

# Primjena Dixon -Coles modela u predviđanju ishoda nogometnih utakmica

---

**Kampić, Karlo**

**Undergraduate thesis / Završni rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **Josip Juraj Strossmayer University of Osijek, School of Applied Mathematics and Informatics / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet primijenjene matematike i informatike**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:126:669458>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2025-03-10**



**mathos**

*Repository / Repozitorij:*

[Repository of School of Applied Mathematics and Informatics](#)



DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJ



SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

FAKULTET PRIMIJENJENE MATEMATIKE I INFORMATIKE

Studij

Sveučilišni prijediplomski studij Matematika i računarstvo

# **Primjena Dixon-Coles modela u predviđanju ishoda nogometnih utakmica**

ZAVRŠNI RAD

Mentor:

**izv. prof. dr. sc.  
Danijel Grahovac**

Student:

**Karlo Kampić**

Osijek, 2024



# Sadržaj

<b>1</b>	<b>Uvod</b>	<b>1</b>
<b>2</b>	<b>Dixon-Coles model</b>	<b>3</b>
2.1	Pregled literature . . . . .	3
2.2	Specifikacije modela . . . . .	4
<b>3</b>	<b>Primjena modela</b>	<b>7</b>
3.1	Prikupljanje podataka . . . . .	7
3.2	Implementacija . . . . .	8
3.3	Brier Score i Log Loss . . . . .	11
<b>4</b>	<b>Osvrt</b>	<b>13</b>
<b>5</b>	<b>Zaključak</b>	<b>15</b>
	<b>Literatura</b>	<b>17</b>
<b>6</b>	<b>Sažetak</b>	<b>19</b>
	<b>Sažetak</b>	<b>19</b>
	<b>Summary</b>	<b>21</b>
	<b>Životopis</b>	<b>23</b>



# 1 | Uvod

Sportsko klađenje razvilo se u globalnu industriju s milijunima aktivnih kladitelja koji sudjeluju i ulažu svoje oklade na svim dostupnim platformama za klađenje. Paralelno s time, pojavila se potreba za razvojem točnog i sigurnog modela koji bi bio sposoban predvidjeti ishode bilo koje utakmice. Tradicionalni modeli temeljeni na Poissonovoj distribuciji stekli su veliku popularnost zbog svoje jednostavnosti i učinkovitosti u modeliranju distribucije golova. Međutim, mnogi od tih modela loše obuhvaćaju intrinzične složenosti povezane s nogometnim utakmicama, uključujući, ali ne ograničavajući se na međuzavisnost između učinkovitosti momčadi i činjenicu da su rezultati s malim brojem golova česti.

Dixon-Coles model, koji su razvili Dixon i Coles 1997. godine, osmišljen je kako bi se prevladali ovi nedostaci prilagođavanjem standardne Poissonove distribucije tako da uzme u obzir složenosti u dinamikama igre tijekom nogometnih utakmica. U ovom radu pokušat će se primijeniti Dixon-Coles model na sezonu Engleske Premier lige 2023. godine, usporediti njegovu izvedbu sa jednostavnijim modelima te raspraviti njegove stvarne primjene u sportskom klađenju i upravljanju nogometnim momčadima.

Ovo istraživanje motivirano je potrebom za složenijim modelima koji bi mogli poboljšati prediktivnu točnost modela za predviđanje ishoda nogometnih utakmica. Korisnost takvih modela mogla bi biti od velikog interesa ne samo za sportske kladitelje, već i za nogometne analitičare, trenere i menadžere koji se moraju osloniti na takve modele kako bi donijeli informirane odluke u vezi s taktikom momčadi i pripremama za utakmicu. Stoga, ovaj rad nastoji usporediti Dixon-Coles model s tradicionalnim metodama, naglašavajući prednosti primjene naprednih statističkih tehnika u nogometnoj analitici.



