

Aplikasi Pemodelan Klaim Asuransi dengan Pendekatan *Mixture Exponential* Untuk Mencari Value-at-Risk Sebagai Threshold Dalam Menentukan Nilai Ekstrim

Fikri Nur Hadiansyah

Jurusan Ilmu Komputasi, Fakultas Informatika, Universitas Telkom, Bandung

e-mail: nurhadiansyahfikri@gmail.com

Abstrak : Klaim asuransi merupakan indikator dalam menentukan tingkat risiko sebuah perusahaan asuransi. Semakin besar peluang klaim terjadi, maka semakin besar pula risiko yang dapat ditanggung oleh perusahaan asuransi. Peaks-Over-Threshold merupakan salah satu metode extreme value theory yang digunakan untuk mencari nilai ekstrim pada data dengan membuat suatu batas dari data tersebut. Selain itu, threshold merupakan nilai VaR sehingga nilai yang melebihi batas dari VaR tersebut dapat dikatakan sebagai nilai ekstrim.

Kata kunci : Asuransi, pemodelan, extreme value theory, Peaks-Over-Threshold, Value-at-Risk

1. Pendahuluan

Asuransi adalah pertanggung jawaban ekonomi pada sebuah kejadian seperti kebakaran atau meninggal dunia agar mendapatkan uang ganti rugi [1]. Asuransi termasuk salah satu produk keuangan yang mempunyai risiko yang cukup besar karena klaim yang terjadi tidak dapat diperhitungkan secara pasti baik dari waktu maupun jumlah. Klaim adalah penggunaan hak pemegang asuransi untuk mendapatkan ganti rugi atas musibah yang diterimanya.

Klaim yang tidak dapat diperhitungkan secara pasti menjadi penyebab utama asuransi menjadi sangat rentan mengalami kerugian, terlebih perusahaan asuransi pada zaman sekarang mempunyai peserta yang sangat banyak yang membuat risiko kebangkrutan menjadi semakin banyak pula. Melihat permasalahan ini, perlu adanya sebuah metode untuk mencari kemungkinan-kemungkinan permasalahan yang akan terjadi karena diakibatkan dari pengajuan klaim.

Peaks-Over-Threshold (POT) adalah salah satu teknik dari metode Extreme Value theory (EVT) yang dapat digunakan untuk mencari sebuah nilai asing dari sebuah data yang kemudian disebut sebagai nilai ekstrim. Nilai ekstrim ini dapat diartikan sebagai *outlier* dari data klaim yang mengindikasikan dapat terjadinya penyimpangan-penyimpangan dalam asuransi tersebut. Cara kerja dari POT adalah dengan mencari sebuah threshold (batas) dari data sehingga dapat dijadikan sebuah batas untuk data lain dikatakan sebagai nilai ekstrim, setelah didapat nilai ekstrim, maka dapat dianalisis untuk memprediksi hal-hal buruk yang bisa terjadi seperti klaim dalam jumlah besar ataupun kecurangan-kecurangan. Maka dari itu akan dicoba untuk menganalisis data-data klaim dengan menggunakan teknik EVT dengan menggunakan distribusi Mixture Eksponensial dari data dengan harapan dapat mendapatkan informasi mengenai hal-hal yang dapat menyebabkan terjadinya kebangkrutan atau hal negative lain sehingga dapat membantu perusahaan asuransi untuk menentukan kebijakan-kebijakan keuangannya.

