

Pemodelan ZIP pada Data Jumlah Gol Tandang Pertandingan Liga 1 Indonesia Tahun 2012-2019

Geovani Silva Ibrahim*, **Aceng Komarudin Mutaqin**

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*geovaniLaquica@gmail.com, aceng.k.mutaqin@gmail.com

Abstract. In the game of football, away team matches are matches against the home team in a league. This study aims to determine whether the Zero Inflated Poisson distribution is suitable for modeling the number of away team goals in an Indonesian League 1. The ZIP distribution is modeled on the data on the number of away team goals scored by each team for 5 seasons of the year and tested for application. The data used is the result of the match results of the Indonesian League 1 away team competition for the 2012-2019 season. The frequency and percentage of away team goals scored are used to describe away team goals scored statistics. The maximum likelihood estimator method is used to estimate the parameters in the discrete distribution. Meanwhile, the chi-square test was used to test the distribution fit. The results of the application show that at the level of significance =1%, the ZIP distribution is suitable for modeling data on the number of away team goals in all seasons of the year. While at the level of significance =5%, the ZIP distribution is suitable for modeling data on the number of away team goals for 4 seasons of the year.

Keywords: Zero inflated Poisson distribution, chi-kuadrat.

Abstrak. Dalam permainan sepak bola, pertandingan tim tandang merupakan pertandingan yang melawan tim tuan rumah dalam suatu Liga. Penelitian ini bertujuan untuk mengetahui apakah distribusi *Zero Inflated Poisson* cocok untuk memodelkan jumlah gol tim tandang dalam suatu Liga 1 Indonesia. Distribusi ZIP dimodelkan kepada data jumlah gol tim tandang yang dicetak setiap tim selama 5 musim tahun dan diuji penerapannya. Data yang digunakan adalah data hasil pertandingan tim tandang kompetisi Liga 1 Indonesia musim tahun 2012-2019. Frekuensi dan persentase dari jumlah gol tim tandang digunakan untuk menggambarkan statistik jumlah gol tim tandang. Metode penaksir kemungkinan maksimum digunakan untuk menaksir parameter pada distribusi diskrit tersebut. Sedangkan uji chi-kuadrat digunakan untuk menguji kecocokan distribusi. Hasil penerapannya menunjukkan bahwa pada taraf nyata $\alpha = 1\%$, distribusi ZIP cocok untuk memodelkan data jumlah gol tim tandang pada semua musim tahun. Sedangkan pada taraf nyata $\alpha = 5\%$, distribusi ZIP cocok untuk memodelkan data jumlah gol tim tandang untuk 4 musim tahun dari kelima musim tahun

Kata Kunci: Distribusi *Zero Inflated Poisson*, chi-kuadrat.

1. Pendahuluan

Pengaplikasian distribusi pada kasus sepak bola sudah banyak dilakukan oleh peneliti-peneliti dari berbagai negara. Para peneliti tidak hanya melihat pertandingan dan hasil suatu pertandingan saja, tetapi melakukan penelitian tentang distribusi jumlah gol dalam suatu pertandingan sepak bola. Dengan hal ini para peneliti bertujuan untuk mengetahui distribusi peluang apa yang cocok untuk memodelkan tim dalam suatu Liga. Distribusi diskrit banyak digunakan secara luas sebagai pendekatan model sederhana untuk distribusi jumlah gol tim tandang. Moroney (1956) menggunakan distribusi Poisson untuk memodelkan jumlah gol di

Liga Negara Korea. Taek (2014) menggunakan distribusi Poisson dan ekstrim untuk memodelkan jumlah gol musim tahun 1983-2012 di Liga Korea.

Berdasarkan latar belakang, maka perumusan masalah dalam penelitian ini sebagai berikut: “bagaimana hasil pemodelan distribusi ZIP pada data jumlah gol tim tandang Liga 1 sepak bola Indonesia untuk musim tahun 2012-2019?” selanjutnya, tujuan yang ingin dicapai dalam penulisan penelitian ini adalah untuk mengetahui hasil pemodelan distribusi ZIP pada data jumlah gol tim tandang Liga 1 sepak bola Indonesia untuk musim tahun 2012-2019.

