 http://dx.doi.org/10.29313/.v0i0.30866

Pemodelan Data Besar Klaim Menggunakan Model *Generalized Pareto Distribution*

Ajri Tsani Lazuardillah*, Aceng Komarudin Mutaqin

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

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

Abstract. Claims size modeling is a way to predict the claim size that can be used to estimate the pricing of insurance premiums in the future. Peaks-Over-Threshold (POT) is a method that can be used for extreme claim big data, where one of the distributions in the POT method is the generalized Pareto distribution (GPD). The focus of this research is the claim size modeling using the generalized Pareto distribution (GPD) model. The data modeling procedures using the GPD model are: (1) matching distribution with existing distribution in eassyfit 5.5 software, (2) determining the limits on the big data claims using the heuristic method, (3) estimating the distribution parameters of the GPD model using the 5.5 easyfit software, (4) distribution fit test using the Kolmogorov-Smirnov test. As an application material, secondary data obtained from the insurance company PT. X 2014, the data contains big data on claims for Partial Loss of Category 5 motor vehicle insurance policy holders (vehicle prices are more than Rp. 800,000,000) Region 3 (other than Region 1 and Region 2) in Indonesia. The calculation results show that the big data for motor vehicle insurance claims for Category 5 Region 3 in Indonesia follows the GPD model.

Keywords: Generalized Pareto Distribution, Heuristik Method, Kolmogorov-Smirnov Test.

Abstrak. Pemodelan besar klaim merupakan cara untuk memprediksi besar klaim yang dapat digunakan untuk memperkirakan penetapan harga premi asuransi di masa datang. *Peaks-Over-Threshold* (POT) adalah salah satu metode yang dapat digunakan untuk data besar klaim yang ekstrim, yang mana salah satu distribusi dalam metode POT adalah *generalized Pareto distribution* (GPD). Fokus pada penelitian ini adalah pemodelan besar klaim menggunakan model *generalized Pareto distribution* (GPD). Prosedur pemodelan data menggunakan model GPD adalah: (1) mencocokan distribusi dengan distribusi yang ada pada *software eassyfit 5.5*, (2) menentukan batas (thereshold) pada data besar klaim menggunakan metode heuristik, (3) menaksir parameter distribusi model GPD menggunakan bantuan *software eassyfit 5.5*, (4) uji kecocokan distribusi menggunakan uji Kolmogorov-Smirnov. Sebagai bahan aplikasi akan digunakan data sekunder yang diperoleh dari perusahaan asuransi PT. X tahun 2014, data tersebut berisi data besar klaim *Partial Loss* pemegang polis asuransi kendaraan bermotor Kategori 5 (harga kendaraan lebih dari Rp. 800.000.000) Wilayah 3 (selain Wilayah 1 dan Wilayah 2) di Indonesia. Hasil perhitungan menunjukan bahwa data besar klaim asuransi kendaraan bermotor Kategori 5 Wilayah 3 di Indonesia mengikuti model GPD.

Kata Kunci: *Generalized Pareto Distribution*, Metode Heuristik, Uji Kolmogorov-Smirnov.

1. Pendahuluan

Setiap manusia selalu dihadapkan pada peristiwa yang tidak tentu dalam kehidupannya. Peristiwa tersebut bisa dalam keuntungan atau kerugian. Keadaan dimana manusia dihadapkan dengan kerugian yang tidak pasti, keadaan inilah yang disebut risiko. Contohnya kecelakaan,

kehilangan, dan lain sebagainya yang mungkin saja terjadi. Tidak ada manusia yang bisa memprediksi dengan tepat peristiwa-peristiwa yang akan terjadi, karena itu murni hanya Tuhan yang mengetahuinya. Untuk meminimalisir risiko yang dihadapi, manusia perlu pihak lain yang dapat mentransferkan risikonya, yang mana itu adalah perusahaan asuransi.

