MATEMATIKA IZVAN MATEMATIKE

Prognoza ishoda nogometne utakmice

Dušan Munđar[1]

Sažetak. U radu su prikazana dva pristupa prognoziranju ishoda nogometne utakmice: prognoza Poissonovim modelom i prognoza binomnim modelom. Modeli su bazirani na parametrima koji predstavljaju procjenu napadačke i obrambene spremnosti ekipa koje sudjeluju u ogledu te prednost domaćeg terena, odnosno hendikep gostovanja. Procjena parametara dobiva se iz utakmica koje su u nedavnoj prošlosti ekipe imale protiv drugih ekipa iste jakosne skupine. Model je prikazan na rezultatima prve hrvatske nogometne lige.

Ključne riječi – Poissonov model, binomni model, prognoza rezultata, nogomet

1. Uvod

Velik dio populacije zainteresiran je za praćenje sportskih događaja, a naročito mlađi dio. Prognoziranje rezultata utakmice često je dio svakodnevnog razgovora pa želja da budu što bolji u tome može pomoći u motiviranju učenika za učenje matematike, posebice statistike i vjerojatnosti. U obrazovnom procesu tijekom srednje škole uvode se pojmovi koji omogućavaju praćenje tematike prezentirane u radu, stoga mogu biti korišteni za produbljenje gradiva srednje škole u dijelu gradiva o podatcima, statistici i vjerojatnosti.

Prognoziranje rezultata sportskih događaja ima dulju povijest, a u današnje vrijeme postaje sastavni dio novinarskih izvještaja prije većih natjecanja ili važnijih nadmetanja. Iako je zbog velike razine neizvjesnosti nezahvalno prognozirati rezultat pojedine utakmice jednim ishodom, vjerojatnosno prognoziranje, tj. prognoziranje slučajnim varijablama korisno je jer u srednjem roku relativna snaga momčadi postane očita. O temi korištenja slučajnih varijabli za stohastičko (vjerojatnosno) modeliranje rezultata bowling nadmetanja mogli ste čitati u radu Munđar, D. i Šamarija, D. (2015.). O upotrebi metode PageRank za rangiranje ekipa te o potencijalnoj upotrebi za prognoziranje pobjednika u rukometnim utakmicama pisali su Munđar, D. i Horvat, D. (2016). Poissonov model korišten u radu Munđar, D. i Šimić, D. (2016.) te jednostavni binomni model bit će prikazani u ovom radu na način da budu prikladni za korištenje u nastavnom procesu viših razreda srednje škole ili kao

[1]Dušan Munđar, Veleučilište Hrvatsko zagorje Krapina

dopunska (seminarska) tema na predmetima vjerojatnosti i statistike na studijima fakulteta.

Ciljevi koji se namjeravaju ostvariti su sljedeći:

1. Predstaviti dva modela za prognoziranje ishoda nogometne utakmice.
2. Odrediti koeficijente prednosti domaćeg terena te koeficijente napadačke i obrambene spremnosti ekipa u natjecanju potrebne za procjenu parametara koji se koriste u modelima.
3. Primijeniti model za procjenu ishoda utakmice.
4. Odrediti vjerojatnosti izvedenih događaja iz ishoda utakmice.

Modeli kojima ćemo se baviti su Poissonov model i binomni model.

2. Modeli prognoze rezultata

Paralelno s razvojem analize podataka o sportskim događajima razvili su se razni načini prognoziranja rezultata budućih događaja. Prognoziranje Poissonovim modelom i binomnim modelom pripadaju jednostavnijim načinima prognoziranja. Od drugih vrlo je popularna metoda *elo*-rangiranja. Prvenstveno je razvijena za rangiranje i prognoziranje u šahovskim natjecanjima, a sve češće koristi se za rangiranje i u drugim sportovima.