## 2 | Dixon-Coles model

### 2.1 Pregled literature

Priča o primjeni statističkih modela u sportu, posebno u nogometu, proteže se još od početka 20. stoljeća. Prvi i jedan od najutjecajnijih bio je rad [5], koji je koristio Poissonovu distribuciju za predviđanje ishoda nogometnih utakmica. Jedna od najvažnijih značajki Maherovog modela bila je njegova jednostavnost, stoga je vrlo snažna osnova koju je ponudio poslužila kao temelj za daljnja istraživanja u sportskoj analitici. Međutim, model je imao niz ograničenja, posebno u pretpostavci neovisnosti u procesima postizanja golova od strane obiju suparničkih momčadi.

Dixon-Coles model predložen je 1997. godine kao model koji uključuje mjere za nadoknađivanje ovog nedostatka, dopuštajući ovisnost između momčadi i činjenicu da nogometne utakmice često karakterizira nizak broj postignutih golova. Posljedično, ovo je prošireni tradicionalni Poissonov model s dodatkom odgovarajućeg termina za prilagodbu vjerojatnosti niskih rezultata, čime je model postao korisniji za stvarno modeliranje rezultata nogometnih utakmica.

Od svog nastanka, model je postao dio nogometne analitike. Njegova su ograničenja prevladana u studijama poput [4], koje su proširile model uvođenjem bivarijantnih Poissonovih distribucija sposobnih uhvatiti ovisnost između rezultata postizanja golova momčadi. Takve dodatke znatno su doprinijele širem prihvaćanju modela, održavajući ravnotežu između složenosti i jednostavnosti, što ga čini prikladnim za znanstvena istraživanja i praktične primjene.

Dixon-Coles model široko je primijenjen, ne samo u akademskom kontekstu, već i u sektoru klađenja. To što model može relativno precizno predvidjeti ishode utakmica učinilo ga je važnim alatom za kladionice i kladitelje. Ovaj aspekt modela, s vremenom, nastavio se razvijati i prilagođavati, uključujući dodatke poput dodatnih varijabli kao što su forma momčadi, ozljede igrača i okolišni čimbenici poput vremenskih uvjeta.

Ostatak rada raspravlja o ovoj literaturi primjenjujući Dixon-Coles model na suvremene nogometne podatke, u ovom slučaju, sezonu Engleske Premier Lige 2023. godine i uspoređujući njegovu robusnost s jednostavnijim modelima, s ciljem demonstriranja prednosti sofisticiranijih statističkih metodologija u nogometnoj analitici. Također se iznose praktične implikacije koje bi takvi modeli mogli imati za primjenu u sportskom klađenju i upravljanju momčadima.



## 2.2 Specifikacije modela

Dixon-Coles model proširuje standardnu Poissonovu distribuciju, koja se često koristi za modeliranje broja događaja koji se rijetko događaju u fiksnom vremenskom razdoblju, poput broja golova postignutih u nogometnoj utakmici. Poissonova distribucija se primjenjuje zbog nekoliko ključnih pretpostavki: rijetkost događaja, neovisnost između događaja, te konstantna stopa događanja kroz vrijeme.

Distribucija slučajne varijable  $X$  koja ima Poissonovu distribuciju s parametrom  $\lambda > 0$  dana je s:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

U okviru modeliranja nogometne utakmice:

- $P(X = k)$  predstavlja vjerojatnost da će se postići  $k$  golova
- $\lambda$  predstavlja očekivani broj golova po utakmici
- $k$  predstavlja broj postignutih golova

Međutim, nogometne utakmice često ne zadovoljavaju u potpunosti pretpostavku neovisnosti golova, jer akcije jedne momčadi izravno utječu na prilike druge momčadi. Također, rezultate s malim brojem golova standardni Poissonov model teško predviđa.

Dixon-Coles model uvodi korekcije na Poissonov model kako bi bolje odražavao međuzavisnost između momčadi i učestalost niskih rezultata. Model zadržava osnovnu strukturu Poissonove distribucije za predviđanje broja golova domaće i gostujuće momčadi, ali uvodi korekcijski faktor kako bi poboljšao preciznost predviđanja.

