

***Zero-inflated poisson regression* untuk memodelkan faktor-faktor yang mempengaruhi terjadinya kebakaran di kabupaten Sidoarjo**

Febri Rahmat Dona¹, Hendro Permadi²

Universitas Negeri Malang

E-mail : febry.rahmat@ymail.com

Abstract : The purpose of this research is determine the factors that affect the rate of fires in Sidoarjo using analysis Zero-Inflated Poisson (ZIP) regression. The procedure in this research through the steps are detecting the distribution of a variable number of fire (Y), identified cases multikolinearitas, establishing models Poisson regression, testing overdispersi or underdispersi, establishing models ZIP regression, testing coefficient simultaneously and partially, and choose a suitable model. The independent variable used are population density (X_1), the number of factories (X_2), sugarcane and paddy land area (X_3), vacant land that overgrown with weeds (X_4), and year (X_5) Based on the regression model Zero-Inflated Poisson (ZIP) obtained best model of rate of fires that affected by the population density (X_1), sugarcane and paddy land area (X_3), and the year (X_5) to distinguish the difference the incidence of fires in every year. The model has a value of AIC (257,4) and the smallest value of R^2 (70,86).

Keywords: Poisson Regression, Zero-Inflated Poisson Regression, Fire.

Kabupaten Sidoarjo termasuk kabupaten yang memiliki potensi tinggi terjadinya kasus kebakaran dari tahun ke tahun. Berdasarkan data yang dimiliki oleh Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Sidoarjo, pada 2006 ada 120 kejadian, 2007 ada 110 kejadian, 2008 ada 112 kejadian, 2009 ada 137 kejadian, 2010 ada 73 kejadian dan jumlah kejadian kebakaran selama tahun 2011 ini telah mencapai tingkat mengkhawatirkan yaitu 167 kejadian. Oleh karena itu, perlu adanya penanganan yang serius guna mengurangi terjadinya kebakaran yang meningkat dengan menganalisis faktor-faktor yang mempengaruhi.

Keterkaitan faktor-faktor penyebab terjadinya jumlah kebakaran yang disebabkan beberapa faktor dapat didekati dengan analisis statistika. Dalam terapan ilmu statistika terdapat suatu metode mengenai hubungan variabel terikat dengan variabel bebas yaitu analisis regresi (Agresti, 2002). Analisis regresi umumnya menggunakan variabel tak bebas yang merupakan peubah acak kontinu berdistribusi normal, tetapi terkadang hubungan variabel terikat merupakan peubah acak diskrit dengan mengikuti Distribusi Poisson. Apabila terdapat variabel terikat yang akan diamati merupakan peubah acak diskrit yang berdistribusi Poisson, maka hubungan antara variabel terikat dan variabel bebas dapat dianalisis dengan regresi Poisson (Myers, 1990).

Model regresi Poisson merupakan salah satu model yang digunakan untuk memodelkan hubungan antara variabel terikat berupa data diskrit dengan variabel bebas dapat berupa data campuran (kontinu dan diskrit). Model regresi Poisson terdapat beberapa asumsi. Salah satu asumsi yang harus dipenuhi adalah variansi dari variabel terikat (Y) sama dengan rata-ratanya (*equidispersion*). Namun pada praktiknya, sering ditemukan data diskrit yang memiliki varian lebih besar dari rata-rata (*overdispersi*) atau varian lebih kecil dari rata-rata (*underdispersi*). Regresi Poisson yang diterapkan pada data yang mengandung *overdispersi* akan mempengaruhi nilai *standard error* yang menjadi turun (Hilbe, 2011). Untuk mengatasi hal tersebut salah satu metode yang digunakan metode regresi *Zero-Inflated Poisson (ZIP)* (Cameron dan Trivedi, 1998).

Menurut Hossain dan Howlader, (2015) distribusi ZIP mempunyai keadaan yaitu keadaan nol dengan probabilitas ω yang hanya menghasilkan data bernilai nol dan keadaan Poisson dengan probabilitas $(1 - \omega)$ yang berdistribusi Poisson dengan rata-rata λ . Dengan dua proses tersebut memberikan distribusi campuran dengan dua komponen yang disebut distribusi ZIP dengan fungsi kepadatan peluang (fkp) yaitu:

