Pemodelan Distribusi Kerugian Siber dengan Pendekatan *Copula* dan Perhitungan Premi Asuransi Siber

Putri Lathifah Idellie dan Raden Mohamad Atok
Departemen Aktuaria, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: moh_atok@statistika.its.ac.id

Abstrak—**Percepatan perkembangan teknologi dan informasi serta kemudahan akses internet mendorong proses digitalisasi semakin luas. Hampir seluruh alat komunikasi baik alat komunikasi pribadi maupun perusahaan menggunakan jaringan internet untuk menerima, mengirimkan serta menyimpan data. Digitalisasi tidak hanya memberikan dampak positif bagi pengguna namun juga terdapat risiko-risiko siber didalamnya. Perkembangan digitalisasi dan meluasnya penggunaan internet membuat risiko atas kerugian yang diakibatkan serangan siber kerap terjadi dan beragam. Beberapa jenis serangan siber yang umumnya terjadi dan dapat terlindungi asuransi diantaranya serangan *ransomware*, *business email compromise* (BEC), *phishing*, dan kelalaian sumber daya manusia (SDM). Salah satu cara untuk mengurangi kerugian atas dampak yang ditimbulkan dari risiko siber yaitu dengan menggunakan jasa asuransi umum berupa produk asuransi siber. Asuransi siber merupakan produk dari asuransi umum yang melindungi pemegang polis dari kejadian tidak menentu terkait kejadian yang berhubungan dengan pengadaan sistem informasi. Perhitungan harga produk asuransi siber berbeda dengan perhitungan premi asuransi tradisional. Asuransi siber tidak memiliki sistem penilaian standar atau tabel aktuaria untuk menentukan harga premi, premi asuransi siber menggunakan variabel *ranking* yang menjadi dasar pengambilan keputusan. Maka dari itu, pada penelitian ini dilakukan perhitungan premi murni asuransi siber dengan metode black scholes dari simulasi Monte Carlo berbasis copula terpilih berdasarkan data historis variabel banyak komputer terpilih (x_k) dan besar kerugian dalam dolar (y_k) sebagai fungsi distribusi kerugian dan waktu tunggu kejadian serangan siber (T) dengan simulasi proses poisson dan menghasilkan distribusi marginal x ~ Geometri (0,053974) dan y ~ Lognormal (11,04501; 2,13956), copula Frank dengan parameter $\hat{\theta}_F$ =11,42 dan didapatkan premi murni asuransi siber sebesar \$0 - \$165.870 untuk tiap perusahaan.**

Kata Kunci—**Asuransi Siber, *Black Scholes*, Copula Archimedean, Simulasi Monte Carlo, Simulasi Proses Poisson.**

I. PENDAHULUAN

PERCEPATAN perkembangan teknologi dan informasi serta kemudahan akses internet mendorong proses digitalisasi semakin luas. Hampir seluruh alat komunikasi baik alat komunikasi pribadi maupun perusahaan menggunakan jaringan internet untuk menerima, mengirimkan serta menyimpan data. Digitalisasi tidak hanya memberikan dampak positif bagi pengguna namun juga terdapat risiko-risiko siber didalamnya. Risiko siber termasuk dalam kegagalan dalam pengamanan serta terjadinya penyerangan sistem seperti peretasan, virus *ransomware*, penipuan *Business email compromise*, *phishing*, dan *human error*. Selain kerugian secara finansial serangan siber juga menyebabkan kerugian atas hilangnya potensi pendapatan, produktivitas, dan biaya legalitas hukum, serta kerugian-kerugian tidak berwujud seperti hilangnya niat baik pelanggan, reputasi, dan peluang bisnis.

Untuk menghindari risiko-risiko tersebut maka perusahaan menggunakan jasa asuransi umum yaitu, produk asuransi siber. Asuransi umum adalah adalah usaha jasa pertanggungan risiko yang memberikan penggantian kepada tertanggung karena kerugian, kerusakan, biaya yang timbul, kehilangan keuntungan, atau tanggung jawab hukum karena terjadinya suatu peristiwa yang tidak pasti. Asuransi siber memberikan jaminan perlindungan kepada lembaga korporasi, pemerintahan dan lembaga lainnya yang memiliki keterkaitan dengan aset data, keamanan jaringan, tanggung jawab privasi media, dan gangguan bisnis dalam kegiatan usahanya [1]. Penetapan premi produk asuransi siber berbeda dengan penetapan premi asuransi tradisional yang bergantung pada tabel aktuaria yang dibangun dari catatan sejarah. Premi asuransi siber dinilai dari klasifikasi dan urutan risiko saat penentuan premi asuransi siber didasarkan pada dua jenis informasi. Pertama, perusahaan asuransi telah menyediakan banyak variabel *ranking* yang menjadi dasar pengambilan keputusan. Kedua, perusahaan asuransi telah menyediakan riwayat pengalaman kerugian pemegang polis, dan riwayat ini dapat memberikan wawasan kepada pemegang polis yang tidak tersedia dari variabel *ranking* [2].

Untuk memodelkan distribusi kerugian pada tugas akhir ini digunakan metode *copula*. Metode *copula* dipilih karena dapat memodelkan variabel multivariat yang tidak saling bebas. Metode ini bersifat fleksibel karena tidak mensyaratkan asumsi normalitas dari data dan dapat menggabungkan beberapa distribusi marginal menjadi distribusi bersama. *Copula* mempunyai kemampuan untuk mendeskripsikan struktur dependensi antar variabel dengan marginal yang berbeda dan memodelkan dependensi *tail*-nya. *Copula* memiliki banyak keluarga, beberapa keluarga *Copula* yang paling popular adalah keluarga *copula* Ellips dan keluarga *copula* Archimedean. *copula* Ellips terbagi menjadi dua kelas yaitu *copula* Normal dan *copula* Student's-t, sedangkan *Copula* Archimedean terdiri dari tiga kelas *Copula* yaitu *copula* Clayton, *copula* Frank, dan *copula* Gumbel [3].