Vjerojatnosti u Dixon-Coles modelu je dane su s:

$$P(X = x, Y = y) = \frac{\lambda_X^x e^{-\lambda_X}}{x!} \cdot \frac{\lambda_Y^y e^{-\lambda_Y}}{y!} \cdot \tau_{X,Y}(x, y),$$

gdje:

- $P(X = x, Y = y)$  predstavlja vjerojatnost da će domaća momčad postići  $x$  golova, a gostujuća momčad  $y$  golova
- $\lambda_X$  i  $\lambda_Y$  predstavljaju očekivani broj golova domaće i gostujuće momčadi
- $\tau_{X,Y}$  predstavlja korekcijski faktor za niske rezultate i međusobnu zavisnost između momčadi

Model uvodi prilagodbu funkcije vjerojatnosti na temelju ishoda utakmica, posebno za utakmice s niskim brojem golova.

Korekcijski faktor definira se kao:

$$\tau_{XY}(x, y) = \begin{cases} 1 - \rho & \text{ako je } x = y = 0 \\ 1 + \lambda_X \rho & \text{ako je } x = 0, y = 1 \\ 1 + \lambda_Y \rho & \text{ako je } x = 1, y = 0 \\ 1 - \rho & \text{ako je } x = y = 1 \\ 1 & \text{inače} \end{cases}$$

Model uzima u obzir činjenicu da su utakmice s niskim brojem golova često rezultat određenih i namjernih taktičkih odluka, poput "*parkiranja autobusa*", ultra-defenzivne igre i slično, te da nisu isključivo rezultat sreće. To omogućuje Dixon-Coles modelu da preciznije podešava svoja predviđanja ishoda utakmica u uvjetima u kojima bi standardni Poissonov model bio izrazito nestabilan.

Očekivani broj golova za domaću momčad  $X$  i gostujuću momčad  $Y$  u modelu se predviđa pomoću sljedećih formula:

$$\lambda_X = \exp(\alpha + \beta_h + \gamma_a + \delta)$$

$$\lambda_Y = \exp(\alpha + \beta_a + \gamma_h)$$

U ovim formulama:

- $\lambda_X$  i  $\lambda_Y$  su očekivani broj golova za domaću ( $X$ ) i gostujuću momčad ( $Y$ )
- $\alpha$  predstavlja osnovnu stopu postizanja golova u cijeloj ligi. Ovaj parametar obuhvaća opću sklonost postizanju golova u utakmicama i služi kao početna točka za model.
- $\beta_h$  i  $\beta_a$  predstavljaju napadačke snage domaće i gostujuće momčadi. Ovi parametri odražavaju ofenzivne sposobnosti svake momčadi i njihovu sposobnost postizanja golova.
- $\gamma_h$  i  $\gamma_a$  predstavljaju obrambene snage domaće i gostujuće momčadi. Ovi parametri odražavaju defanzivne sposobnosti svake momčadi i njihovu sposobnost sprječavanja suparničke momčadi da postigne golove.
- $\delta$  je korekcijski faktor za utakmice s niskim brojem golova, koje su u nogometu česte i često ih je teško precizno predvidjeti.

Kombinirajući i prilagođavajući ove varijable, Dixon-Coles model može pružiti precizniji uvid u složenost nogometnih utakmica. Dakle, u usporedbi s jednostavnijim modelima, ovaj model je snažniji u predviđanju ishoda utakmica.



## 3 | Primjena modela

### 3.1 Prikupljanje podataka

Podaci su prikupljeni s Football-Data.co.uk, najcjelovitije arhivske baze podataka o nogometu. Fokus ovog istraživanja bio je na analizi Engleske Premier lige u 2023. godini, za koju je pružena iscrpna lista rezultata utakmica.

Prikupljeni podaci su organizirani u *Excel* tablicu.

HomeTeam	AwayTeam	FTHG	FTAG
Burnley	Man City	0	3
Arsenal	Nott'm Forest	2	1
Bournemouth	West Ham	1	1
Brighton	Luton	4	1
Everton	Full Ham	0	1

Tablica 3.1: Dio korištene tablice.