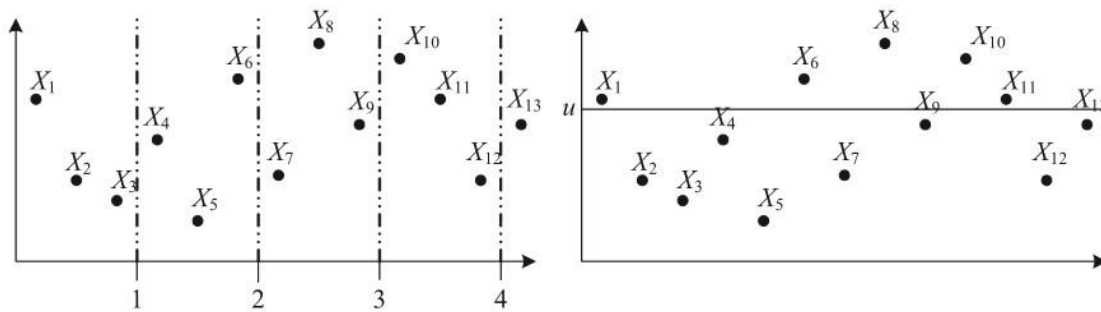
2. Landasan Teori

2.1. Extreme Value Theory

Extreme Value Theory adalah sebuah metode statistik yang digunakan untuk mencari dan menganalisis observasi data ekstrim [6]. EVT dapat diaplikasikan dalam berbagai macam bidang seperti cuaca, keuangan, risiko bencana dan lain sebagainya, dimana jika terdapat data yang menyimpang maka akan mengakibatkan risiko yang cukup besar. Terdapat dua macam pemodelan dalam teori EVT, yaitu Block Maxima (BM) dan Peaks-Over-Threshold (POT). Metode Block Maxima adalah teknik pengambilan suatu data dengan mengambil nilai yang maksimum dalam suatu periode. Nilai maksimum dari data dianggap sebagai nilai ekstrim. Metode BM mengikuti distribusi Gumbel, Frechet, dan Weibull. Bentuk standar dari ketiga distribusi tersebut disebut dengan distribusi Generalized Extreme Value [13].

$$\begin{aligned}
 & 1 - \left(1 + \left(\frac{x - \mu}{\sigma}\right) \left(\frac{1}{\xi}\right)\right)^{-1/\xi} \exp\left(-\left(1 + \left(\frac{x - \mu}{\sigma}\right) \left(\frac{1}{\xi}\right)\right)^{-1/\xi}\right), \xi \neq 0 \\
 & \frac{1}{\sigma} \exp\left(-\frac{x - \mu}{\sigma}\right) \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right), \xi = 0
 \end{aligned} \tag{1}$$

Metode yang kedua adalah Peaks-Over-Threshold, yaitu pemodelan dengan mencari nilai batas (threshold). Nilai-nilai yang dikatakan ekstrim adalah nilai yang melebihi batas yang telah ditemukan. Nilai threshold ini bersifat asymptotic mengikuti distribusi Generalized Pareto Distribution.



Gambar 2.1 Metode BM (kiri) dan Metode POT (kanan)

2.2. Generalized Pareto Distribution

Generalized Pareto Distribution (GPD) adalah model distribusi yang dapat mengatasi permasalahan ekstrim pada data [14]. Model distribusi ini merupakan heavy tail yaitu distribusi yang memiliki ekor lebih panjang dari pada distribusi normal. Metode GPD diturunkan dari konsep pengambilan data losses yang melebihi suatu nilai yang disebut threshold value, dan sering pula disebut metode Excesses Over Threshold value (EOT) [3].

$$\begin{aligned}
 F(x; \sigma, \gamma) &= 1 - \left[1 + \frac{x - u}{\sigma} \right]^{-1/\gamma}, & \gamma \neq 0, \\
 &\{ 1 - \exp\left\{-\frac{x - u}{\sigma}\right\} \}, & \gamma = 0
 \end{aligned}
 \tag{2}$$

Keterangan :
 u=threshold
 γ=parameter bentuk dari distribusi
 σ=parameter skala
 x=data

2.3. Distribusi Eksponensial

Distribusi eksponensial merupakan distribusi kontinu yang digunakan untuk memodelkan waktu penggunaan klaim. Distribusi eksponensial merupakan kasus khusus dari distribusi gamma dengan parameter bentuk α = 1 [6]

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}
 \tag{3}$$

Keterangan :
 λ=mean
 x=data

2.4. Mixture Distribution

Dalam probabilistik, *Mixture Distribution* adalah kumpulan dari variabel acak yang terkumpul dari variabel-variabel acak lainnya. Dengan kata lain, *Mixture Distribution* adalah menggabungkan beberapa distribusi yang sama menjadi sebuah distribusi model baru dengan tujuan mencari hasil yang lebih baik. Dalam tugas akhir ini, Mixture Distribution yang digunakan adalah Mixture Exponential dengan marginal distribusi eksponensial.