2.1. Poissonov model

Metoda prognoziranja Poissonovim modelom modelira vjerojatnost broja golova pojedine ekipe. Pretpostavljamo da su brojevi golova jedne i druge ekipe nezavisni pa se množenjem vjerojatnosti broja golova za svaku ekipu dobije vjerojatnost ishoda nogometne utakmice.

Neka $(X_{i,j}, Y_{i,j})$ predstavlja slučajni ishod utakmice, pri čemu je $X_{i,j}$ predstavlja broj golova koje je zabila ekipa i koja igra na domaćem terenu protiv ekipe j koja igra u gostima i koja je zabila $Y_{i,j}$ golova. Rezultat se modelira dvjema slučajnim varijablama s Poissonovom distribucijom, pri čemu očekivani broj golova domaće ekipe iznosi λ, a gostujuće ekipe μ.

$$X_{i,j} \sim Poisson(\lambda)$$

$$Y_{i,j} \sim Poisson(\mu)$$

$$\lambda = \lambda_p(1+\gamma)(1+\alpha_i)(1-\beta_j)$$

$$\mu = \lambda_p(1-\gamma)(1+\alpha_j)(1-\beta_i)$$

Parametri λ i μ dobiju se pomoću povijesnih rezultata utakmica koje su te ekipe odigrale protiv ekipa sličnih snaga. Parametri α_i, β_i predstavljaju spremnost napada i obrane momčadi i te analogno za ekipu j. Parametar λ_p predstavlja očekivani broj golova za ekipu u jednoj utakmici, a parametar γ iskazuje prednost ekipe koja igra na domaćem terenu, odnosno hendikep ekipe koja igra na gostujućem terenu. Više o izračunu spomenutih parametara bit će u cjelini Procjena parametara.

Vjerojatnost ishoda $(X_{i,j}, Y_{i,j}) = (m, n)$ tada možemo dobiti formulom

$$P((X_{i,j}, Y_{i,j}) = (m,n)) = P(X_{i,j} = m) \cdot P(Y_{i,j} = n) = \frac{\lambda^m}{m!} \mathrm{e}^{-\lambda} \frac{\mu^n}{n!} \mathrm{e}^{-\mu} = \frac{\lambda^m \mu^n}{m!n!} \mathrm{e}^{-\lambda-\mu}.$$

2.2. Geometrijski model

Drugi pristup koji ćemo prikazati pristup je modeliranjem raspodjele golova na utakmici, uz poznatu distribuciju golova na utakmici. Neka je Z slučajna varijabla ukupnog broja golova na utakmici. Broj Z tada moramo podijeliti na dvije skupine, prvu skupinu golova – golove koje je postigla domaća ekipa, i drugu skupinu – golove koje je postigla gostujuća ekipa. Vjerojatnost da je ekipa domaćin postigla gol uzima u obzir napadačke i obrambene sposobnosti domaće i gostujuće ekipe te prednost domaćeg terena. Navedene aspekte ekipa u ogledu, kao i prednost domaćeg terena, modeliramo parametrima prikazanima u Poissonovom modelu, a njihov izračun, kao što je prethodno navedeno, bit će prikazan u cjelini Procjena parametara. Vjerojatnost da pojedini gol, iz skupa ukupnog broja golova na utakmici, postigne prva (domaća) ekipa određujemo formulom

$$p = \frac{\lambda}{\lambda+\mu} = \frac{(1+\gamma)(1+\alpha_i)(1-\beta_j)}{(1+\gamma)(1+\alpha_i)(1-\beta_j)+(1-\gamma)(1+\alpha_j)(1-\beta_i)}.$$