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)e^{-\lambda}, & y = 0 \\ (1 - \omega) e^{-\lambda} \left(\frac{\lambda^y}{y!}\right), & y \neq 0 \end{cases}$$

dimana $0 \leq \omega \leq 1$ dan $\lambda \geq 0$. Distribusi ZIP dituliskan dengan $ZIP(\omega, \lambda)$ dengan parameter ω dan λ dengan rata-rata dan variannya adalah $E(Y) = (1 - \omega)\lambda$ dan $Var(Y) = \lambda(1 - \omega)(1 + \lambda\omega)$. Lambert (1992) menyarankan hubungan model untuk λ dan ω adalah sebagai berikut:

$$\log(\lambda) = \mathbf{X}\boldsymbol{\beta} \text{ dan } \text{logit}(\omega) = \log\left(\frac{\omega}{1-\omega}\right) = \mathbf{X}\boldsymbol{\gamma}$$

Sehingga dapat ditulis sebagai persamaan berikut:

$$\lambda_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \text{ dan } \omega_i = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\gamma})}$$

Tujuan penelitian ini adalah untuk mengetahui model regresi ZIP dalam menentukan hubungan antara variabel terikat dengan variabel bebas pada kasus angka kebakaran di Kabupaten Sidoarjo tahun 2013-2014 serta faktor-faktor yang dominan.

METODE

Data yang digunakan dalam penelitian ini adalah data sekunder. Data diperoleh dari Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Sidoarjo berupa data kejadian kebakaran yang terjadi di Kabupaten Sidoarjo pada tahun 2013 dan 2014, serta data yang berasal dari Badan Pusat Statistik (BPS) Sidoarjo yang berupa data kepadatan penduduk, jumlah industri, luas lahan tebu dan sawah, luas daerah tidak ditanami yang ada di setiap wilayah kecamatan yang berada di Kabupaten Sidoarjo. Variabel terikat yang diamatai adalah angka kebakaran (Y) dan variabel bebasnya yaitu kepadatan penduduk (X_1), jumlah industri (X_2), luas lahan tebu dan sawah (X_3), luas lahan kosong yang ditumbuhi ilalang (X_4), tahun (X_5) dengan kode 1 untuk kejadian di tahun 2014 dan kode 0 untuk kejadian di tahun 2013. Dalam penelitian ini, langkah-langkah yang dilakukan adalah:

- a) Menguji angka kebakaran apakah mengikuti distribusi poisson atau tidak dengan uji *Kolmogorov-Smirnov*
- b) Menguji apakah terjadi multikolinieritas atau tidak dengan melihat nilai korelasi *spearman rank-order correlation* antar variabel bebas dan variabel bebas dengan variabel terikat.
- c) Melakukan pendugaan parameter regresi Poison
- d) Menguji apakah *equaldispersi*?
 - 1) Jika ya maka regresi Poison telah sesuai.
 - 2) Jika tidak maka harus di deteksi apakah terjadi *overdispersi* atau *underdispersi*. Selanjutnya itu dimodelkan dengan model regresi ZIP.
- e) Pengujian koefisien regresi yang pertama yaitu apakah model signifikan secara seluruh?
 - 1) Jika ya maka dilakukan pengujian koefisien regresi secara parsial (Uji *Wald*),
 - 2) Jika tidak maka model yang dibuat tidak didukung oleh data sehingga penganalisisan selesai.
- f) Pengujian koefisien regresi secara parsial (Uji *Wald*) apakah nilai koefisien parameternya signifikan?
 - 1) Jika ya maka dilakukan pemilihan model terbaik.

- 2) Jika tidak maka menguji model regresi ZIP yang lain.
- g) Pemilihan model terbaik dengan melihat nilai AIC terkecil dan R^2 yang cukup besar.
- h) Selesai.

HASIL DAN PEMBAHASAN

1. Uji Kesesuaian Distribusi

Hasil uji kesesuaian distribusi untuk variabel angka kebakaran dapat dibuat Tabel 1 sebagai berikut:

Tabel 1 Hasil Uji Kesesuaian Distribusi

Distribusi	Kolmogorov Smirnov			Daerah Kritis (0,05)	Keputusan
	Statistik	Urutan	P_{value}		
Uniform	0,19841	1	0,10204	0,22119	Terima H_0
Geometri	0,20041	2	0,09616	0,22119	Terima H_0
Neg. Binomial	0,3181	4	9,7213E-4	0,22119	Terima H_0
Poison	0,21962	3	0,05265	0,22119	Terima H_0

Berdasarkan pada Tabel 4.1 variabel angka kebakaran mengikuti Distribusi Poison berada di ranking 4 dengan nilai P_{value} (0,05265) > dari nilai $\alpha(0,05)$ dan nilai statistiknya (0,21962) < nilai kritisnya (0,22119) sehingga terima H_0 yang berarti variabel angka kebakaran mengikuti Distribusi Poison.