2.5. Value-at-Risk

Klaim asuransi merupakan hal yang paling berpengaruh dan berisiko pada bidang asuransi, dimana apabila klaim yang terjadi tidak terkendali, maka akan memberikan dampak yang sangat buruk bagi perusahaan asuransi. Untuk mengatasi hal ini, dapat digunakan salah satu metode yang bernama Value at Risk (VaR).

Value at Risk (VaR) adalah nilai statistik yang menunjukkan nilai risiko dari sebuah kejadian. VaR merupakan metode untuk menilai risiko dengan teknik pendekatan statistika [7]. VaR pada tingkat kepercayaan q% adalah invers dari fungsi distribusi dengan fungsi sebagai berikut :

$$F^{-1}(q) = VaR_q \tag{4}$$

Dimana F adalah fungsi distribusi kumulatif (cdf) dari nilai total loss.

VaR merupakan threshold antara nilai yang berisiko dengan bebas berisiko sehingga VaR sering dikaitkan dengan nilai threshold dengan tingkat q%.

3. Pembahasan
3.1. Data

Data yang digunakan dalam tugas akhir ini adalah data pengajuan klaim asuransi dari perusahaan A sebanyak 309068 periode 01 Januari 2013 sampai 23 Desember 2014.

Data kemudian ditransformasi untuk memudahkan perhitungan untuk beberapa tahap seperti untuk mencari parameter yang nantinya dikembalikan menjadi nilai asli.

3.2. Mencari Threshold

Pada tahap ini, threshold ditentukan dengan metode persentase yaitu dengan mengurutkan data dari yang terbesar hingga terkecil, kemudian menghitung 10% dari banyak data (k). Threshold adalah data dengan urutan k+1.

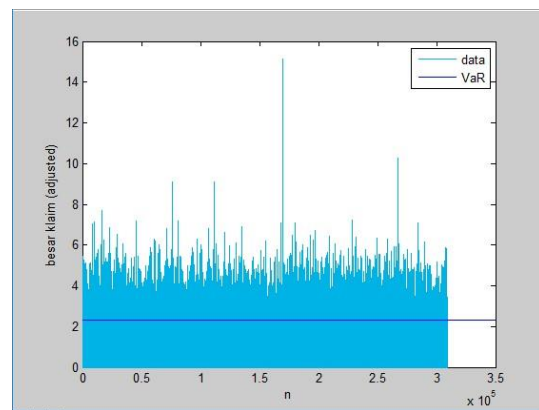
Tabel 1 Parameter Threshold

Jumlah data	309068
K	30907
U	2.302585093

3.3. Analisa Data Ektrim

Setelah diketahui nilai threshold, dapat dicari tahu nilai nilai yang diindikasikan sebagai nilai ekstrim dengan membandingkan dengan nilai threshold, apabila nilai tersebut melebihi nilai threshold, maka nilai tersebut diindikasikan sebagai nilai ekstrim.

Persamaan (2) adalah pengembangan model pareto dengan metode peaks-over-threshold dimana model tersebut membutuhkan sebuah batas (threshold) untuk bisa menentukan sebuah nilai dikatakan ekstrim. Simbol γ menyatakan parameter untuk indeks nilai ekstrim, sementara $\sigma > 0$ adalah skala parameter, dan x adalah nilai asimtotik konvergen untuk generalized Pareto distribution (GPD) dengan fungsi distribusi untuk model klaim asuransi ini. Melihat dari model tersebut, untuk dapat membaca full data set maka diasumsikan dengan distribusi sebagai berikut :



Gambar 1 Plot Data Dan Threshold U

$$f(x) = \begin{cases} p(x) & , x \leq u \\ p(x) + (1 - p(u))q(x) & , x > u \end{cases} \tag{5}$$

Dimana P(.) adalah fungsi distribusi. Dalam penelitian ini, akan diuji dengan menggunakan mixture model satu lambda dan dua lambda.