Sljedeće su ključne varijable koje su procijenjene iz ovog skupa podataka:

- HomeTeam: Naziv domaće momčadi.
- AwayTeam: Oznaka gostujuće momčadi.
- FTHG - Full-Time Home Goals: Ukupan broj golova koje je domaća momčad postigla tijekom utakmice.
- FTAG - Full-Time Away Goals: Broj golova koje je gostujuća momčad postigla u regularnom vremenu.

Jedina svrha ovih odabranih podataka je predstavljanje jednostavnih varijabli potrebnih u Dixon-Colesovom modelu za predviđanje broja golova koje bi momčad mogla postići. Podaci poput ovih korišteni u ovom istraživanju morali su biti pažljivo očišćeni kako bi se izbjegle pogreške ili nekonzistentnosti, jer bi one najvjerojatnije utjecale na točnost rezultata.

Ovaj rad se fokusira na Englesku Premier Ligu zbog njegove konkurentne prirode i rezultirajuće varijabilnosti u ishodima utakmica što ju čini savršenim slučajem za testiranje učinkovitosti prediktivnih modela poput Dixon-Colesovog. Nadalje, Engleska Premier Liga je jedna od najgledanijih nogometnih liga na svijetu, što ovo istraživanje čini korisnim za mnoge kategorije čitatelja, uključujući sportske kladitelje, nogometne analitičare i navijače.

## 3.2 Implementacija

Implementacija Dixon-Coles modela provedena je u **Python-u**, programskom jeziku koji je široko korišten za analizu podataka i statističko modeliranje. Korištene su različite biblioteke, uključujući **numpy** za numeričke izračune, **pandas** za manipulaciju podacima i **scipy** za optimizaciju.

Implementacijski proces se može podijeliti u sljedeće dijelove:

**Priprema podataka:** Podaci su očišćeni i strukturirani kako bi se osigurala točnost i konzistentnost. To je uključivalo uklanjanje eventualnih nedostajućih vrijednosti i pogrešnih unosa, te preformatiranje podataka za analizu i modeliranje.

**Podjela podataka:** Kako bismo izbjegli korištenje istih podataka za procjenu parametara i testiranje modela, podaci su podijeljeni na trening skupove. Trening set koristi se za procjenu parametara modela, dok se test set koristi za provjeru performansi modela na neviđenim podacima.

Ova podjela omogućava modelu da "uči" iz trening seta, dok se simulacija i evaluacija modela provode na nepoznatim podacima iz test seta. To sprječava preveliku prilagodbu modela podacima na kojima je treniran i daje realniju procjenu prediktivne točnosti modela.

**Procjena parametara:** Dixon-Coles model koristi metodologiju koja uključuje procjenu očekivanog broja golova za obje momčadi, domaću i gostujuću. Očekivani broj golova nije stvaran broj postignutih golova, već statistička procjena koja odražava vjerojatnost koliko će golova određena momčad postići, uzimajući u obzir snagu napada, obrane i ostale relevantne čimbenike.

Procjena očekivanog broja golova temelji se na sljedećim parametrima:

- $\lambda_x$ : Očekivani broj golova domaće momčadi
- $\lambda_y$ : Očekivani broj golova gostujuće momčadi

Parametri koji se koriste za procjenu očekivanog broja golova uključuju:

- $\alpha$ : Osnovna stopa postizanja golova u cijeloj ligi
- $\beta_H$ : Napadačka snaga domaće momčadi

- $\beta_a$ : Napadačka snaga gostujuće momčadi
- $\gamma_h$ : Obrambena snaga domaće momčadi
- $\gamma_a$ : Obrambena snaga gostujuće momčadi
- $\delta$ : Korekcija za niske rezultate

Nakon podjele podataka, parametri se procjenjuju pomoću trening seta koristeći metodu maksimalne vjerojatnosti (Maximum Likelihood Estimation - MLE). Ovaj postupak traži optimalne vrijednosti parametara koji najbolje objašnjavaju zabilježene rezultate utakmica u trening setu.