2. Pemeriksaan Multikolinearitas

Pemeriksaan uji multikolinearitas dengan melihat matrik korelasinya seperti pada Tabel 2 berikut:

Tabel 2 Matrik Korelasi

	X_1	X_2	X_3	X_4	X_5
X_2	0,743 0,000				
X_3	-0,450 0,006	-0,614 0,000			
X_4	0,240 0,158	0,453 0,316	-0,172 0,006		
X_5	0,070 0,687	0,075 0,816	-0,070 0,687	0,040 0,664	
Y	-0,289 0,088	0,004 0,983	0,082 0,634	-0,100 0,561	0,054 0,756

Berdasarkan Tabel 2 nilai korelasi antara X_1 dan X_2 cukup tinggi yaitu 0,743 sedangkan nilai korelasi antara X_2 dan Y sangatlah rendah yaitu 0,004. Hal ini bisa menyebabkan terjadinya kasus multikolinearitas sehingga harus ada penghapusan variabel bebas. Variabel bebas yang memungkinkan untuk dihapus yaitu variabel jumlah pabrik (X_2) karena nilai korelasinya dengan Y sangat kecil sehingga jika masuk dalam model hanya berpengaruh sangat sedikit. Variabel kepadatan penduduk (X_1), luas lahan tebu dan sawah (X_3), luas lahan kosong yang ditumbuhi ilalang (X_4) dan tahun (X_5) dapat digunakan untuk memodelkan regresi ZIP.

3. Pemodelan Regresi Poisson dan Pemeriksaan *Overdispersion* atau *Underdispersion*

Hubungan antara angka kebakaran (Y) dengan kepadatan penduduk (X_1), luas lahan tebu dan sawah (X_3), luas lahan kosong yang ditumbuhi ilalang (X_4) dan tahun (X_5) dapat dibuat regresi Poisson seperti Tabel 3 berikut:

Tabel 3 Hasil Pemodelan Regresi Poisson

Parameter	Koefisien	SE Koefisien	P_{value}
β_0	3,083	0,251	0,000
β_1	-0,000215	0,000044	0,000
β_3	-0,02811	0,00826	0,001
β_4	-0,0174	0,0200	0,380
β_5	0,119	0,129	0,355
	<i>Pearson χ^2</i>	160,30	
	Db	31	
	\hat{u}	5,171	

Berdasarkan Tabel 3 dapat dibuat model regresi Poisson sebagai berikut:

$$y = \exp (3,083 - 0,000215x_1 - 0,02811 x_3 - 0,0174 x_4 + 0,119 x_5)$$

Nilai *Pearson χ^2* pada hasil regresi Poisson dengan derajat bebas 31 menyebabkan nilai \hat{u} (5,171) >1. Kejadian ini mengidentifikasi terjadinya *overdispersi* yang menyebabkan model regresi Poisson tidak layak untuk memodelkan sehingga perlu dilakukan pemodelan lain dengan menggunakan regresi ZIP sebagai alternatif memperbaiki model karena variabel angka kebakaran ada yang bernilai nol.

4. Pemodelan Regresi ZIP

Model regresi ZIP yang menjelaskan hubungan antara angka kebakaran (Y) dengan kepadatan penduduk (X_1), luas lahan tebu dan sawah (X_3), luas lahan kosong yang ditumbuhi ilalang (X_4), tahun (X_5) yang dikombinasikan sehingga terbentuk model-model pada Tabel 4 berikut:

Tabel 4 Model-model Regresi ZIP

Parameter	Estimasi	Uji Simultan		Uji Parsial		R^2	AIC
		G_{value}	Keputusan	$ t_{value} $	Keputusan		
γ_0	-33,9425			2,14E6	Tolak H_0		
γ_1	0,03704			0,31	Terima H_0		
γ_3	-29,7996			259260	Tolak H_0		
γ_5	-42,2151	241,4	Tolak H_0	2,67E6	Tolak H_0	70,86	257,4
β_0	2,7614			11,60	Tolak H_0		
β_1	-0,00009			1,81	Terima H_0		
β_3	-0,03145			3,84	Tolak H_0		
β_5	0,05046			1,00	Terima H_0		

Berdasarkan Tabel 4 bahwa model signifikan secara simultan sehingga secara keseluruhan angka kebakaran disetiap wilayah kecamatan Kabupaten Sidoarjo dipengaruhi oleh kepadatan penduduk, luas lahan tebu dan sawah, dan tahun. Sesuai uji signifikan secara parsial, model memiliki parameter yang signifikan lebih 50% dari parameter yang ada pada model tersebut.