Oleh karena itu, pada tugas akhir ini akan dilakukan pemodelan kerugian siber dengan struktur dependensi menggunakan metode *copula* yang paling sesuai. Tugas akhir ini menggunakan data "*Dataset of Data Breaches and Ransomware Attacks over 15 Years from 2004 to 2020*" [4].

Pada tugas akhir ini variabel yang akan digunakan yaitu data banyaknya komputer terpengaruh *ransomeware* dan kerugian dalam dolar akibat serangan *ransomeware* sebagai titik awal untuk menggambarkan distribusi kerugian [5].

Tabel 1.
Distribusi kegagalan variabel banyak komputer terpengaruh (x)

Distribusi (x)	Persamaan Distribusi
Poisson	$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$
Geometri	$P(x) = p(1-p)^{x-1}$
Binomial Negatif	$f(x) = \begin{cases} \binom{x-1}{k-1} p^k (1-p)^{x-k} \\ 0 \end{cases}$

Tabel 2.
Distribusi frekuensi variabel banyak komputer terpengaruh (y)

Distribusi (y)	Persamaan Distribusi
Eksponensial	$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
Log-normal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left(\frac{(lnx-\mu)^2}{2\sigma^2}\right)$
Weibull	$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{k}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$

II. TINJAUAN PUSTAKA

A. Kejahatan Siber

Cybercrime atau kejahatan siber adalah penggunaan komputer atau alat lainnya yang terhubung dengan jaringan internet sebagai instrumen untuk tujuan ilegal seperti melakukan penipuan, pencurian kekayaan intelektual atau identitas, dan melakukan pelanggaran privasi. Kejahatan siber, terutama melalui Internet, semakin sering ditemui seiring perkembangan fungsi internet sebagai pusat perdagangan, hiburan, dan alat bantu pemerintahan. Perkembangan teknologi serta jaringan internet berawal mula dari Amerika Serikat. Dulu, sebagian besar korban dan pelaku kejahatan siber adalah orang Amerika namun, saat ini hampir semua bagian di dunia berisiko mengalami kejahatan siber [6].

Teknologi baru tidak hanya memberikan dampak positif namun juga menciptakan peluang kriminal. Hal yang membedakan kejahatan siber dari kegiatan kriminal tradisional adalah seiring bertambahnya peluang kejahatan siber, karena pertumbuhan jumlah pengguna, jenis kejahatan siber tidak terlalu berkembang. Pada umumnya, kejahatan siber merupakan perpanjangan dari perilaku kriminal transional yang melibatkan media internet seperti pencurian secara digital, penipuan secara digital, pencucian uang secara digital dan perdagangan ilegal secara digital.

Mayoritas kejahatan siber adalah serangan terhadap informasi tentang individu, perusahaan, atau pemerintah. Meskipun serangan tidak terjadi secara fisik namun, kumpulan identitas informasi digital merupakan bagian yang vital bagi perseorangan maupun institusi perusahaan. Identitas informasi virtual baik merupakan kumpulan angka dan pengenal di database-database komputer yang dimiliki oleh pemerintah dan perusahaan digital yang menghimpun berbagai input dari informasi-informasi personal. Kejahatan siber juga dapat diartikan sebagai kejahatan yang melibatkan pelanggaran mendasar terhadap privasi pribadi atau perusahaan, seperti serangan terhadap integritas informasi yang disimpan dalam penyimpanan digital dan penggunaan informasi digital yang diperoleh secara ilegal untuk memeras perusahaan atau individu [7].

B. Asuransi Siber

Asuransi siber merupakan produk asuransi yang digunakan untuk melindungi bisnis dari risiko siber. Asuransi siber dirancang untuk membantu perusahaan dalam memitigasi risiko dengan mengimbangi biaya yang dibutuhkan untuk melakukan pemulihan setelah terjadinya serangan yang terkait dengan keamanan siber atau sejenisnya. Asuransi siber akan memberikan perlinduangan darii risiko pelanggaran atau kebocoran data, biaya remediasi dalam penanganan serangan keamanan siber, dan denda hukuman atas pelanggaran terkait keamanan siber, Klasifikasi dan urutan risiko serangan siber saat penentuan premi didasarkan pada dua jenis informasi. Pertama, variabel *ranking* yang menjadi dasar pengambilan keputusan. Kedua, perusahaan asuransi telah menyediakan riwayat pengalaman kerugian pemegang polis, dan riwayat ini dapat memberikan wawasan kepada pemegang polis yang tidak tersedia dari variabel *ranking* [2].

C. Distribusi Marginal

Distribusi marginal adalah distribusi nilai untuk satu variabel yang mengabaikan kumpulan variabel terkait yang lebih luas dalam kumpulan data. Fungsi distribusi marginal memainkan peran penting dalam karakterisasi independensi antara variabel acak: dua variabel acak independen jika dan hanya jika fungsi distribusi bersamanya sama dengan produk dari fungsi distribusi marginalnya [8].

Distribusi banyak komputer terpengrauh (x) digunakan distribusi kegagalan; Distribusi poisson [9], Distribusi Geometri [10], dan Distribusi Binomial negatif seperti terlihat di Tabel 1 [11].

Distribusi kegagalan mengukur kemungkinan komputer pada suatu perusahaan akan gagal atau terpengaruh virus pada kurun waktu berikutnya, sedangkan sampai saat waktu tunggu kejadian (T) masih dalam keadaan baik.

Distribusi besar kerugian dalam dolar (y) digunakan distribusi frekuensi; Distribusi eksponensial [12], distribusi log-normal [13], dan distribusi Weibull seperti terlihat di Tabel 2 [14].

Distribusi frekuensi besar kerugian akibat serangan siber dalam dolar (y) membantu menjelaskan tingkat sebaran besaran kerugian yang dialami akibat kejadian siber.