2. Metodologi

Distribusi Zero Inflated Poisson (ZIP)

Peubah acak diskrit X dikatakan berdistribusi ZIP dengan parameter $0 \leq \tau \leq 1$ dan $\lambda \geq 0$ apabila fungsi peluangnya sebagai berikut:

$$f(x; \tau, \lambda) = \begin{cases} \tau + (1 - \tau)e^{-\lambda}, & x = 0 \\ (1 - \tau) \frac{e^{-\lambda} \lambda^x}{x!}, & x = 1, 2, \dots \end{cases}$$

(1)

Ekspektasi dan varians dari distribusi ZIP adalah

$$E(X) = \lambda(1 - \tau)$$

$$V(X) = \lambda(1 - \tau)(1 + \lambda\tau)$$

untuk kasus data *overdispersi* dan *underdispersi* yang data nya disebabkan frekuensi nol terlalu banyak dapat dimodelkan dengan distribusi ZIP (Widjajati, Saputri, & Asiyah, 2015).

Misalkan X_1, X_2, \dots, X_n adalah suatu sampel acak berukuran n dari distribusi ZIP dengan parameter τ dan λ , dengan nilai sampel acak x_1, x_2, \dots, x_n . Penaksir kemungkinan maksimum untuk τ dan λ dari distribusi ZIP adalah solusi dari dua persamaan berikut (Beckett dkk.2014):

$$\frac{n\bar{x}}{\lambda} = \frac{n_0(1-\tau)e^{-\lambda}}{\tau+(1-\tau)e^{-\lambda}} + n - n_0,$$

(2)

$$n - n_0 = \frac{n_0(1-\tau)(1-e^{-\lambda})}{\tau+(1-\tau)e^{-\lambda}}$$

(3)

dimana n_0 adalah banyaknya data yang bernilai 0.

Uji Kecocokan Distribusi Chi-Kuadrat

Uji kecocokan distribusi adalah suatu pengujian hipotesis statistik yang digunakan untuk mengetahui apakah x_1, x_2, \dots, x_n adalah nilai dari sampel acak X_1, X_2, \dots, X_n yang berdistribusi (\cdot). Uji kecocokan distribusi dapat digunakan untuk menguji hipotesis berikut:

H_0 : x_1, x_2, \dots, x_n . merupakan nilai dari sampel acak yang berdistribusi dengan fungsi distribusi $F(\cdot)$

H_1 : x_1, x_2, \dots, x_n . Merupakan bukan nilai dari sampel acak yang berdistribusi dengan fungsi distribusi $F(\cdot)$

Salah satu uji kecocokan distribusi adalah uji chi-kuadrat. Statistik uji untuk uji chi-kuadrat adalah

$$\chi^2 = \sum_{x=1}^m \frac{(n_x - np_x)^2}{np_x} \quad (9)$$

dimana n_x adalah banyaknya pengamatan untuk kategori x , n adalah total banyaknya pengamatan, p_x adalah peluang untuk kategori x , dan m adalah banyaknya kategori. Statistik uji di atas berdistribusi chi-kuadrat dengan derajat bebas $- r - 1$, dimana r menyatakan banyaknya parameter yang ditaksir dari distribusi. Pada taraf nyata α , tolak hipotesis nol jika nilai statistik uji yang ada pada Persamaan (9) lebih besar dari nilai kuantil dari distribusi chi-kuadrat dengan derajat bebas $m - r - 1$ atau tolak hipotesis nol jika $\chi^2 \geq \chi_{(m-r-1)(1-\alpha)}^2$. Jika ada kategori yang nilai harapannya (np_x) kurang dari 5, maka kategori-kategori tersebut digabungkan untuk menghasilkan kategori baru yang mempunyai nilai np_x lebih besar sama dengan 5. Hal ini bertujuan untuk menghindari penerimaan hipotesis nol yang bias.