5. Pemilihan Model Terbaik dan Penjelasan Model

Model ZIP terbaik yang diperoleh dalam pemodelan kasus kebakaran disetiap wilayah kecamatan di Kabupaten Sidoarjo dari Tabel 4.4 adalah model yang dipengaruhi variabel prediktor kepadatan penduduk (X_1), dan luas lahan tebu dan sawah (X_3), dan kode tahun (X_5) karena nilai AIC (257,4) terkecil. Nilai R^2 (70,86) cukup besar yang berarti 70,86% angka kebakaran dipengaruhi variabel-variabel tersebut an sisanya 29,14% dipengaruhi hal lain. Model ZIP sebagai berikut:

$$\lambda = \exp(2,7614 - 0,00009x_1 - 0,03145x_3 + 0,05046x_5)$$

dan

$$\omega = \frac{\exp(-33,9425 + 0,03704x_1 - 29,7996x_3 - 42,2151x_5)}{1 + \exp(-33,9425 + 0,03704x_1 - 29,7996x_3 - 42,2151x_5)}$$

Artinya bahwa:

1. Setiap perubahan kepadatan penduduk, satu penduduk per kilometer persegi menyebabkan penurunan nilai harapan terjadinya kebakaran sebesar 1,00009 kali dan kenaikan peluang tidak terjadinya kebakaran sebesar 0,509259 kali,
2. Setiap perubahan luas lahan sawah, satu kilometer persegi menyebabkan penurunan nilai harapan terjadinya kebakaran sebesar 1,03195 kali dan penurunan peluang tidak terjadinya kebakaran sebesar 0 kali,
3. Tahun menjelaskan perbedaan nilai harapan angka kebakaran disetiap kecamatan pada tahun yang berbeda adalah 1,05175 atau bisa dikatakan 1 kejadian dan mempengaruhi peluang terjadinya kebakaran sebesar 0 kali.

KESIMPULAN

Berdasarkan hasil analisis terhadap model regresi *Zero-Inflated Poisson* (ZIP) maka dapat disimpulkan:

1. Model regresi *Zero-Inflated Poisson* (ZIP) yang terbaik untuk menentukan hubungan variabel angka kebakaran dengan faktor-faktornya yaitu:

$$\lambda = \exp(2,7614 - 0,00009x_1 - 0,03145x_3 + 0,05046x_5)$$

dan

$$\omega = \frac{\exp(-33,9425 + 0,03704x_1 - 29,7996x_3 - 42,2151x_5)}{1 + \exp(-33,9425 + 0,03704x_1 - 29,7996x_3 - 42,2151x_5)}$$

2. Faktor-faktor yang paling dominan dengan nilai $R^2(70,86)$ yaitu kepadatan penduduk (X_1) menyebabkan penurunan nilai harapan terjadinya kebakaran sebesar 1,00009 kali dan kenaikan peluang tidak terjadinya kebakaran sebesar 0,509259 kali, luas lahan tebu dan sawah (X_3) menyebabkan penurunan nilai harapan terjadinya kebakaran sebesar 1,03195 kali dan penurunan peluang tidak terjadinya kebakaran sebesar 0 kali, sedangkan tahun (X_5) menjelaskan perbedaan nilai harapan angka kebakaran disetiap kecamatan pada tahun yang berbeda adalah 1,05175 atau bisa dikatakan 1 kejadian dan mempengaruhi peluang terjadinya kebakaran sebesar 0 kali.

SARAN

Penelitian selanjutnya bisa menggunakan variabel terikat dengan data yang lebih banyak terdapat nilai nol serta variabel bebas yang lebih banyak tetapi saling bebas dan memiliki korelasi yang tinggi terhadap variabel terikat.

DAFTAR RUJUKAN

- Agresti, A. 2002. *Categorical Data Analysis*. New York: John Wiley and Sons.
- Cameron, A.C, & Trivedi, P.K. 1998. *Regression Analysis of count data*. Cambridge: Cambridge University Press.
- Hilbe, J. M. 2011. *Negative Binomial Regression*. Edisi Kedua. Cambridge: Cambridge University Press.
- Hossain, S., & Howlader, H. A. 2015. Estimation Techniques for Regression Model with Zero-inflated Poisson Data. *International Journal of Statistics and Probability*, 4 (4): 64-76.
- Lambert, D. 1992. Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, 34 (1):1-14.
- Myers, R. H. 1990. *Classical and Modern Regression with Applications*. Boston: PWS-KENT Publishing Company.