D. Copula Archimedean

Copula digunakan untuk menganalisis hubungan dependensi antara variabel berdistribusi Uniform[0,1]. Copula Archimedean didefinisikan sebagai, $\phi: [0,1] \rightarrow [0,\infty)$ sebagai fungsi generator *copula* Archimedean dan ϕ^{-1} invers dari generator, sehingga $\phi(0) = \infty$ dan $\phi(1) = 0$. Fungsi *copula* Archimedean didefinisikan pada Tabel 3 [15].

Tabel 3 menjelaskan penulisan model copula untuk copula Archimedean bivariat. Plot copula Archimedean diilustrasikan pada Gambar 1 [16].

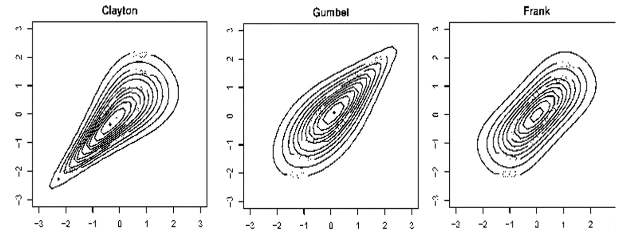
Gambar 1 menunjukkan tail dependensi dari copula Archimedean dimana, copula Clayton memiliki *tail* dependensi di bagian bawah, copula Gumbel memiliki *tail* dependensi di bagian atas, dan copula Frank idak meiliki *tail* [3].

E. *Transformasi ke Domain Uniform* [0,1]

Untuk melakukan analisis copula diperlukan transformasi variabel acak ke domain Uniform [0,1]. Distribusi uniform menghasilkan nilai variabel acak memperoleh nilai peluang

Tabel 3.
Keluarga copula Archimedean

Keluarga	Generator	$C(u_1, u_2)$
Clayton	$\theta^{-1}(u^{-\theta} - 1)$, $\theta > 0$	$(u_i{}^{-\theta} + m - 1)^{-\frac{1}{\theta}}$
Gumbel	$(-\ln u)^{\theta}, \theta > 1$	$\exp\{-[(-ln\, u_i)^{\theta} + (-ln\, u_2)^{\theta}]^{\frac{1}{\theta}}\}$
Frank	$ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$, $\theta \in \mathbb{R}\,\{0\}$	$\frac{1}{\theta} ln\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{\theta}-1}\right)$



(a) (b) (c)

Gambar 1. *Scatterplot* dependensi (a) copula Clayton, (b) copula Gumbel, (c) copula Frank

yang sama [9]. Sehingga persamaan tranformasi Uniform [0,1] didefinisikan sebagai Persamaan 1 [16].

$$\mathrm{C}(\mathrm{u}_1, \ldots, \mathrm{u}_\mathrm{m}) = \frac{1}{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathbf{1}\left(\frac{\mathrm{R}_1^{(\mathrm{j})}}{\mathrm{n}+1} \leq \mathrm{u}_1, \ldots, \frac{\mathrm{R}_\mathrm{m}^{(\mathrm{j})}}{\mathrm{n}+1} \leq \mathrm{u}_\mathrm{m}\right) \quad (1)$$

Transformasi variabel acak ke domain uniform dilakukan untuk melihat apakah variabel-variabel tersebut mengikuti sebuah fungsi distribusi tertentu dan sebagai *input* pemodelan copula.

F. Uji Dependensi

Uji dependensi adalah metode statistik yang digunakan untuk mengukur besarnya hubungan linier antara dua variabel dengan satu nilai yang dinamakan koefisien korelasi.. Pada tugas akhir ini digunakan korelasi Tau Kendal karena data diasumsikan tidak berdistribusi normal [15]

Korelasi Tau-Kendall merupakan statistik non-parametrik dengan skala pengukuran data minimal ordinal. Korelasi tau kendall digunakan untuk menganalisis korelasi copula Archimedean Asumsi lain yang harus terpenuhi adalah data merupakan sampel acak yang terdiri atas *n* pasangan hasil pengamatan (X_i,Y_i) [16]. Korelasi Tau Kendall dapat dituliskan sebagai Persamaan 2.

$$\tau = \frac{2(P-Q)}{n(n-1)} \quad (2)$$

Dengan hipotesis:
H_0 : Tidak ada korelasi antara *x* dan *y*
H_1 : Ada korelasi antara *x* dan *y*
Dan Statistik Uji:

$$Z = \sqrt{\frac{9n(n-1)}{2(2n+5)}}|\tau| \quad (3)$$

Dengan taraf signifikansi $\alpha = 0{,}05$ tolak H_0 jika Z > 0,05

Estimasi parameter copula Archimedean dengan pendekatan korelasi Tau-Kendall didefinisikan pada Tabel 4 [16].

G. Estimasi Parameter Copula

Metode *Maximum Likelihood Estimation* (MLE) digunakan untuk mendapatkan parameter *copula.* Pemilihan model *copula* terbaik dilakukan dengan memilih distribusi

Tabel 4.
Tabel tau kendall *copula* Archimedean

Kelas *Copula*	Estimasi (θ)
Clayton	$\tau = \frac{\theta_C}{\theta_C+2}$ maka $\theta_C = \frac{2\tau}{1-\tau}$
Frank	$\tau = 1 - \frac{1}{\theta_G}$ maka $\theta_G = \frac{1}{1-\tau}$
	$\tau = \frac{1-4(1-D_1(\theta_G))}{\theta_F}$
Gumbel	Dimana $D_k(x)$ =fungsi Debye
	$D_k(x) = \frac{k}{x^k}\int_0^x \frac{u^k}{e^x-1} du$

dengan nilai parameter terbesar [17]. *Log-likelihood* dapat didefiniskan pada Persamaan 4.

$$ln\, L = \sum_{i=1}^{n} lnc\{F_{X_1}(x_1^i), \ldots, F_{X_m}(x_m^i)\} + \sum_{j=1}^{m}\sum_{i=1}^{n} ln\left(f_{X_j}\left(x_j^{(i)}\right)\right) \quad (4)$$