3. Pembahasan dan Diskusi

Data yang digunakan adalah data sekunder hasil pencatatan yang diperoleh dari beberapa sumber yaitu *rsssf.com*, *Flashscore.com*, *id.Soccerway.com*, *Wikipedia.com* mengenai hasil-hasil pertandingan Liga 1 Indonesia dari mulai tahun 2012-2019. Data yang digunakan tersebut berisi informasi mengenai hasil skor akhir setiap pertandingan tim tandang yang dipertandingkan pada musim tahun 2012-2019. Distribusi yang diuji dikatakan cocok apabila nilai *p-value* nya lebih besar dari nilai taraf nyata, α

Tabel 1. Nilai Frekuensi Untuk Data Jumlah Gol Tim Tandang Musim Tahun 2012-2019

Tahun	Jumlah Gol Tandang										
	0	1	2	3	4	5	6	7	8	9	10
2012-2013	103	113	60	22	5	2	1	0	0	0	0
2013-2014	93	86	45	9	7	3	0	0	1	0	0
2016-2017	107	110	59	20	7	3	0	0	0	0	0
2017-2018	116	103	59	17	9	2	0	0	0	0	0
2018-2019	100	109	69	22	5	1	0	0	0	0	0

Tabel 2. Nilai Persentase Untuk Data Jumlah Gol Tim Tandang Musim Tahun 2012-2019

Tahun	Jumlah Gol Tandang	
	0	Lebih Dari 0
2012-2013	103 (33,66%)	203 (66,34%)
2013-2014	93 (38,11%)	151 (61,89%)
2016-2017	107 (34,97%)	199 (65,03%)
2017-2018	116 (37,91%)	190 (62,09%)
2018-2019	100 (32,68%)	206 (67,32%)

Tabel 3. Nilai *P-value* Uji Kecocokan Distribusi ZIP Untuk Data Jumlah Gol Tim Tandang Musim Tahun 2012-2019

Tahun	<i>P-value</i>
	ZIP
2012-2013	0.9659
2013-2014	0.0206
2016-2017	0.6575
2017-2018	0.3178
2018-2019	0.4860

4. Kesimpulan

Berdasarkan hasil analisis pemodelan *Zero Inflated Poisson* pada data jumlah gol tim tandang hasil pertandingan sepak bola Liga Indonesia musim tahun 2012-2019 dapat disimpulkan bahwa:

1. Secara Keseluruhan distribusi ZIP cocok untuk memodelkan data jumlah gol tim tandang.
2. Terbukti dalam tabel 4.1 dan 4.2 untuk semua musim tahun Liga 1 Indonesia, secara keseluruhan nilai frekuensi jumlah gol 0 tim tandang lebih dari 90 dan secara keseluruhan nilai persentase lebih dari 30% . Kondisi di atas sangat realistis, mengingat tim tandang selalu kesulitan dalam mencetak gol ke tim kandang. Hal ini menyebabkan data jumlah gol tim tandang banyak bernilai 0.
3. Pada taraf nyata $\alpha = 1\%$, distribusi ZIP cocok untuk memodelkan data jumlah gol tim tandang pada semua musim tahun. Sedangkan pada taraf nyata $\alpha = 5\%$, distribusi ZIP cocok untuk memodelkan data jumlah gol tim tandang untuk 4 musim tahun dari kelima musim tahun.

Acknowledge

With a very deep sense of gratitude and my gratitude to Allah SWT for His grace and guidance alhamdulillah, I was able to complete this research on time and at the right time. I dedicate this thesis to two great people in my life, namely my parents. Don't forget to also say thank you and my love to my mentor, girlfriends and all good people involved who always provide help, prayers and support to the author. I will forever be so grateful to have you all in my life.

Daftar Pustaka

- [1] Moroney, M.J. (1956). Facts From Figures. Pelican Books Inc, Harmondsworth Middlesex, USA.
- [2] Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A., Pal, N. (2014). Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. *Involve a Journal of Mathematics*, Vol. 7, No. 6, 751-767.
- [3] Taek, Y.J. (2014), Distribution Goal in K-League Soccer, *Journal of the Korean, Data & Information Science Society* 2014, 25(6), 1231-1239.
- [4] Widjajati, A.F, Saputri, D.M, Asiyah, N. (2015). Sifat-sifat Generalisasi Distribusi Binomial yang Bertipe COM-Poisson. *J. math. and Its Appl*, Vol. 12, No. 1, 13-22
- [5] Utama Muhammad Bangkit Riksa, Hajarisman Nusar. (2021). *Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg*. *Jurnal Riset Statistika*, 1(1), 35-42.