Seperti dijelaskan oleh David Lee (2012), dari persamaan (5) nilai P(.) diubah kedalam bentuk eksponensial $p(x) = 1 - e^{-\lambda x}$ sehingga fungsi densitas untuk satu lambda menjadi :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \leq u \\ \lambda e^{-\lambda x} + (1 - e^{-\lambda u})\lambda e^{-\lambda x} & , x > u \end{cases} \tag{6}$$

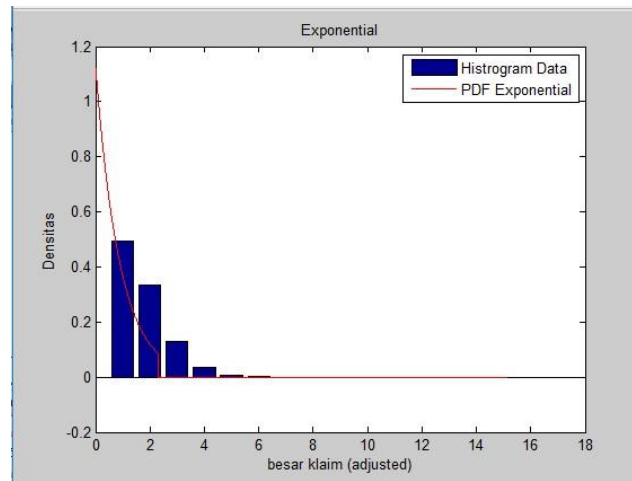
Keterangan

x : fungsi distribusi model

λ : mean untuk fungsi x

γ : parameter untuk indeks nilai ekstrim

σ : skala parameter



Gambar 2 Curve Fitting distribusi mixture exponential dengan satu lambda.

untuk persamaan dengan dua lambda, sama halnya seperti Mixture Exponential dengan satu lambda, nilai P diubah kedalam bentuk eksponensial $p(x) = 1 - e^{-\lambda x}$ sehingga fungsi densitas untuk dua lambda menjadi :

$$f(x) = \begin{cases} p_1(x) + (1 - p_1(u))p_2(x) & , x \leq u \\ (p_1(u) + (1 - p_1(u))p_2(u))p_1(x) + (1 - p_1(u) + p_1(u) - p_1(u)p_2(u))p_2(x) & , x > u \end{cases} \tag{7}$$

dimana $\lambda_1 > 0$, $\lambda_2 > 0$, dan $\lambda = (\lambda_1, \lambda_2)$, dari keseluruhan model yang telah dijelaskan, maka telah didapat 5 parameter yang akan digunakan untuk analisis data dalam model.

Keterangan

x : fungsi distribusi model

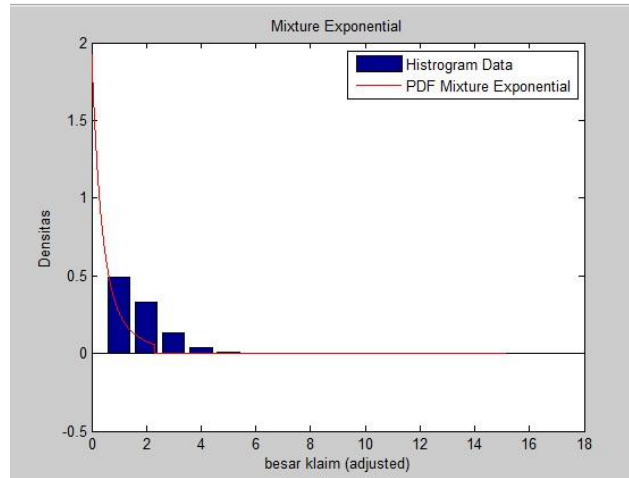
p : fungsi distribusi spesifik

λ_1 : mean untuk fungsi $x > u$

λ_2 : mean untuk fungsi $x \leq u$

γ : parameter untuk indeks nilai ekstrim

σ : skala parameter



Gambar 3 Curve Fitting distribusi mixture exponential dengan dua lambda

Pada gambar (1) dan (2) dapat dilihat bahwa dengan menggunakan model (7), data lebih cocok dibandingkan dengan menggunakan model (6). Terlihat dari garis merah yang melewati histogram data klaim lebih mendekati batang histogram dibanding dengan model satu lambda.