Dan *maximum log-likelihood* didefinisikan dengan Persamaan 5.

$$\hat{\theta}_{MLE} = {}^{max}_{\theta} ln\, L \quad (5)$$

H. Perhitungan Premi dengan Fungsi Copula

Pada penelitian ini, definisi dan teori dari penelitian dijelaskan dalam deskripsi model, di mana deskripsi model terdiri atas:

1) Copula sebagai fungsi biaya kerugian (L)

Jika F fungsi distribusi bersama dengan $u_1 = F_x(x)$ dan $u_2 = F_y(y)$ adalah distribusi marjinalnya. Fungsi copula dapat dituliskan sebagai Persamaan 6 dan Persamaan 7 [15].

$$F(x,y) = C(F_x(x), F_y(y)) = C(x_k, y_k) \quad (6)$$

$$f(x,y) = c(F_x(x), F_y(y)) f_x(x), F_y(y) = c(x_k, y_k) \quad (7)$$

Untuk setiap $x, y \in \mathbb{R}$ [2].

2) Biaya Total Kerugian

Model distribusi kerugian digambarkan dengan fungsi *payoff* dari fungsi copula yang didapatkan pada Persamaan 8 [5].

$$L = \Pi = g(x,y) \quad (8)$$

Salah satu elemen risiko kontrak asuransi adalah biaya ganti rugi atau total kerugian yang dibayarkan. Besarnya *L* akan bergantung pada jumlah total kerugian pihak pertama yang didapat dari menghitung banyaknya komputer yang terpengaruh (*x*) dan jumlah kerugian dalam dolar akibat insiden (*y*) [5].

3) Perhitungan Premi Murni

Perhitungan premi murni asuransi siber merupakan modifikasi dari persamaan Black Scholes pada Persamaan 9 [5].

$$P = e^{-rT} L \quad (9)$$

Perhitungan tersebut didapatkan berdasarkan tinjauan dua elemen risiko pada kontrak asuransi, yaitu biaya ganti rugi yang dibayarkan (*L*) dan waktu tunggu kejadian (*T*). Variabel tingkat bunga (*r*) yang diasumsikan 6%. Variabel T adalah waktu tunggu kejadian diasumsikan bahwa jeda waktu antara kejadian dan biaya ganti rugi dibayarkan sangat pendek atau sama dengan nol hal ini dikarenakan pada penelitian ini hanya dihitung kerugian pihak pertama. *T* dapat dimodelkan dengan

Tabel 5.
Variabel tugas akhir

Variabel	Nama Variabel	Skala
x	Banyak komputer terdampak *ransomeware* dalam unit	Rasio
y	Besar kerugian serangan siber dalam dolar ($)	Rasio

proses Poisson. Distribusi Poisson banyak digunakan dalam model kedatangan instruksi per satuan waktu [18].

I. Pembangkitan Data Waktu dengan Proses Poisson

Memodelkan kejadian pertama pelanggaran keamanan atau waktu hingga terjadi pelanggaran pertama menggunakan prosedur simulasi proses Poisson [19].

$$T_k = t_k - \frac{1}{\lambda}\ln u_k \quad (10)$$

Untuk memodelkan waktu tunggu kejadian diasumsikan untuk spesifik suatu perusahaan dalam setahun terjadi dua kali kejadian $\lambda = 2$ dimana satu kejadian dan kejadian lain diasumsikan independen.

Algoritma untuk membangkitkan waktu kejadian (T) yaitu:
1. Bangkitkan nilai U dari distribusi Uniform[0,1].
2. Hasilkan $t_i = t_{i-1} - (\frac{1}{\lambda})\ln u$ dengan $t_0 = 0$.
3. Hasil bangkitan waktu tunggu kejadian sebanyak n=5000 kejadian.

III. METODOLOGI

A. Sumber Data

Data yang digunakan "*Dataset of Data Breaches and Ransomware Attacks over 15 Years from 2004 to 2020*" yaitu set data berisi observasi banyaknya komputer yang terserang *ransomeware* serta kerugiannya di Amerika Serikat dari tahun 2004-2020 [4].

B. Variabel Tugas akhir

Variabel yang akan digunakan pada tugas akhir ini dirangkum Tabel 5.

C. Tahapan Tugas akhir

Langkah yang dilakukan untuk mendapatkan nilai premi murni asuransi siber yaitu:
1. Memperoleh data dari set data "*Dataset of Data Breaches and Ransomware Attacks over 15 Years from 2004 to 2020*" yang diakses pada *website* https://espace.library.uq.edu.au/
2. Melakukan deskripsi data dan statistika deskriptif dengan minimum, maksmimum, mean, median, *skewness*, kurtosis.
3. Memodelkan distribusi data banyaknya komputer yang terpengaruh virus (x) dengan distribusi frekuensi, yaitu distribusi Poisson, distribusi Geometri, dan distribusi Negatif Binomial.
4. Memodelkan distribusi data jumlah kerugian dalam dolar akibat insiden (y) dengan distribusi kegagalan, yaitu distribusi Eksponensial, distribusi Log Normal, dan distribusi Weibull.
5. Uji kesesuaian model distribusi x dengan Perason's Chi Square dan distribusi y dengan uji Kolmogorov Smirnov.
6. Memilih model distribusi terbaik dari uji kesesuaian dengan *Akaike Information Criterion* (AIC).