Slučajne varijable broja golova pojedine ekipe, ukoliko je ukupno palo n golova, tada iznose:

$$X_{i,j} \mid Z \sim Binom(n, p), \qquad Y_{i,j} = Z - X_{i,j}.$$

Dakle, uvjetna vjerojatnost ishoda s ukupno n golova na utakmici iznosi

$$P(X_{i,j} = k, Y_{i,j} = n-k \mid Z = n) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Vjerojatnost ishoda $(X_{i,j}, Y_{i,j}) = (m,n)$ tada možemo dobiti množenjem uvjetne vjerojatnosti s vjerojatnosti $p_n = P(Z = n)$ koju ćemo izračunati empirijski na temelju povijesnih rezultata utakmica:

$$P\Big((X_{i,j}, Y_{i,j}) = (k, n-k)\Big) = P(X_{i,j} = k, Y_{i,j} = n-k \mid Z = n) \cdot P(Z = n) = \Big(\quad \Big)\ {}^{k}(1-p)^{n\ \ k} \cdot p_n$$

Za primjer ćemo izračunati vjerojatnosti pojedinih rezultata u slučaju da su na utakmici postignuta tri gola ($Z = 3$). Pretpostavimo da vjerojatnost da su u utakmici postignuta tri gola iznosi 0.3, a da su snage momčadi $\lambda = 1.8$ i $\mu = 1.2$. Vjerojatnost da pojedini gol zabije prva (domaća) ekipa tada iznosi $p = 1.8/(1.8 + 1.2) = 0.6$, a vjerojatnost da pojedini gol zabije druga (gostujuća) ekipa iznosi $1 - 0.6 = 0.4$.

Ishod	Uvjetna vjerojatnost ishoda (uvjet: ukupan broj golova je 3)	Vjerojatnost konkretnog ishoda
	$P(X_{i,j}=k, Y_{i,j}=3-k \mid Z=3)$	$P(X_{i,j}=k, Y_{i,j}=3-k \mid Z=3) \cdot P(Z=3)$
3:0	$\binom{3}{3} p^3(1-p)^0 = 0.6^3 \cdot 0.4^0 = 0.216$	0.0864
2:1	$\binom{3}{2} p^2(1-p)^1 = 0.6^2 \cdot 0.4^1 = 0.432$	0.1728
1:2	$\binom{3}{1} p^1(1-p)^2 = 0.6^1 \cdot 0.4^2 = 0.288$	0.1152
0:3	$\binom{3}{0} p^0(1-p)^3 = 0.6^0 \cdot 0.4^3 = 0.064$	0.0256

Tablica 1. Prikaz vjerojatnosti pojedinog ishoda uz uvjet da su na utakmici postignuta tri gola

2.3. Procjena parametara modela

Kako bismo mogli koristiti predložene modele, potrebno je izračunati parametre korištene u modelima. U nastavku opisujemo način izračuna parametara i provodimo izračun na skupu konkretnih podataka.

Prosječan broj golova po ekipi i po nogometnoj utakmici λ_p jednak je zbroju svih golova domaćih ekipa X i golova gostujućih ekipa Y podijeljen brojem $2n$, gdje je n broj odigranih utakmica.

$$\lambda_p = \frac{\sum x + \sum y}{2 \cdot n}$$

Snaga napada ekipe dobije se kao omjer prosječnog broja golova koje je ekipa postigla po utakmici i prosječnog broja golova po ekipi i utakmici umanjen za 1.

$$\alpha_i = \frac{\overline{x_i}}{\lambda_p} - 1$$

Parametar pokazuje za koliko je posto ekipa uspješnija ili slabija od prosjeka ako se promatra snaga napada. Primjerice, rezultat *0.2*, odnosno 20 %, označava da ekipa u prosjeku postiže 20 % više golova od prosječnog iznosa za sve ekipe.

Snaga obrane momčadi β_i dobije se kao omjer prosječnog broja golova po ekipi i prosječnog broja golova koje ekipa primi po utakmici umanjen za 1.

$$\beta_i = \frac{\lambda_p}{\overline{y_i}} - 1$$

Parametar pokazuje za koliko je posto ekipa uspješnija ili slabija od prosjeka ako se promatra snaga obrane. Primjerice, rezultat 0.2, odnosno 20 %, označava da je prosječan iznos golova dobivenih po ekipi po utakmici u prosjeku 20 % veći od broja golova koje dobiva analizirana ekipa.