Asuransi merupakan sebuah mekanisme pemindahan risiko dari suatu pihak tertanggung (*insured*) kepada pihak penanggung (*insurer*) dengan tanda jadi pembayaran sejumlah dana atau yang disebut premi kepada pihak penanggung dengan bukti yang disebut polis. Polis berfungsi sebagai bukti tertulis perjanjian antara pihak penanggung dan pihak tertanggung. Selain polis dan premi dalam asuransi ada juga yang disebut besar klaim, yaitu besar nominal yang harus diberikan oleh pihak penanggung kepada pihak tertanggung jika terjadi kerugian yang sesuai dengan yang dipertanggungkan berdasarkan polis. Agar pihak penanggung dapat menentukan jumlah premi yang tepat, pihak penanggung perlu memprediksi besar klaim di masa yang akan datang.

Pemodelan besar klaim merupakan cara untuk memprediksi besar klaim yang dapat digunakan untuk memperkirakan penetapan harga premi asuransi di masa datang. Cadangan biaya klaim dari perusahaan asuransi cenderung lebih besar dipengaruhi oleh besar klaim dari pada frekuensi klaim. Kerugian yang besar merupakan salah satu bagian utama yang harus diperhatikan dalam mengganti biaya yang dikeluarkan perusahaan asuransi. Kerugian besar sering disebut kerugian ekstrim yang mana bisa terjadi sesekali atau beberapa kali. Peristiwa ekstrim sangat menarik bagi aktuaris karena berperan penting dalam prosedur pengambilan keputusan untuk asuransi, misalnya untuk menghitung premi.

Memilih model atau distribusi untuk besar klaim sangatlah penting, terutama jika terdapat data besar klaim yang ekstrim. Nilai ekstrim ini dapat diartikan sebagai outlier dari data klaim yang mengindikasikan dapat terjadinya penyimpangan-penyimpangan dalam asuransi tersebut. *Peaks-Over-Threshold* (POT) adalah salah satu metode yang dapat digunakan untuk data besar klaim yang ekstrim. Cara kerja dari POT adalah dengan mencari sebuah *threshold* (batas) dari data sehingga dapat dijadikan sebuah batas untuk data lain dikatakan sebagai nilai ektrim, setelah didapat nilai ekstrim, maka dapat dianalisis untuk memprediksi hal-hal buruk yang bisa terjadi. Salah satu distribusi dalam metode POT adalah *generalized Pareto distribution* (GPD).

Dalam penelitian ini model *Generalized Pareto distribution* akan diterapkan pada data besar klaim asuransi umum di Indonesia.

2. Metodologi

Mencocokan Model

Metode yang akan digunakan pada penelitian ini adalah model *Generalized Pareto Distribution* (GPD). Fungsi dari distribusi GPD adalah :

$$G_u(x;\ \gamma,\sigma) = \begin{cases} 1 - \left[1 + \dfrac{\gamma(x-u)}{\sigma}\right]^{-1/\gamma} & ,\ \gamma \neq 0 \\ 1 - exp\left\{-\dfrac{x-u}{\sigma}\right\} & ,\ \gamma = 0 \end{cases}$$

dimana $x > u$ jika $\gamma \geq 0$ dan $0 \leq x -< \sigma / \gamma$ jika $\gamma < 0$.

Untuk mempermudah penelitian ini dalam penaksiran dan uji kecocokan digunakan *software eassyfit 5.5*. Sebelum dapat menggunakan bantuan *software eassyfit 5.5* perlu dilakukan pencocokan distribusi terlebih dahulu antara distribusi yang akan dipakai dengan distribusi yang tertera dalam *software eassyfit 5.5*. Fungsi distribusi Gen. Pareto pada *software eassyfit 5.5* adalah:

$$F(x) = \begin{cases} 1 - \left[1 + \dfrac{k(x-\mu)}{\sigma}\right]^{-1/k} & ,\ \gamma \neq 0 \\ 1 - exp\left\{-\dfrac{x-\mu}{\sigma}\right\} & ,\ \gamma = 0 \end{cases}$$

Model *Generalized Pareto Distribution (GPD)* cocok dengan distribusi Gen. Pareto yang terdapat pada *software eassyfit 5.5*.