Tabel 6.
Statistika deskriptif

		Jumlah Unit Terpengaruh (x)	Besar Kerugian (y)
Banyak Data (N)	Valid	1,057.00	1,057.00
	Hilang	-	-
Mean		1.851,74	3.018.855,55
Median		797,00	50.000,00
Skewness		7,10	14,68
Kurtosis		74,84	246,98
Minimum		26,00	253,00
Maksimum		45.557,00	500.000.000,00
Total		1.957.292,00	3.190.930.319,00

7. Melakukan transformasi U[0,1] pada data model distribusi x dan model distribusi y terpilih.
8. Mendapatkan nilai hubungan korelasi x dan y dengan uji korelasi Tau-Kendall.
9. Mencari estimasi parameter copula dengan *Maximum Likelihood Estimation*.
10. Memilih model copula terbaik dengan nilai *Maximum Likelihood Estimation* terbesar.
11. Membangkitkan data banyak komputer (x_k) dan besar kerugian (y_k) menggunakan fungsi copula terbaik.
12. Menghitung biaya ganti rugi (L) dengan data x_k dan y_k dengan mengacu Persamaan 8.
13. Membangkitkan waktu kejadian (T) menggunakan proses Poisson dengan mengacu Persamaan 10.
14. Menghitung premi (P) dari data yang telah dibangkitkan dengan Persamaan 9.
15. Melakukan interpretasi hasil analisis.
16. Penarikan kesimpulan dan saran.

IV. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

Pada tugas akhir ini digunakan data kejadian serangan siber (*cyberattacks*) di Amerika Serikat pada periode tahun 2004-2020. Pada pemodelan dependensi copula digunakan dua variabel dari set data yaitu variabel "Jumlah komputer terpengaruh" sebagai "x" dan "Besar kerugian" sebagai "y". Kedua variabel merupakan variabel numerik dengan skala rasio. Statistika deskriptif yang akan dijelakan meliputi mean, median, minimum, maksimum, *skewness* dan kurtosis dari kedua variabel tugas akhir yang dapat dilihat pada Tabel 6.

Banyak data yang diolah yaitu sebanyak 1056 observasi. Variabel x memiliki nilai mean atau nilai rata-rata sebesar 1.851,74-unit komputer pada dan variabel y memiliki nilai mean atau nilai rata-rata sebesar $ 3.018.855,55. Untuk nilai minimum atau nilai terkecil pada variabel x yaitu sebesar 26-unit komputer dan pada variabel y sebesar $253. Nilai maksimum atau nilai tertinggi pada variabel x yaitu sebesar 45.557-unit komputer dan pada variabel y yaitu sebesar $500.000.000. Hasil statistika deskriptif yang didapatkan nilai *skewness* dan nilai kurtosis. *Skewness* memberikan gambaran distribusi data apakah miring ke kiri, ke kanan atau simetris sedangkan kurtosis merupakan statistik yang digunakan dalam memberikan gambaran apakah distribusi data cenderung rata atau runcing Variabel x memiliki nilai *skewness* sebesar 7,10 dan variabel y memiliki nilai skewness sebesar 14,68. Nilai kurtosis pada variabel x sebesar 74,84 yang berarti data pada variabel x memiliki ekor distribusi berada di sebelah kanan dan tidak berdistribusi normal dan

Tabel 7.
Hasil penentuan parameter distribusi frekuensi

Distribusi Frekuensi	Simbol	Parameter
Poisson	$\hat{\lambda}$	1.851,742000
Geometri	$\hat{p}$	0,053974
Negatif Binomial	$\hat{\beta}$	0,894998
	$\hat{r}$	1.851,743000

Tabel 8.
Hasil penentuan parameter distribusi kegagalan

Distribusi Kegagalan	Simbol	Parameter
Eksponensial	$\hat{\lambda}$	0,000003
Log Normal	$\hat{\mu}$	11,04501
	$\hat{\sigma}$	2,13956
Weibull	$\hat{\beta}$	0,36890
	$\hat{\eta}$	173.747,00000

variabel y 246,98 yang berarti data pada variabel y memiliki ekor distribusi berada di sebelah kanan dan tidak berdistribusi normal [8].

B. Penentuan Distribusi Marginal

Untuk mencari parameter dari distribusi marginal Banyak Komputer Terpengaruh (x) yang diduga menyebar Poisson, Geometri, atau Binomial Negatif digunakan dan untuk mencari parameter dari distribusi marginal Besar Kerugian (y) yang diduga menyebar Eksponensial, Lognormal atau Weibull [10]. Distribusi poisson merupakan suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadian tergantung pada selang waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel prediktor saling independen [14].

Binomial Negatif merupakan alternatif dari distribusi Poisson dimana jika varians sampelnya melebihi rataan sampel disebut sebagai masalah overdispersi. Salah satu distribusi yang mampu menangani masalah overdispersi tersebut adalah distribusi Binomial negatif. Distribusi Eksponensial, Lognormal atau Weibull digunakan karena terdapat frekuensi yang tinggi terhadap besar kerugian dengan tingkat kerugian tertentu. Distribusi Eksponensial, Lognormal atau Weibull dapat digunakan pada data tidak berdistribusi normal [20].

1) Penentuan Nilai Parameter

Pada tahap ini, dilakukan penentuan nilai parameter distribusi Poisson, distribusi Geometri dan Binomial Negatif dengan persamaan tertera pada Tabel 1. Hasil penentuan parameter distribusi marginal Banyak komputer terpengaruh (x) dan Besar Kerugian (y) dapat dilihat pada Tabel 7.

Berdasarkan Tabel 7 hasil dugaan parameter untuk distribusi frekuensi didapatkan hasil parameter $\hat{\lambda} = 1.851,74$ untuk distribusi Poisson, $\hat{p}$ =0,053974 untuk distribusi Geometri dan $\hat{\beta}$ =0,894998 dan $\hat{r}$ =1851,743 untuk distribusi Negatif Binomial.

Berdasarkan hasil Tabel 8 hasil dugaan parameter untuk distribusi kegagalan didapatkan $\hat{\lambda} = 0{,}000003$ untuk distribusi Eksponensial, $\hat{\mu} = 11{,}04501$ dan $\hat{\sigma} = 2{,}13956$ untuk distribusi Log Normal dan $\hat{\beta}$ =0,36890 dan $\hat{\eta}$ =173.747 untuk distribusi Weibull.

2) Uji Kesesuaian Distribusi

Uji kesesuaian merupakan pengujian kebaikan sesuai antara hasil pengamatan (frekuensi pengamatan) tertentu dengan frekuensi yang diperoleh berdasarkan nilai harapannya (frekuensi teoritis) [16]. Uji kesesuaian dilakukan dengan uji Kolmogorov smirnov pada distribusi kegagalan.