Prednost domaćeg terena - pretpostavljamo da je ista za sve ekipe i jednaka je omjeru prosječnog broja golova koje postiže domaća ekipa naspram prosječnog broja golova po ekipi po utakmici.

$$\gamma = \frac{\overline{h}}{\lambda_p} - 1$$

Empirijska distribucija golova po utakmici računa se kao relativni broj utakmica s k golova u ukupnom broj odigranih utakmica n. Da bismo izbjegli ekstremne situacije, uzimat ćemo u obzir samo utakmice u kojima je palo najviše sedam golova.

$$p_k = \frac{f_k}{n}$$

3. Izračun prognoze ishoda utakmice

3.1. Procjena parametara za modele

Koristit će se povijesni podatci o rezultatima odigranih dvadeset kola HNL sezone 2022./2023., odnosno za sve utakmice odigrane od početka sezone do početka veljače 2023. godine. U natjecanju sudjeluje deset ekipa pa to znači da je svaka ekipa trebala odigrati 20 utakmica. No, budući da je jedna utakmica odgođena (Dinamo – Istra 1961) i nije odigrana do trenutka pisanja rada, promatrat će se ukupno 99 utakmica. Prikaz ekipa u natjecanju, sažeti rezultati prvih 20 kola te broj golova u analiziranim utakmicama predstavljen je u Tablici 2. Nakon prikupljanja povijesnih podataka određeni su parametri potrebni za modele.

Ekipa	Broj pobjeda/ neodlučenih/ izgubljenih/ odigranih utakmica	Broj postignutih pogodaka	Broj primljenih pogodaka
Dinamo	14/ 4/ 1/ 19	51	19
Gorica	1/ 6/ 13/ 20	11	29
Hajduk	11/ 5/ 4/ 20	36	25
Istra 1961	7/ 6/ 6/ 19	20	19
Lokomotiva	6/ 4/ 10/ 20	25	31
Osijek	10/ 4/ 6/ 20	32	23
Rijeka	6/ 4/ 10/ 20	22	30
Slaven Belupo	7/ 7/ 6/ 20	17	28
Šibenik	2/ 10/ 8/ 20	15	23
Varaždin	7/ 6/ 7/ 20	24	26

Tablica 2. Prikaz ekipa u natjecanju, sažeti rezultati, broj postignutih i primljenih golova u prvih 20 kola HNL-a u sezoni 2022./23. (izvor podataka: https://www.rezultati.com/, obrada: autor)

Prosječan broj golova po ekipi i nogometnoj utakmici omjer je ukupnog broja golova u odigranih 99 utakmica, koji je 253, i broja 2 · 99 = 198.

$$\lambda_p = \frac{253}{2 \cdot 99} = 1.28$$

Parametri spremnosti napada i obrane ekipa prikazani su u Tablici 3.

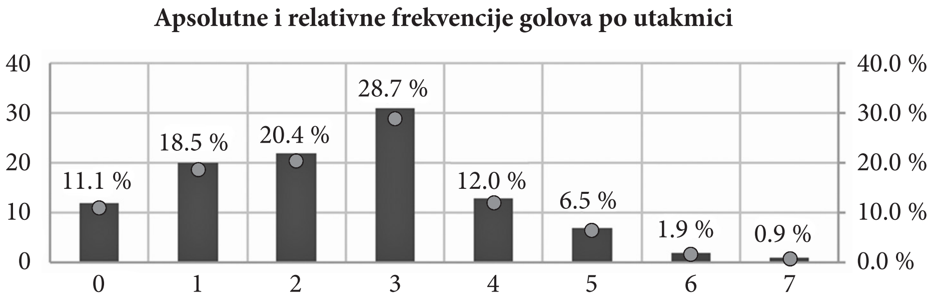
Ekipa	**Parametar spremnosti napada α_i**	**Parametar spremnosti obrane β_i**
Dinamo	1.101	-0.217
Gorica	-0.570	0.135
Hajduk	0.409	-0.022
Istra 1961	-0.176	-0.217
Lokomotiva	-0.022	0.213
Osijek	0.252	-0.100
Rijeka	-0.139	0.174
Slaven Belupo	-0.335	0.096
Šibenik	-0.413	-0.100
Varaždin	-0.061	0.017

