao, već sam sve podatke sam ručno unosio za potrebe simulacije. S obzirom da skup podataka koji mi je bio potreban i nije bio pretjerano velik isti nisam spremao u dataset, već sam ga za svaku grupu pojedinačno unio u samom programskom kodu. Ipak, nakon obavljenih simulacija podatke s imenima reprezentacija i njihovom rasporedu po grupama ručno sam unio u samostalno kreirani dataset i spremio (exportao) ga kao csv datoteku (Slika 2) kako bih istu mogao priložiti uz dokumentaciju, te koristiti eventualno za neke kasnije simulacije sličnog tipa.

1	country	group
2	Italy	A
3	Wales	A
4	Switzerland	A
5	Turkey	A
6	Belgium	B
7	Russia	B
8	Finland	B
9	Denmark	B
10	Netherlands	C
11	Ukraine	C
12	Austria	C
13	Republic of North Macedonia	C
14	England	D
15	Czech Republic	D
16	Croatia	D
17	Scotland	D
18	Sweden	E
19	Slovakia	E
20	Spain	E
21	Poland	E
22	France,	F
23	Germany	F
24	Portugal	F
25	Hungary	F

Slika 2 - Kreirani CSV dataset

3. Konceptualni model

Na prikazanom konceptualnom modelu (Slika 3) je jednostavni prikaz „etapa“ kreiranja cjelokupne simulacije odnosno projekta.



Slika 3 - Konceptualni model

4. Izrada računalnog modela i rezultati simulacije

Kao što je već rečeno cilj je kreirati simulaciju koja će dati vjerojatnost da neka od reprezentacija osvoji svoju grupu, pritom prvi put pokrećemo simulaciju bez dodjeljivanja prednosti reprezentacijama unutar grupe, a potom svakoj reprezentaciji dodjeljujemo određeni koeficijent tj. vjerojatnost da će upravo ona biti osvajač svoje skupine.

S obzirom da nemamo nikakav dataset iz kojeg možemo izvući podatke, iste unosimo ručno unutar RStudio u kojem i izrađujemo simulaciju (Slika 4).

```
Group_A <- c("Italy", "Wales", "Switzerland", "Turkey")
Group_B <- c("Belgium", "Russia", "Finland", "Denmark")
Group_C <- c("Netherlands", "Ukraine", "Austria", "Republic of North Macedonia")
Group_D <- c("England", "Czech Republic", "Croatia", "Scotland")
Group_E <- c("Sweden", "Slovakia", "Spain", "Poland")
Group_F <- c("France", "Germany", "Portugal", "Hungary")
```

Slika 4 - Izvadak iz koda u kojem kreiramo grupe i unosimo reprezentacije

Nakon kreiranja grupa potrebno je unesenim reprezentacijama dodijeliti i vjerojatnosti za osvajanje njihovih skupina za potrebe kasnije simulacije (Slika 5).

```
vjerojatnostA <- c(6/10, 1/10, 1/10, 2/10)
vjerojatnostB <- c(5/10, 2/10, 1/10, 2/10)
vjerojatnostC <- c(7/10, 1.5/10, 1/10, 0.5/10)
vjerojatnostD <- c(5/10, 1.5/10, 3/10, 0.5/10)
vjerojatnostE <- c(1/10, 0.5/10, 7/10, 1.5/10)
vjerojatnostF <- c(4/10, 3/10, 2/10, 1/10)
```

Slika 5 - Unos vjerojatnosti osvajanja skupine za svaku pojedinu reprezentaciju

Sljedeće što je potrebno su funkcije simulacija. Izrađuju se funkcije simulacija za svaku pojedinu grupu, prva funkcija služi za simuliranje bez dodijeljivanja prednosti (koeficijenta) reprezentacijama, a druga funkcija simulira potencijalne pobjednike skupina prema dodijeljenom koeficijentu vjerojatnosti za osvajanje skupine (Slika 6). Svaka simulacija “vrti se” 1000 puta, te izlaze možemo jednostavno prikazati print funkcijom (Slika 7).

```
plasmanA <- sample(Group_A,
                    replace=TRUE,
                    size= 1000)
plasmanA_vjerojatnost <- sample(Group_A,
                                prob= vjerojatnostA,
                                replace=TRUE,
                                size= 1000)
```

Slika 6 - Prikaz funkcija za simulacije na primjeru grupe A

[6]	"Italy"	"Italy"	"Wales"	"Italy"	"Italy"
[11]	"Switzerland"	"Italy"	"Italy"	"Italy"	"Switzerland"
[16]	"Italy"	"Italy"	"Italy"	"Wales"	"Italy"
[21]	"Italy"	"Turkey"	"Italy"	"Italy"	"Italy"
[26]	"Turkey"	"Wales"	"Italy"	"Turkey"	"Switzerland"
[31]	"Turkey"	"Italy"	"Wales"	"Switzerland"	"Turkey"
[36]	"Turkey"	"Italy"	"Italy"	"Turkey"	"Italy"
[41]	"Italy"	"Turkey"	"Turkey"	"Turkey"	"Turkey"
[46]	"Italy"	"Italy"	"Wales"	"Italy"	"Wales"
[51]	"Turkey"	"Italy"	"Switzerland"	"Turkey"	"Italy"