Menyusun data

Sebelum melakukakan analisis lanjutan diharuskan Menyusun data terlebih dahulu. Data disusun dari mulai data yang bernilai kecil hingga data yang bernilai besar. Pada penelitian ini data yang diaplikasikan yaitu data besar klaim Kategori 5 Wilayah 3.

Memilih Ambang Batas

Pada tahapan ini dicari ambang batas yaitu dengan menggunakan metode heuristik. Metode heuristik pada dasarnya adalah suatu aturan praktis. Metode ini memiliki keunggulan yaitu sangat mudah untuk diimplementasikan. Mungkin itulah sebabnya metode heuristik sering digunakan dalam praktik oleh perusahaan asuransi.

Semua metode heuristik menaksir ambang batas u berdasarkan persamaan $\hat{u} = x_{([n-k])}$, dimana $[n-k]$ menyatakan bilangan integer diantara $1,\ldots,n$ yang paling dekat ke $n-k$ dan $k = \sqrt{n}$.

Uji Kecocokan

Pada tahapan ini setelah taksiran parameter didapat dilakukan uji kecocokan, yang mana pada penelitian ini uji kecocokan yang digunakan adalah uji Kolmogorov-Smirnof. Uji Kolmogorov-Smirnov merupakan salah satu uji kecocokan model untuk peubah acak kontinu. Hipotesis uji Kolmogorov-Smirnov adalah:

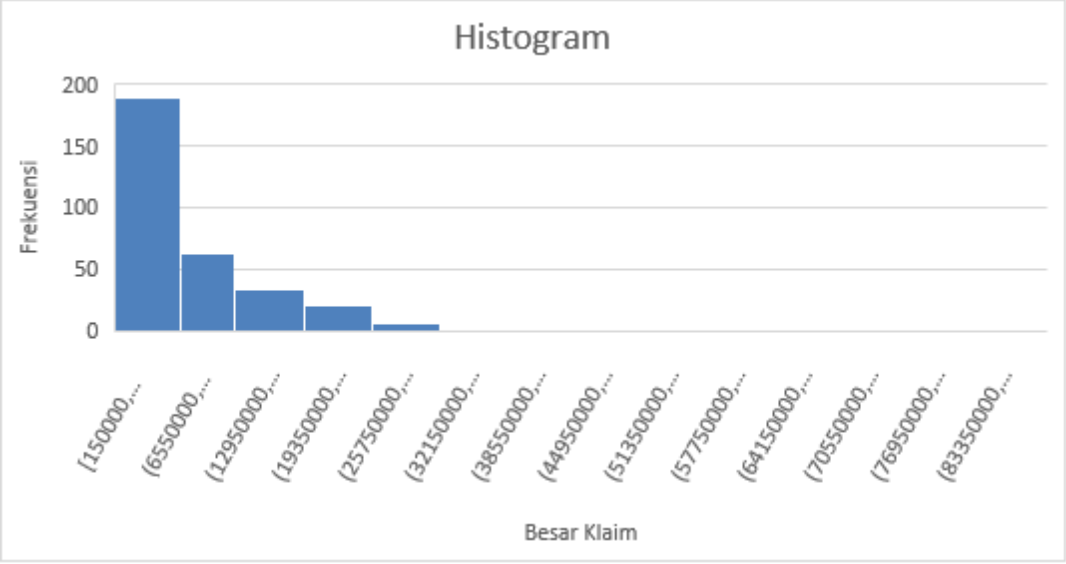
H_0: Data berasal dari suatu populasi berdistribusi $F(.)$

H_1: Data tidak berasal dari suatu populasi $F(.)$

Untuk membantu menghitung statistik uji Kolmogorv-Smirnov dihitung dengan bantuan *software eassyfit 5.5*. untuk statistic uji nya tolak H_0 jika $p - value \leq \alpha$.