**Simulacija rezultata** Nakon što su procijenjeni parametri modela, simuliraju se ishodi utakmica u test setu. Simulacija koristi očekivani broj golova za svaku momčad. Ovi očekivani brojevi golova, koji su procijenjeni na temelju trening seta, simuliraju rezultate u test setu te se uspoređuju s stvarnim brojem postignutih golova u tim utakmicama.

Simulirani rezultati pomažu nam procijeniti koliko se točno model može generalizirati na nepoznate podatke. Očekivani broj golova služi kao statistička procjena koja daje informaciju o tome koliko bi momčad trebala postići golova, dok stvaran broj golova koji se uspoređuje s tim predikcijama predstavlja stvarne ishode utakmica u test setu.

**Procjena performansi** Nakon što su simulirani rezultati uspoređeni sa stvarnim rezultatima u test setu, procjenjuju se performanse modela. Kako bismo točno izmjerili uspješnost predikcija modela, koriste se metrički pokazatelji kao što su Brier Score i Log Loss, koji omogućuju detaljnu analizu preciznosti probabilističkih predikcija modela.

Ispod je prikazan Python kod korišten za implementaciju koraka simulacije i procjene izvedbe:

```
1 import numpy as np
2 import pandas as pd
3 from scipy.optimize import minimize
4 from sklearn.model_selection import train_test_split
5 from sklearn.metrics import brier_score_loss, log_loss
6
7 # Ucitavanje podataka
8 data = pd.read_csv('EO.csv')
9 data = data[['HomeTeam', 'AwayTeam', 'FTHG', 'FTAG']]
10
11 # Funkcija za izracun vjerojatnosti po Dixon-Coles modelu
12 def dixon_coles_likelihood(params, data):
13     alpha, beta_H, beta_A, gamma_H, gamma_A, delta = params
14     lambda_X = np.exp(alpha + beta_H + gamma_A + delta)
15     lambda_Y = np.exp(alpha + beta_A + gamma_H)
16
```

```
17     likelihood = 0
18     for i, row in data.iterrows():
19         X_i = row['FTHG']
20         Y_i = row['FTAG']
21         p_X_Y = (lambda_X**X_i * np.exp(-lambda_X) / np.math.
factorial(X_i)) * \
22             (lambda_Y**Y_i * np.exp(-lambda_Y) / np.math.
factorial(Y_i))
23         likelihood += np.log(p_X_Y)
24
25     return -likelihood
26
27 # Podjela podataka na trening i test set (80% trening, 20% test)
28 train_data, test_data = train_test_split(data, test_size=0.2,
random_state=42)
29
30 # Procjena parametara na trening setu
31 initial_params = [0.1, 0.1, 0.1, 0.1, 0.1, 0.1]
32 result = minimize(dixon_coles_likelihood, initial_params, args=(
train_data,), method='Nelder-Mead')
33
34 # Optimalni parametri modela
35 alpha_opt, beta_H_opt, beta_A_opt, gamma_H_opt, gamma_A_opt,
delta_opt = result.x
36
37 # Funkcija za simulaciju rezultata
38 def simulate_match_outcomes(params, row):
39     alpha, beta_H, beta_A, gamma_H, gamma_A, delta = params
40     lambda_X = np.exp(alpha + beta_H + gamma_A + delta)
41     lambda_Y = np.exp(alpha + beta_A + gamma_H)
42
43     predicted_home_goals = np.random.poisson(lambda_X)
44     predicted_away_goals = np.random.poisson(lambda_Y)
45
46     return predicted_home_goals, predicted_away_goals
47
48 # Simulacija rezultata na test setu
49 predictions = []
50 for i, row in test_data.iterrows():
51     predicted_home_goals, predicted_away_goals =
simulate_match_outcomes(result.x, row)
52     predictions.append((predicted_home_goals, predicted_away_goals)
)
53
54 # Dodavanje predikcija u test set
55 test_data['Predicted_HomeGoals'] = [p[0] for p in predictions]
56 test_data['Predicted_AwayGoals'] = [p[1] for p in predictions]
57
58 # Brier Score i Log Loss
59 brier_scores = []
60 log_losses = []
61
62 for i, row in test_data.iterrows():
63     # Brier Score: Usporedba predvidene vjerojatnosti pobjede
domace momcadi s stvarnim ishodom
64     home_prob = 1 if row['Predicted_HomeGoals'] > row['
```

```

    Predicted_AwayGoals'] else 0
65     actual_home_win = 1 if row['FTHG'] > row['FTAG'] else 0
66     brier_scores.append(brier_score_loss([actual_home_win], [
    home_prob]))
67
68     # Log Loss: Usporedba stvarnog i predviđenog broja golova
69     predicted_prob = np.exp(-row['Predicted_HomeGoals']) * np.exp(-
    row['Predicted_AwayGoals'])
70     actual_result = [row['FTHG'], row['FTAG']]
71     log_losses.append(log_loss([actual_result], [[row['
    Predicted_HomeGoals'], row['Predicted_AwayGoals']])))
72
73 # Izracun prosječnog Brier Score-a i Log Loss-a
74 average_brier_score = np.mean(brier_scores)
75 average_log_loss = np.mean(log_losses)
76
77 print(f"Average Brier Score: {average_brier_score}")
78 print(f"Average Log Loss: {average_log_loss}")
79
80 # Provjera uspjeha simulacije (usporedba sa stvarnim rezultatima)
81 print(test_data[['HomeTeam', 'AwayTeam', 'FTHG', 'FTAG', '
    Predicted_HomeGoals', 'Predicted_AwayGoals']].head())