3.4. Value-at-Risk

Setelah mendapat nilai yang dinyatakan sebagai ekstrim, maka dapat dicari nilai VaR untuk mengetahui risiko yang terjadi dari data klaim tersebut, dimana VaR dalam data klaim ini menunjukkan nilai statistik kemungkinan terjadinya 1 klaim.

Untuk mencari VaR dari nilai ekstrim digunakan pendekatan GPD dimana dari persamaan (2.4) Jika $F(x)$ adalah distribusi nilai total loss x dan u merupakan nilai threshold, maka nilai Excess Over Threshold (EOT) adalah $x-u$. Dalam hal ini hanya kondisi dengan $x > u$, yaitu EOT positif yang diperhatikan. Maka distribusi untuk EOT adalah:

$$F(x-u) = \frac{P(X-u \leq X < u)}{P(X > u)} = \frac{F(u) - F(x-u)}{1 - F(u)} \tag{8}$$

Dan dengan pendekatan Mixture Exponential, maka akan terpenuhi persamaan

$$\begin{aligned} F(x-u) &= F(u) + \lambda \int_u^x f(t) dt \\ &= 1 - \hat{F}(u) + \lambda \int_u^x \hat{F}(t) dt \\ &= 1 + \hat{F}(u) [\lambda \int_u^x \hat{F}(t) dt - 1] \\ &= 1 + \hat{F}(u) + \lambda \int_u^x \hat{F}(t) dt \tag{9} \end{aligned}$$

$$= 1 + \frac{\lambda}{\gamma} \left(\hat{F}(u) \left(1 + \frac{\gamma(x-u)}{\sigma} \right)^{-\gamma} \right)$$

Maka untuk mendapat VaR dari persamaan (4.16) adalah dengan melakukan invers terhadap persamaan tersebut, sehingga didapat nilai VaR untuk satu lambda:

$$VaR = u + \frac{\hat{\sigma}}{\gamma} \left(\left(\frac{\lambda u}{\gamma} \right)^{\frac{n}{\gamma}} (1 - \alpha)^{\frac{1}{\gamma}} - 1 \right) \tag{10}$$

Setelah didapat persamaan tersebut, maka didapat nilai VaR sebagai berikut :

Tabel 3 Nilai VaR satu lambda untuk beberapa nilai kepercayaan

	90%	95%	99%
VaR	4.044	4.4615	5.3811

Untuk dua mean didapat persamaan nilai VaR:

$$VaR = u + \frac{\hat{\sigma}}{\gamma} \left(\left(\left(\frac{\lambda u}{\gamma} \right)^{\frac{n}{\gamma}} + (1 + \frac{\lambda u}{\gamma}) \right)^{\frac{n}{\gamma}} (1 - \alpha)^{\frac{1}{\gamma}} - 1 \right) \tag{11}$$

Setelah didapat persamaan tersebut, maka didapat nilai VaR sebagai berikut :

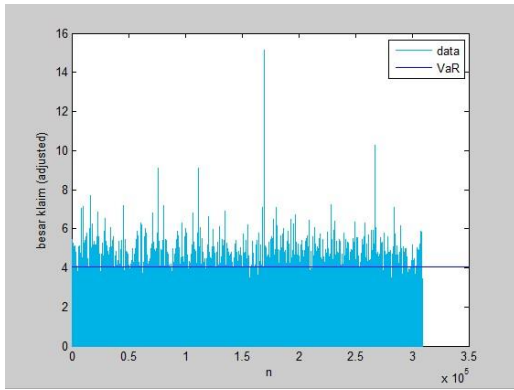
Tabel 4 Nilai VaR dua lambda untuk beberapa nilai kepercayaan