Tabel 9.
Hasil uji Kolmogorov Smirnov untuk variabel *x* variabel distribusi frekuensi

Variabel	Distribusi Frekuensi	*p-value*
Banyak Komputer	Poisson	0,00000
	Geometri	0.92018
	Negatif Binomial	0,00700

Tabel 10.
Hasil uji Kolmogorov Smirnov untuk variabel *y* variabel distribusi kegagalan

Variabel	Distribusi Kegagalan	*p-value*
Besar Kerugian	Eksponensial	0,86090
	Log Normal	0,85820
	Weibull	0,72250

Tabel 11.
Nilai Akaike Information Criterion (AIC) distribusi kegagalan

Variabel	Distribusi Kegagalan	AIC
Besar Kerugian	Eksponensial	28468,37000
	Lognormal	27960,71000
	Weibull	33657,70000

Dengan hipotesis:

H_0 : Data mengikuti distribusi tersebut.

H_1 : Data tidak mengikuti distribusi tersebut.

Dengan daerah tolak H_0 jika, *p-value*<α atau 0,05.

Berdasarkan Tabel 9, nilai *p-value* dari variabel banyaknya komputer lebih besar dari α = 0.05 terdapat pada distribusi Geometri yang mengindikasikan tidak ada bukti untuk menolak H_0. Dengan demikian, banyaknya komputer mengikuti distribusi Geometri.

Berdasarkan Tabel 10, nilai *p-value* dari variabel besarnya kerugian lebih besar dari α = 0.05 terdapat pada distribusi Eksponensial, Lognormal dan Weibull yang mengindikasikan tidak ada bukti untuk menolak H_0. Dengan demikian, besar kerugian mengikuti distribusi Eksponensial, Lognormal, dan Weibull.

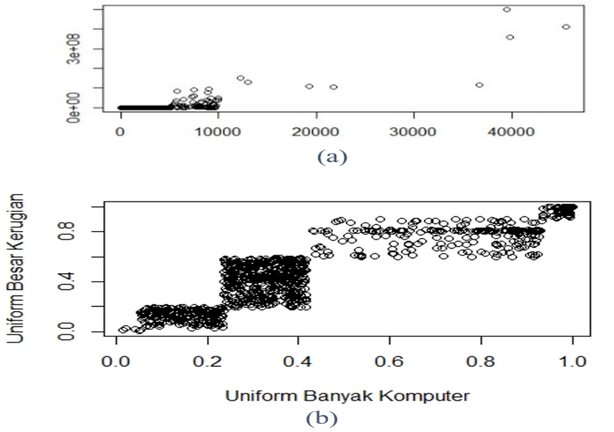
3) Akaike Information Criterion (AIC)

Untuk mencari distribusi yang paling sesuai dengan data maka dipilih distribusi dengan nilai AIC terkecil pada variabel besar kerugian [14]. Hasil perhitungan nilai AIC dirangkum pada Tabel 11.

Pada Tabel 11 yang didapat, distribusi marginal y dipilih Lognormal karena memiliki nilai AIC terkecil yaitu sebesar 27960,71. Dari hasil pemilihan nilai AIC, dapat disimpulkan bahwa bangkitan dengan fungsi copula akan digunakan parameter x ~ Geometri (0,053974) dan y ~ Lognormal (11,04501; 2,13956).

C. Analisis Copula

Pada penelitian ini fungsi copula digunakan sebagai dasar pembangkitan copula dimana pembangkitan data dengan metode Simulasi Monte Carlo menggunakan parameter copula terpilih [20]. Copula memungkinkan penggabungan semua jenis distribusi marginal untuk variabel (x) dan (y).



Gambar 2. (a) *Scatterplot* sebelum transformasi dan (b) *Scatterplot* sesudah tranformasi.

Tabel 12.
Estimasi parameter copula

Copula	Estimasi Parameter	*Log-likehood*
Clayton	$\hat{\theta}_C = 3{,}44$	693,3
Gumbel	$\hat{\theta}_G = 2{,}59$	644,3
Frank	$\hat{\theta}_F = 11{,}42$	759,3

1) Uji Korelasi Tau Kendall

Pola hubungan antara jumlah unit terdampak (x) dan besar kerugian (y) yang didapatkan mampu dijelaskan dengan baik menggunakan korelasi Tau-Kendall [16]. Hasil pengujian korelasi memberikan kesimpulan bahwa jumlah unit terdampak (x) dan besar kerugian (y) memiliki hubungan erat dengan besar kerugian yang diakibatkan berdasarkan *p-value* menunjukkan signifikansi terhadap *p-value* = 0,000.

2) Transformasi ke Uniform [0,1]

Untuk melakukan analisis model copula Archimedean, variabel terlebih dahulu ditransformasi ke Uniform [0,1]. Transformasi dilakukan untuk melihat hubungan dependensi antar variabel antara jumlah unit terdampak (x) dan besar kerugian (y).

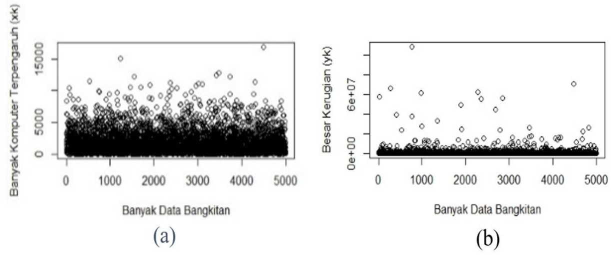
Jika dibandingkan Gambar 2 (a) dibandingkan dengan gambar 2 (b), Gambar 1 (a) kurang dapat menunjukkan struktur dependensi variabel jumlah unit terdampak (x) dan besar kerugian (y). Pada Gambar 2 (b) merupakan *scatterplot* jumlah unit terdampak (x) dan besar kerugian (y). setelah dilakukan transformasi ke Uniform [0,1]. Struktur dependensi memberikan gambaran model copula yang paling sesuai [16].