Tablica 3. Prikaz parametara spremnosti napada i obrane ekipa u natjecanju

Prednost domaćeg terena, a pretpostavljamo da je ista za sve ekipe i jednaka je omjeru prosječnog broja golova koje postigne domaća ekipa naspram prosječnog broja golova po ekipi i po utakmici, u ovom slučaju iznosi $\gamma = 0.059$.

Empirijska distribucija broja golova Z po utakmici[2] iznosi:

$$Z \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0.111 & 0.185 & 0.204 & 0.287 & 0.120 & 0.065 & 0.019 & 0.009 \end{pmatrix}.$$



Grafikon 1. Apsolutne i relativne frekvencije broja golova na nogometnim utakmicama u prvih 20 kola HNL-a u sezoni 2022./23. (izvor podataka: https://www.rezultati.com/, obrada: autor)

[2]Iz analize je izuzeta situacija s devet postignutih golova jer se može smatrati neobičajenim događajem.

3.2. Prognoza rezultata Poissonovim modelom

Napravimo izračun prognoze rezultata za proizvoljne dvije ekipe iz analizirane lige. Neka za nas to budu *Osijek* i *Rijeka,* pri čemu se utakmica igra u Osijeku. Budući da smo izračunali parametre snage napada i obrane odabranih ekipa, možemo za ekipe izračunati očekivani broj golova. Za Osijek očekivani broj golova iznosi 1.40, a za Rijeku 1.14. U tablici su prikazane vjerojatnosti pojedinih ishoda utakmice. Tri najvjerojatnija ishoda bila su 1:1 (12.6 %), 1:0 (11.1 %) i 0:1 (9.0 %). 21. siječnja 2023. odigrana je utakmica u Osijeku koja je završila 1:1, što je prema modelu bio najvjerojatniji ishod.

		Rijeka				
		0	1	2	3	4
Osijek	0	7.9 %	9.0 %	5.1 %	1.9 %	0.6 %
	1	11.1 %	12.6 %	7.2 %	2.7 %	0.8 %
	2	7.7 %	8.8 %	5.0 %	1.9 %	0.5 %
	3	3.6 %	4.1 %	2.3 %	0.9 %	0.3 %
	4	1.3 %	1.4 %	0.8 %	0.3 %	0.1 %

Tablica 4. Prikaz vjerojatnosti ishoda utakmice Osijeka i Rijeke dobivenih Poissonovim modelom

3.3. Prognoza rezultata binomnim modelom

Napravimo sada izračun prognoze rezultata za iste dvije ekipe, ali koristeći binomni model. Vjerojatnost da pojedini gol postigne domaća ekipa, u ovom slučaju Osijek, iznosi $p = \lambda/(\lambda + \gamma) = 0.552$.

Vjerojatnosti ishoda utakmice izračunate korištenjem formula prikazanih u opisu binomnog modela uz korištenje empirijske distribucije prikazane su u Tablici 5. Tri najvjerojatnija ishoda bila su 2:1 (12.1 %), 0:0 (11.2 %) i 1:1 (11.1 %). Kao što smo prethodno naveli, rezultat je bio 1:1 i u ovom slučaju taj rezultat bio je treći najvjerojatniji.