	90%	95%	99%
VaR	4.1835	4.5962	5.5064

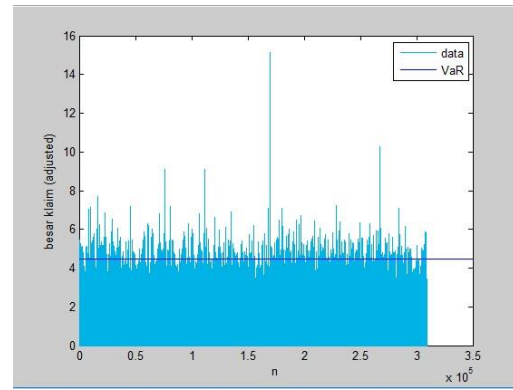
Dari tabel 3 dan 4 terlihat beberapa nilai VaR dari data klaim dengan tingkat kepercayaan 90%, 95% dan 99%. Nilai VaR untuk satu lambda dengan tingkat kepercayaan 90% adalah 4.044. Hal ini menandakan bahwa 90% dari jumlah keseluruhan data berada dibawah angka 4.044, begitupula untuk nilai dengan tingkat kepercayaan yang lainnya. Nilai yang berada diatas angka VaR tersebut dapat diindikasikan sebagai nilai ekstrim yang artinya apabila besar klaim yang terjadi melebihi batas nilai tersebut, maka perusahaan tersebut dapat mengalami kerugian.

3.5. Uji Validasi VaR

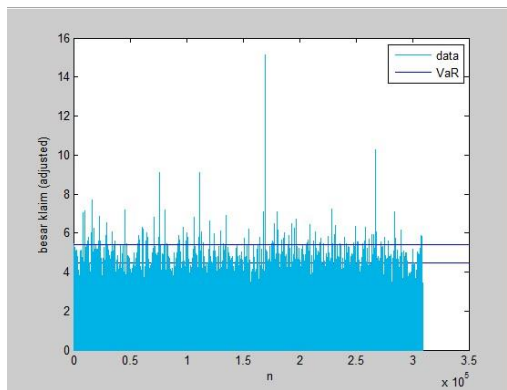
Setelah didapat nilai VaR untuk beberapa tingkat kepercayaan seperti terlihat pada tabel 4.4 dan 4.5, diuji nilai validasinya untuk menentukan VaR terbaik. Nilai VaR pada tabel tersebut menunjukan batas maksimal nilai klaim yang tidak berisiko. Artinya apabila ada nilai klaim yang diatas nilai tersebut, maka perusahaan asuransi itu berisiko mengalami kerugian.



Gambar 6 Grafik Data Frekuensi Terhadap Nilai Dengan VaR 90%



Gambar 7 Grafik Data Frekuensi Terhadap Nilai Dengan VaR 95%



Gambar 8 Grafik Data Frekuensi Terhadap Nilai Dengan VaR 99%

Pada gambar (6) terdapat data yang melewati batas nilai VaR, untuk VaR dengan tingkat kepercayaan 90%, artinya klaim asuransi tersebut seharusnya memiliki 10% ($\alpha=0.1$) nilai klaim asuransi yang menyebabkan perusahaan asuransi mengalami kerugian. Pada gambar (7) juga terdapat data yang melewati batas nilai VaR, untuk VaR dengan tingkat kepercayaan 95%, artinya klaim asuransi tersebut seharusnya memiliki 10% ($\alpha=0.05$) nilai klaim asuransi yang menyebabkan perusahaan asuransi mengalami kerugian dan begitu pula pada gambar (8) terdapat data yang melewati batas nilai VaR, untuk VaR dengan tingkat kepercayaan 99%, artinya klaim asuransi tersebut seharusnya memiliki 10% ($\alpha=0.01$) nilai klaim asuransi yang menyebabkan perusahaan asuransi mengalami kerugian. Maka akan dihitung tingkat validasinya dengan menghitung banyak nilai klaim yang melebihi batas VaR dibagi dengan total banyak nilai klaim.