3. Pembahasan dan Diskusi

Data yang digunakan dalam skripsi ini adalah data besar klaim Partial Loss pemegang polis asuransi kendaraan bermotor Kategori 5 (harga kendaraan lebih dari Rp.800.000.000,) Wilayah 3 (Jakarta, Jawa barat, dan Banten) di Indonesia yang bersumber dari perusahaan asuransi PT. X tahun 2014. Datanya terdiri dari 323 pengamatan besar klaim *partial loss*. Nilai minimum, maksimum, dan rata-ratanya masing-masing adalah Rp. 150.000, Rp. 88.615.700, dan Rp. 8.949.752 dilihat dati Tabel 1 yang disajikan. Histogram untuk data besar klaim *partial loss* di atas disajikan dalam gambar 1.



Gambar 1. Histogram untuk data besar klaim *partial loss*

Statistic	Value
Sample Size	323
Range	88.466.000
Mean	8.949.800
Variance	$1{,}5743 \times 10^{14}$
Std. Deviation	12.547.000
Coef. of Variation	1,402
Std. Error	698.140
Skewness	3,5007
Excess Kurtosis	15,917

Percentile	Value
Min	150.000
5%	553.000
10%	831.400
25% (Q1)	1.615.000
50% (Median)	4.875.000
75% (Q3)	11.219.000
90%	21.525.000
95%	27.301.000
Max	8.8616.000

Tabel 1. Deskriptif data menggunagan *software easstfit 5.5*

Berdasarkan data diketahui banyaknya pengamatan besar klaim adalah, $n = 323$. Kemudian nilai $k = \sqrt{n} = \sqrt{323} = 17{,}9722$. Dengan demikian nilai taksiran ambang batasnya, $\hat{u} = x_{([n-k])} = x_{(305)} = 24.864.500$

Setelah diperoleh nilai taksiran ambang batas, data dibagi menjadi dua bagian yaitu data di bawah ambang batas yaitu data yang lebih kecil atau sama dengan u dan data di atas ambang batas yaitu data yang lebih besar dari $\hat{u}$.

Hasil Pengujian Kecocokan Distribusi

Data di bawah Ambang Batas

Dalam bagian ini akan dilakukan pengujian kecocokan distribusi standar untuk data besar klaim asuransi kendaraan bermotor Indonesia kategori 5 wilayah 3 yang berada di bawah ambang batas menggunakan uji kecocokan Kolmogorov-Smirnov. Pengujiannya dibantu dengan mengunakan perangkat lunak *EasyFit 5.5*. Distribusi standar yang akan diuji kecocokannya adalah Gen. Pareto (GPD). Hipotesis untuk pengujian tersebut adalah:

H_0 : Data besar klaim asuransi kendaraan bermotor kategori 5 wilayah 3 yang berada di bawah ambang batas berasal dari populasi yang berdistribusi GPD.

H_1 : Data besar klaim asuransi kendaraan bermotor kategori 5 wilayah 3 yang berada di bawah ambang batas bukan berasal dari populasi yang berdistribusi GPD.

Dengan bantuan perangkat lunak *EasyFit 5.5* dapat diperoleh nilai $p - value = 0{,}0656$. Dengan taraf nyata $\alpha = 5\%$, terlihat bahwa nilai *p-value* lebih besar dari taraf nyata $(0{,}0656 > 0{,}05)$. Dengan demikian hipotesis nol diterima dan disimpulkan bahwa data besar klaim asuransi kendaraan bermotor kategori 5 wilayah 3 yang berada di bawah ambang batas berasal dari populasi yang berdistribusi GPD. Nilai taksiran parameter GPD standar berdasarkan perangkat lunak *EasyFit 5.5* adalah $\gamma = 0{,}0128$ dan $\sigma = 6.955.300$.