```

Sljedeći kod se koristio za prikupljanje prosječnog broja golova:

```

1 import pandas as pd
2
3 data = pd.read_csv('E0.csv')
4
5 average_goals_per_match = (data['FTHG'] + data['FTAG']).mean()
6
7 average_home_goals = data['FTHG'].mean()
8
9 average_away_goals = data['FTAG'].mean()
10
11 print(f"Avg golovi: {average_goals_per_match:.2f}")
12 print(f"Avg home: {average_home_goals:.2f}")
13 print(f"Avg away: {average_away_goals:.2f}")

```

### 3.3 Brier Score i Log Loss

Kvalitetu predviđanja možemo mjeriti s različitim metričkim pokazateljima, od kojih su najpoznatiji **Brier Score** [1] i **Log Loss** [3]. Ove mjere omogućuju evaluaciju vjerojatnosnih predviđanja, te pomažu u kvantificiranju koliko su predviđanja modela točna i koliko su blizu stvarnim ishodima.























# Application of the Dixon-Coles model in predicting football match outcomes

## Summary

This paper analyzes the application of the Dixon-Coles model for predicting football match outcomes, with a particular focus on the 2023 English Premier League season. The Dixon-Coles model introduces adjustments to the traditional Poisson regression model to account for team interactions and the frequent occurrence of low-scoring matches, which are often inadequately addressed by conventional models. The collected data is analyzed and modeled using the maximum likelihood estimation method, while the model's effectiveness is evaluated using metrics such as the Brier Score and Log Loss.

The results demonstrate that the Dixon-Coles model provides more accurate predictions compared to simpler models, especially in matches with a low number of goals. This advantage makes the model practically applicable in sports betting and football analytics, where prediction accuracy can significantly contribute to strategic decision-making.

## Keywords

Dixon-Coles model, Poisson model, Brier Score, Log Loss, Python



# Životopis

Rođen sam 2000. godine u Vinkovcima. Pohađao sam Osnovnu školu "August Cesarec" u Ivankovu, nakon koje sam upisao Gimnaziju "Matija Antun Reljković" u Vinkovcima, matematički smjer. Godine 2019. upisujem sveučilišni prije-diplomski studij Matematike i računarstva Odjela za matematiku (sada Fakultet primjenjene matematike i informatike).