Tabel 4 Uji Validasi Nilai VaR dengan Satu Lambda

α	Banyak data yang melebihi VaR	Persentase data yang melebihi nilai VaR
10%	1537	0.497308001 %
5%	741	0.239756167%
1%	171	0.055328342%

Tabel 5 Uji Validasi Nilai VaR dengan Dua Lambda

α	Banyak data yang melebihi VaR	Persentase data yang melebihi nilai VaR
10%	1207	0.389239769 %
5%	583	0.188634069%

1%	138	0.044650956%
----	-----	--------------

Berdasarkan nilai VaR pada tabel (4) dan (5) diperoleh nilai VaR terbaik untuk satu lambda adalah ketika tingkat kepercayaan 1% dengan nilai persentase = 0.055328346%. Untuk nilai VaR terbaik dengan dua lambda juga diperoleh ketika tingkat kepercayaan 1% dengan persentase = 0.044650946%.

4. Kesimpulan Dan Saran

Fitting data claim size model dengan menggunakan dua lambda lebih cocok dibandingkan dengan menggunakan satu lambda, hal ini terlihat dari gambar hasil fitting antara nilai pdf dengan histogram bahwa dengan dua lambda, kurva lebih mendekati sebaran data.

Akurasi VaR dengan tingkat kepercayaan 90%, 95% dan 99% menunjukkan bahwa VaR dengan dua lambda dengan tingkat kepercayaan 99% lebih akurat dengan persentase = 0.044650946%. Oleh karena itu, perlu diuji dengan menggunakan metode yang berbeda dan lebih akurat untuk menentukan *threshold* dalam POT.

5. Daftar Pustaka

- [1] Malika, Rosna., Sutikno. (2010). Declustering Peaks Over Threshold Pada Data Curah Hujan Ekstrem Dependen di Sentra Produksi Padi Jawa Timur. Intitute Teknologi Nasional
- [2] Adiperdana A; Suwignjo P; Rusdiansyah A. (2009). Analisis Value at Risk Menggunakan Metode Extreme Value Theory-Generalized Pareto Distribution dengan Kombinasi Algoritma Meboot dan Teori Samad-Khan. Journal of Insurance, 1-5
- [3] Lee, David., Keung, Wai., Tony S. (2012). Modeling insurance claims via a mixture exponential model combined with .peaks-over-threshold approach. Journal of Insurance: Mathematics and Economics, 1-7.
- [4] Bolviken, Erik . (2014). Computation and Modelling in Insurance and Finance. Cambridge : University Printing House.
- [5] Quiroga, Hector. (2008). The Auto Policy and The Claim Investigation Process. Washingtone : Spokane USA.
- [6] Dipak K; Dey,Jun Yan. (2016). Extreme Value Modeling and Risk Analysis: Methods and Applications. New york : CRC press
- [7] Harinaldi M. (2005). Prinsip Statistik U/teknik & Sains. Bandung : Erlangga
- [8] Mengersen ,Kerrie., Robert,Mike, Christian.(2011). Mixtures: Estimation and Applications. West Sussex
- [9] Zahara, Umami;M. Sjahid Akbar; Haryono. (2012). Penggunaan Metode VaR (Value at Risk) dalam Analisis Risiko Investasi Saham dengan Pendekatan Generalized Pareto Distribution (GPD). JURNAL SAINS DAN SENI ITS Vol. 1, No. 1, (Sept. 2012) ISSN: 2301-928X
- [10] J. McNeil ,Alexander., Frey Rudiger. (1999). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. 3-9.
- [11] Alamah, Fauzah. (2013). Analisis Risiko pada Portofolio Syariah dengan Pemodelan Value At Risk (Var) Block Maxima Generalization Extreme Value, journal of Konvegensi, 2-4.
- [12] Gede, I;Dharmawan, Komang; Putu, Luh. (2015). Penentuan Nilai Value At Risk Pada Saham Ihsg Menggunakan Model Geometric Brownian Motion Dengan Lompatan . Jurnal Matematika Vol. 4 67-73