Slika 7 - Primjer dobivenih “pobjednika” nakon pokretanja simulacije

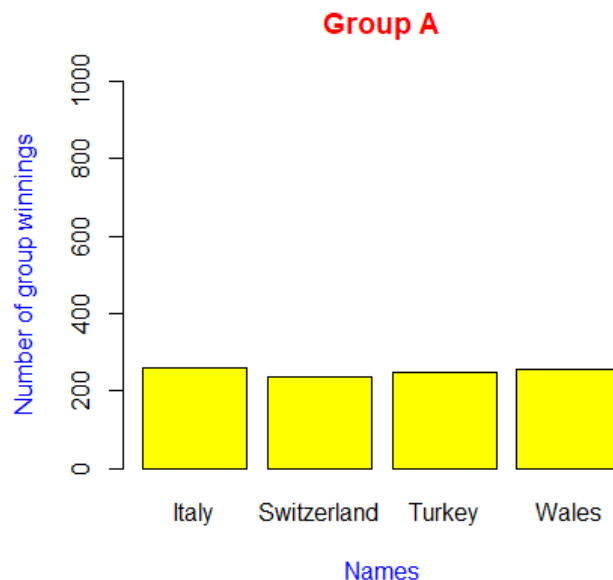
5. Vizualizacija i tumačenje rezultata

Nakon uspješno odrađenih simulacija jedino što je preostalo je vizualizirati dobivene rezultate. Dobivene rezultate najlakše je prikazati putem histograma, pritom nije korištena funkcija “hist()” iz razloga što na X osi nemamo brojčane podatke nego imena reprezentacija (character podatke), stoga je korištena funkcija “barplot()” koja ne zahtjeva korištenje brojčanih podataka na X osi (Slika 8).

```
barplot(table(plasmanA),  
        ylim = c( 0, 1000 ),  
        ylab= "Number of group winnings",  
        xlab= "Names",  
        col = "yellow",  
        col.main = "red",  
        col.lab = "blue",  
        main = "Group A")
```

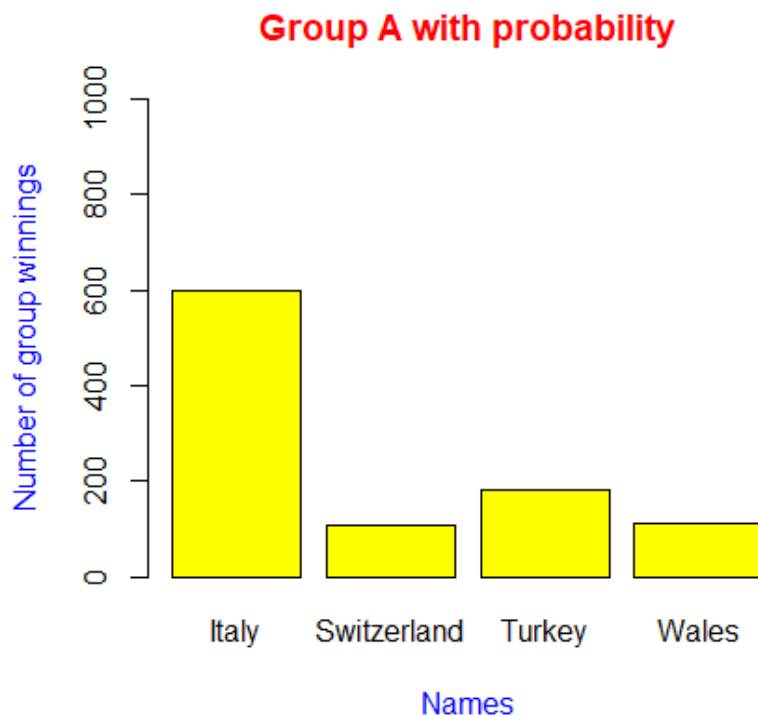
Slika 8 - Primjer koda za izradu grafa(histograma) korištenjem funkcije barplot

Na sljedeća dva histograma vidimo primjere vizualizacije dobivenih ishoda za grupu A. Na prvom grafu očekivano dobivamo podjednake vrijednosti (izglede) za osvajanje grupe svake od 4 reprezentacije (Slika 9) pošto u toj simulaciji nismo davali prednost nijednoj od reprezentacija.



Slika 9 - Vizualizacija podataka simulacije bez davanja prednosti

Na drugom histogramu vidimo vizualizaciju simulacije u kojoj smo dali svakoj reprezentaciji određene izgleda za osvajanje grupe, te smo sukladno tim dodijeljenim vjerojatnostima dobili prikladan graf sa približno očekivanim vrijednostima koje smo prethodno dodijelili (Slika 10).



Slika 10 - Vizualizacija podataka simulacije s davanjem vjerojatnosti

Za potrebe ovog projekta i vizualizacija nisu korišteni rezultati ni trenutna stanja na tablici u stvarnom vremenu, već su vjerojatnosti za plasman unutar skupina kreirane proizvoljno prema favoriziranju reprezentacija od strane stručnjaka i kladionica prije početka samog europskog prvenstva.

6. Literatura

1. Monte Carlo Simulations in R, URL: <https://www.countbayesie.com/blog/2015/3/3/6-amazing-trick-with-montecarlo-simulations>, 2015
2. Introducing Monte Carlo Methods with R, C.P. Robert, G. Casella, 2010.
3. Monte Carlo simulacija, URL: https://hr.wikipedia.org/wiki/Monte_Carlo_simulacija
4. Uefa Euro 2020, Official page, URL: <https://www.uefa.com/uefaeuro-2020/>