3) Estimasi Parameter

Estimasi parameter copula dapat menggunakan *Maximum Log-Likelihood* atau MLE dimana, model *copula* terbaik dipilih dengan mempertimbangkan jumlah parameter dalam model. Semakin besar nilai MLE, maka model semakin baik dan layak untuk digunakan.

Estimasi parameter copula dengan estimasi *maximum log-likelihood* untuk melihat model copula manakah yang paling sesuai untuk memodelkan hubungan dependensi antara jumlah unit terdampak (x) dan besar kerugian (y). Parameter *copula* yang dipilih adalah parameter dengan hasil estimasi *maximum log-likelihood* yang paling besar [17].

Dari Tabel 12 hasil Estimasi *copula* yang dilakukan pada *copula* Archimedean; *copula* Clayton, *copula* Gumbel dan *copula* Frank, didapatkan nilai *maximum log-likelihood* terbesar pada Estimasi parameter *copula* sebesar 759,3 dan didapatkan nilai estimasi parameter bernilai 11,42. Dari hasil tersebut menghasilkan kesimpulan model *copula* terbaik umtuk distribusi kerugian siber di Amerika Serikat dari tahun 2004-2020 yaitu *copula* Frank.



Gambar 3. *Scatterplot* (a) data bangkitan (x_k) dan (b) data bangkitan (y_k).

4) Pemodelan Copula Distribusi Kerugian Siber

Dari hasil pemilihan model *copula* terbaik dari nilai MLE, selanjutnya dilakukan pembentukan model *copula* Gumbel untuk distribusi kerugian serangan siber di Amerika Serikat tahun 2004-2020. Persamaan copula Gumbel dituliskan:

$$C_{11,42}^{Fr} = (1 + 11{,}42)(x \cdot y)^{-1-11{,}42} \cdot \left(\frac{1}{11{,}42} ln \left(1 + \frac{(e^{-11{,}42x}-1)(e^{-11{,}42y}-1)}{e^{11{,}42}-1}\right)^{-2}\right)$$

D. Analisis Perhitungan Premi Murni Asuransi Siber

Setelah melakukan analisis copula dan didapatkan parameter dari copula Archimedean terbaik yaitu copula Frank sebesar 11,42 maka dilakukan bangkitan data dengan parameter copula Frank ($\hat{\theta}_F$= 11,42) untuk mendapatkan bangkitan data banyak komputer terpengaruh (x_k), besar kerugian (y_k) dan waktu tunggu kejadian (T) yang selanjutnya digunakan untuk menghitung premi murni asuransi.

1) Pembangkitan Data (x) dan (y) dengan Simulasi Monte Carlo

Pada penelitian ini akan dibangkitkan data (x_k, y_k) dengan Persamaan 7. Data dibangkitkan menggunakan fungsi kepekatan bersama copula terbaik yaitu copula Frank dengan persamaan:

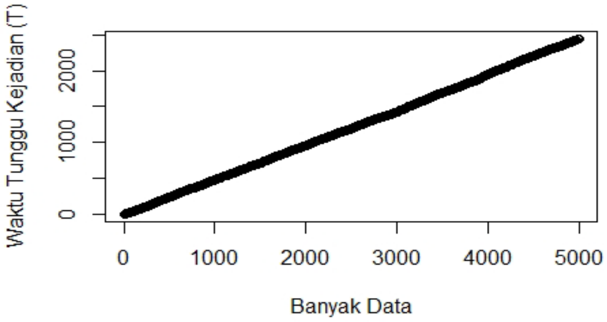
$$f(x, y) = c(u_1, u_2) = (1 + 11{,}42)(x \cdot y)^{-1-11{,}42} \left(\frac{1}{11{,}42} ln \left(1 + \frac{(e^{-11{,}42x}-1)(e^{-11{,}42y}-1)}{e^{11{,}42}-1}\right)^{-2}\right)$$

Dengan (u_1, u_2) adalah data berpasangan (x, y) dan $\hat{\theta}_F$ =11,42 adalah parameter copula Frank. Hasil dari bangkitan data (x_k, y_k) terlihat pada *scatterplot* pada Gambar 3.

Gambar 3 menjelaskan penyebaran data bangkitan banyak komputer terpengaruh (x_k) dan besar kerugian (y_k) sebanyak n=5000 selanjutnya data ini akan menjadi elemen dari total biaya kerugian (L).

2) Pembangkitan Waktu Tunggu Kejadian (T) dengan Proses Poisson

Pada penelitian ini diasumsikan sistem jaringan komputer perusahaan gagal mencegah ketika virus pertama kali menyerang dan klaim hanya akan dibayarkan sekali. Waktu tunggu antara insiden satu dengan yang lainnya adalah $T_i = t_1 - t_i - 1$ yang merupakan variabel acak eksponensial



Gambar 4. *Scatterplot* waktu tunggu kejadian.

dengan nilai harapan $\frac{1}{\lambda}$. Hasil Bangkitan waktu tunggu kejadian sebanyak 5000 kejadian.

Scatterplot dari hasil bangkitan waktu kejadian digambarkan pada Gambar 4.

Sehingga dapat dikatakan bahwa waktu tunggu kejadian serangan siber *ransomware* dalam kurun waktu 15 tahun mengalami peningkatan yang konstan dari satu kejadian ke kejadian lain Selanjutnya pembangkitan data Waktu Tunggu Kejadian (T) digunakan untuk melakukan perhitungan premi.