Data di atas Ambang Batas

Dalam bagian ini akan dilakukan pengujian kecocokan distribusi standar untuk data besar klaim asuransi kendaraan bermotor Indonesia kategori 5 wilayah 3 yang berada di atas ambang batas menggunakan uji kecocokan Kolmogorov-Smirnov. Pengujiannya dibantu dengan mengunakan perangkat lunak *EasyFit 5.5*. Distribusi standar yang akan diuji kecocokannya adalah Gen. Pareto (GPD). Hipotesis untuk pengujian tersebut adalah:

H_0 : Data besar klaim asuransi kendaraan bermotor kategori 5 wilayah 3 yang berada di atas ambang batas berasal dari populasi yang berdistribusi GPD.

H_1 : Data besar klaim asuransi kendaraan bermotor kategori 5 wilayah 3 yang berada di atas ambang batas bukan berasal dari populasi yang berdistribusi GPD.

Dengan bantuan perangkat lunak *EasyFit 5.5* dapat diperoleh nilai $p - value = 0{,}1447$. Dengan taraf nyata α = 5%, terlihat bahwa nilai *p-value* lebih besar dari taraf nyata

$(0{,}1447 > 0{,}05)$. Dengan demikian hipotesis nol diterima dan disimpulkan bahwa data besar klaim asuransi kendaraan bermotor kategori 5 wilayah 3 yang berada di atas ambang batas berasal dari populasi yang berdistribusi GPD. Nilai taksiran parameter GPD standar berdasarkan perangkat lunak *EasyFit 5.5* adalah $\gamma = 0{,}20955$ dan $\sigma = 33.488.000$.

4. Kesimpulan

Berdasarkan hasil penelitian ini dapat disimpulkan bahwa model *generalized Pareto distribution* (GPD) dapat digunakan untuk pemodelan data besar klaim asuransi kendaraan Kategori 5 Wilayah 3.

Acknowledge

Dengan kerendahan hati dan rasa hormat yang sangat dalam, penulis mengucapkan terima kasih kepada:

Tuhan Yang Maha Esa Allah SWT atas rahmat dan hidayah-Nya, penulis dapat menyelesaikan penelitian ini dengan lancar.

Orang tua dan keluarga serta kerabat yang senantiasa selalu mendo'akan serta memberikan dukungan moral maupun materi kepada penulis.

Seluruh keluarga besar Statistika Unisba, terimakasih atas segala dukungan, semangat, dan motivasi.

Kepada semua pihak yang tidak dapat disebutkan satu persatu yang telah berkenan memberikan bantuan serta dukungan kepada penulis untuk menyelesaikan penelitian baik secara langsung maupun tidak langsung.

Semoga Allah SWT senantiasa memberikan karunia, rahmat dan hidayah nya untuk memberi balasan atas semua bantuan yang telah diberikan kepada penulis. Besar harapan bahwa penelitian ini dapat memberikan manfaat bagi penulis dan bagi para pembaca.

Daftar Pustaka

[1] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete. *Journal of the Royal Statistical Society*, Series B 39.

[2] Klugman, S. A., Panjer, H. H., & Willmot, G. A. (2008). *Loss Models.* Canada: Edisi Ketiga. Jhon Wiley and Sons, Inc.

[3] Law, A. M., & Kelton, W. D. (1991). *Simulation Modeling and Analysis*. New York: Edisi kedua. McGraw-Hills, Inc.

[4] Leadbetter, M. R. (1990). On a basis for "peaks over thereshold" modeling. *Center for Stochastic Processes*, 1-10.

[5] Lee, D., Li, W. K., & Wong, T. S. (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-thereshold approach. *Insurance: Mathematics and Economics*, 538-550.

[6] Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 3*, 119–131.

[7] Safitri, R.P., & Mutaqin, A. K. (2020). Pemodelan Besar Klaim Asuransi Kendaraan Bermotor Indonesia Menggunakan Model Komposit Log-Logistik – Generalized Pareto Distribution. *Statistika, 1-27.*

[8] Yulianto Anggi Priliani, Darwis Sutawanir. (2021). *Penerapan Metode K-Nearest Neighbors (kNN) pada Bearing*. Jurnal Riset Statistika, 1(1), 10-18.