		Rijeka				
		0	1	2	3	4
Osijek	0	11.2 %	6.9 %	4.5 %	2.7 %	0.5 %
	1	8.4 %	11.1 %	9.8 %	2.2 %	0.8 %
	2	6.8 %	12.1 %	4.1 %	2.0 %	0.4 %
	3	5.0 %	3.4 %	2.4 %	0.6 %	0.2 %
	4	1.0 %	1.5 %	0.6 %	0.3 %	0.0 %

Tablica 5. Prikaz vjerojatnosti ishoda utakmice Osijeka i Rijeke dobivenih binomnim modelom

3.4. Prognoza izvedenih događaja

U nekim situacijama nećemo bazirati svoje prognoze na temelju najvjerojatnijeg događaja nego na temelju skupine događaja. Primjerice, može nas zanimati vjerojatnost pobjede domaćina p_1, vjerojatnost izjednačenog rezultata p_0 ili pobjede gostujuće ekipe p_2. Do tih vjerojatnosti može se doći sumiranjem vjerojatnosti koje pripadaju opisanom događaju. Tako su:

- vjerojatnost pobjede domaćina
$$p_1 = P(X_{ij} > Y_{ij}) = \sum_{(x_{ij}, y_{ij}) | x_{ij} > y_{ij}} p(x_{ij}, y_{ij}),$$
- vjerojatnost neodlučenog rezultata
$$p_0 = P(X_{ij} = Y_{ij}) = \sum_{(x_{ij}, y_{ij}) | x_{ij} = y_{ij}} p(x_{ij}, y_{ij}),$$
- vjerojatnost pobjede gosta
$$p_2 = P(X_{ij} < Y_{ij}) = \sum_{(x_{ij}, y_{ij}) | x_{ij} < y_{ij}} p(x_{ij}, y_{ij}).$$

U slučaju našeg para *Osijek – Rijeka*, pri čemu je Osijek domaćin, navedene vjerojatnosti dobivene Poissonovim modelom iznose $(p_1, p_0, p_2) = (0.424, 0.270, 0.306)$, dok vjerojatnosti ishoda dobivene binomnim iznose (p_1, p_0, p_2)=(0.421, 0.275, 0.304. Na sličan način mogu se dobiti i vjerojatnosti drugih izvedenih događaja, primjerice vjerojatnost da je ekipa pobijedila s dva ili više golova razlike.

4. Zaključak

Prognoziranje rezultata sportskih događaja za mnoge je zanimljiva tema. U ovom smo radu prikazali dvije relativno jednostavne metode prognoziranja rezultata nogometne utakmice. Slične metode mogu se koristiti i za druge sportove sličnog sustava bodovanja, kao što su rukomet ili vaterpolo. Metode smo prikazali za dvije ekipe iz hrvatske nogometne lige, a parametre smo izračunali iz prethodnih dvadesetak utakmica koje su te ekipe odigrale prije spomenute utakmice prognoziranja. Prognoza rezultata pogodna je za prikaz u nastavnom procesu u sklopu gradiva matematike kako bi se učenicima viših razreda srednjih škola približilo gradivo statistike i vjerojatnosti. Također, omogućuju širenje analize procjene vjerojatnosti izvedenih događaja te testiranja uspješnosti prognoziranja s istim ili drugačijim načinom definiranja parametara modela.

Literatura:

1. Munđar, D., i Horvat, D. (2016.). *Rangiranje ekipa i prognoziranje ishoda u rukometu korištenjem PageRank algoritma*. Poučak: časopis za metodiku i nastavu matematike, 17(67), 35-42. https://hrcak.srce.hr/clanak/266675

2. Munđar, D., & Šamarija, D. (2015.). *Primjena simulacijskih modela za prognoziranje rezultata natjecanja u bowlingu*. Poučak: časopis za metodiku i nastavu matematike, 16(64), 73-79. https://hrcak.srce.hr/clanak/234056

3. Munđar, D. i Šimić, D. (2016.). *Croatian First Football League: Teams' performance in the championship*. Croatian Review of Economic, Business and Social Statistics, 2 (1), 15-23. https://doi.org/10.1515/crebss-2016-0006