3) Perhitungan Premi

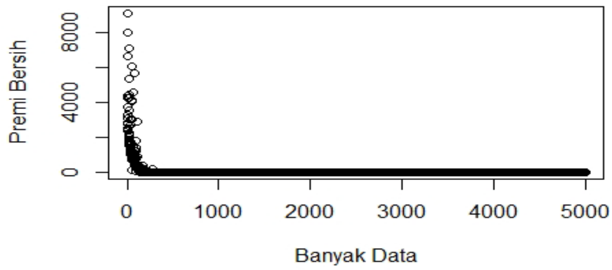
Perhitungan premi dilakukan dengan menghitung total kerugian yang didapatkan dari Hasil bangkitan simulasi Monte Carlo Banyak Komputer Terpengaruh (x_k) dan Besar Kerugian (y_k), *rate* premi sebesar 6% dan hasil bangkitan proses poisson Waktu Tunggu Kejadian (T) dengan Persamaan10. *Scatterplot* dari sebaran harga premi asuransi siber digambarkan pada Gambar 5.

Didapatkan rata-rata harga premi sebesar \$41,96 dengan minimum \$0 dan maksimum \$9.099,53. Dengan total premi sebesar \$209.770,48. Dari data premi murni dengan mengabaikan *outlier* dihasilkan nilai varians yang lebih kecil dari sebelumnya varians premi murni asuransi siber sebesar 5.637.429,86 kini dihasilkan nilai yang lebih kecil yaitu sebesar 137.722,50.

V. KESIMPULAN

Hasil dari pemodelan banyaknya komputer dan besarnya kerugian menunjukkan bahwa distribusi marginal dari masing-masing variabel bukan normal. Oleh sebab itu metode copula lebih tepat digunakan untuk memodelkan dependensi distribusi frekuensi banyak komputer terpilih yaitu distribusi x ~ Geometri (0,053974) dan distribusi kegagalan, distribusi y ~ Lognormal (11,04501; 2,13956) yang selanjutnya digunakan untuk mencari hubungan dependensi dengan copula Archimedean terbaik, yaitu copula Frank dan didapatkan parameter sebesar $\hat{\boldsymbol{\theta}}_{\boldsymbol{F}}$ =11,42.

Perhitungan premi asuransi siber dilakukan metode Black Scholes dengan persamaan $P = L.e^{-rT}$ dari bangkitan data banyak komputer terpengaruh berbasis copula terpilih, yaitu copula Frank dan besar kerugian dalam dollar berbasis copula Frank. Didapatkan premi murni asuransi siber sebesar \$0 hingga \$165.870 untuk tiap perusahaan.



Gambar 5. Scatterplot hasil perhitungan premi murni asuransi siber.

DAFTAR PUSTAKA

[1] K. Awiszus, T. Knispel, I. Penner, G. Svindland, A. Voß, and S. Weber, *Modeling and Pricing Cyber Insurance*. Germany: Leibniz Universität Hannover, 2021.

[2] R. S. Betterley, *The Betterley Report - Cyber and Privacy Insurance*. Dallas: International Risk Management Institute, Inc, 2018.

[3] R. B. Nelsen, *An Introduction to Copulas*. New York: Springer, 2006. ISBN: 978-0387286594.

[4] R. Ko, E. Tsen, and S. Slapnicar, "Dataset of data breaches and ransomware attacks over 15 years from 2004," *J. UQ Espac.*, 2020, doi: 10.14264/e4c8c65.

[5] K. Dharmawan, W. Widia, and L. P. E. Yuni, "Penerapan Metode Penilaian Kontrak Opsi Dalam Penentuan Nilai Premi Asuransi Pertanian Berbasis Indeks Curah Hujan," in *Prosiding Konferensi Nasional Matematika XVIII*, Pekanbaru, 2017, pp. 4–9.

[6] Casualty Actuarial Society, "Insurance Rating Variables: What They Are and Why They Matter," in *National Council of Insurance Legislators (NCOIL)*,Texas, 2019, pp. 1-12.

[7] I. Jeremy, "Pemodelan dan Perhitungan Premi Asuransi Keamanan Siber dengan Model Markov dan Model Non-Markov," Departemen Matematika, Universitas Katolik Parahyangan, 2021.

[8] S. I. Abramowitz, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1972. ISBN: 978-0486612720.

[9] G. Casella and R. L. Berger, *Statistical Inference, 2nd Edition*. Massachusetts: Cengage Learning, 2021. ISBN: 978-0534243128.

[10] R. V. Hogg, E. A. Tanis, and D. L. Zimmerman, *Probability and Statistical Inference, 9th Edition*. New Jersey: Pearson, 2015. ISBN: 978-0-321-92327-1.

[11] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. London: Chapman & Hall, 2020. ISBN: 9781138087446.

[12] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. New Jersey: Pearson, 2017. ISBN: 978-0131877153.

[13] O. Thorin, "On the infinite divisibility of the lognormal distribution," *Scand. Actuar. J.*, vol. 1977, pp. 121–148, 1977, doi: 10.1080/03461238.1977.10405635.

[14] R. E. Walpole, *Pengantar Statistik*. Jakarta: Gramedia, 1995.

[15] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*. New Jersey: John Wiley & Sons Inc, 2019. ISBN: 9781118315323.

[16] C. M. Cuadras, J. Fortiana, and J. A. Rodriguez-Lallena, *Distributions With Given Marginals and Statistical Modelling*. Berlin: Springer-Science+Business Media, 2002. ISBN: 978-1-4020-0914-3.

[17] C. Genest, J. Nešlehová, and J. Ziegel, "Inference in multivariate Archimedean copula models," *TEST*, vol. 20, pp. 223–256, 2011, doi: 10.1007/s11749-011-0250-6.

[18] D. N. Gujarati, *Basic Econometrics Fourth Edition*. New York: The McGraw-Hill Companies, 2004. ISBN: 978-0070597938.

[19] N. T. Thomopoulos, *Essentials of Monte Carlo Simulation: Statistical Methods for Building Simulation Models*. New York: Springer, 2013. ISBN: 978-1461460213.

[20] H. S. B. Herath and T. C. Herath, "Copula-based actuarial model for pricing cyber-insurance policies," *Insur. Mark. Co. Anal. Actuar. Comput.*, vol. 2, no. 1, pp. 7–20, 2011.