Jur. Ris. & Apl. Mat. 2 (2018), no. 1, 1-64
Jurnal Riset dan Aplikasi Matematika
e-ISSN: 2581-0154
URL: journal.unesa.ac.id/index.php/jram

ANALISIS PENGELOMPOKAN *K-MEANS* UNTUK DATA BIVARIAT LAJU KUNJUNGAN DAN RASIO RUJUKAN

(Studi Data pada Seluruh Fasilitas Kesehatan Tingkat I di BPJS Kesehatan Surakarta)

Oktaviana Reni Kristyaningrum[1], Adi Setiawan[2], Leopoldus Ricky Sasongko[3]

Program Studi Matematika, Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana

email: 662014015@student.uksw.edu[1], adi_setia_03@yahoo.com[2], leopoldus.sasongko@staff.uksw.edu[3]

ABSTRAK

Pengelompokkan fasilitas kesehatan (faskes) tingkat I berdasarkan tinggi atau rendah laju kunjungan dan rasio rujukan perlu dilakukan. Hal ini berguna bagi BPJS Kesehatan untuk menentukan arah kebijakannya maupun pemerintah Indonesia dalam mensejahterakan kehidupan masyarakat Indonesia. Metode pengelompokan yang diterapkan pada data laju kunjungan dan rasio rujukan di 356 faskes tingkat I BPJS Kesehatan Surakarta adalah metode pengelompokkan *k-means*. Pengelompokkan *k-means* yang telah dilakukan pada data menghasilkan empat kelompok dengan tingkat laju kunjungan dan rasio rujukannya yaitu kelompok 1 dengan tingkat rendah-rendah, kelompok 2 dengan tinggi-rendah, kelompok 3 dengan tingkat rendah-tinggi, dan kelompok 4 dengan tingkat rendah-sedang yang mana secara berurutan masing-masing kelompok beranggotakan 164, 129, 20, dan 43 faskes tingkat I. Setelah data dikelompokan, analisis hubungan antara data laju kunjungan dan rasio rujukan menjadi perhatian dalam paper ini. Analisis hubungan didasarkan pada data di tiap kelompok hasil pengelompokkan sehingga selanjutnya dapat diketahui bagaimana keterhubungan data laju kunjungan dan ratio rujukan faskes-faskes tingkat I di tiap kelompok. Analisis hubungan dilakukan dengan menggunakan pendekatan copula (bivariat). Ukuran keterhubungan yang digunakan untuk mengetahui keterhubungan data dalah Kendall's *Tau* dan Spearman's *Rho*. Copula yang terbaik untuk menjelaskan keterhubungan data didasarkan pada *p-value* tertinggi dari uji kecocokan statistik Cramér-von Mises yang diperoleh melalui simulasi *parametric bootstrap*. Paper ini memberikan hasil bahwa dua kelompok (kelompok 1 dan kelompok 4) yang mana keterhubungan data di dalamnya dapat digambarkan melalui copula, sedangkan dua kelompok yang lain tidak dapat digambarkan dari lima copula yang diujikan dalam paper ini. Kelompok 1 digambarkan melalui copula Gumbel dengan marginal-marginalnya adalah Normal-Normal, sedangkan kelompok 4 digambarkan melalui copula Ali-Mikhail-Haq (AMH) dengan marginal-marginalnya Weibull-Gamma.

Kata Kunci: *Faskes tingkat I BPJS Kesehatan,Laju Kunjungan, Rasio Rujukan, Pengelompokan K-means,Copula*

1. Pendahuluan

Pentingnya asuransi kesehatan telah mendorong pemerintah Indonesia untuk membentuk Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan, suatu badan hukum publik yang memiliki tugas untuk menyelenggarakan Jaminan Kesehatan Nasional (JKN) bagi rakyat Indonesia. Dalam penyelenggaraannya, BPJS Kesehatan menggunakan sistem rujukan berjenjang kepada pesertanya. Setiap peserta BPJS Kesehatan terdaftar di tepat satu fasilitas kesehatan (faskes) tingkat I, tempat pertama yang harus peserta kunjungi sewaktu peserta

Tanggal Masuk: 24-04-18; Revisi: 28-04-18; Diterima: 30-04-2018

memerlukan pelayanan kesehatan. Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Kota Surakarta berfungsi sebagai penyelenggara program jaminan kesehatan di wilayah Surakarta, yang mana meliputi Kota Surakarta, Kabupaten (Kab.) Karanganyar, Kab. Sragen, Kab. Sukoharjo dan Kab. Wonogiri. Di wilayah tersebut terdapat 132 dokter praktik perorangan, 58 klinik pratama, 109 puskesmas, 43 dokter gigi, 5 klinik POLRI, dan 9 klinik TNI yang telah bekerjasama dengan BPJS Kesehatan sebagai faskes tingkat I. Banyak peserta yang berkunjung tiap bulan di tiap faskes tingkat I tersebut tercatat. Dari semua peserta yang berkunjung tersebut, terdapat peserta yang dirujuk ke faskes tingkat selanjutnya. Banyak peserta yang dirujuk tersebut juga tercatat. Berdasarkan hal-hal tersebut diperoleh data mengenai laju kunjungan dan rasio rujukan per bulan untuk tiap faskes tingkat I. Laju kunjungan diperoleh dari banyaknya kunjungan pada bulan itu dikali 1000 jiwa, lalu dibagi dengan jumlah peserta yang terdaftar pada suatu faskes tingkat I. Sedangkan rasio rujukan diperoleh dari jumlah peserta yang dirujuk dikali 1000 jiwa, lalu dibagi jumlah peserta yang terdaftar pada suatu faskes tingkat I.

Informasi mengenai seberapa tinggi atau rendah laju kunjungan dan rasio rujukan di suatu faskes tingkat I merupakan hal yang penting untuk diperhatikan. Informasi tersebut dapat bermanfaat bagi BPJS Kesehatan guna menentukan arah kebijakannya pada waktu mendatang. Selain itu, informasi tersebut penting bagi pemerintah Indonesia dalam mengambil kebijakan dalam rangka mensejahterakan kehidupan masyarakat Indonesia seperti pemberian penyuluhan, konsultasi kesehatan, atau pengobatan gratis bagi masyarakat (terutama peserta) di faskes tingkat I yang memiliki laju kunjungan atau rasio rujukan pada tingkat tertentu. Untuk memperoleh informasi tersebut, pengelompokan (*clustering*) data mengenai seberapa tinggi atau rendah laju kunjungan dan rasio rujukan faskes-faskes tingkat I perlu dilakukan. Pengelompokan data dalam penelitian ini menggunakan *clustering* dengan metode *k-means* dimana faskes-faskes tingkat I di BPJS Kesehatan Surakarta dikelompokan menjadi *k* kelompok sesuai dengan karakteristik tingkat laju kunjungan dan rasio rujukan. Penelitian mengenai pengelompokan data antara lain Dhuhita [1] melakukan analisis *clustering* dengan metode *k-means* untuk menentukan status gizi balita dan Lesnussa [2] yang juga melakukan analisis *clustering* dengan metode *k-means* pada data Indeks Pembangunan Manusia (IPM) tahun 2004-2010 provinsi Maluku Utara.

Setelah data dikelompokan, analisis hubungan antara laju kunjungan dan rasio rujukan menjadi perhatian penelitian. Analisis hubungan didasarkan pada data tiap kelompok hasil analisis *clustering* sehingga selanjutnya dapat diketahui bagaimana hubungan laju kunjungan dan ratio rujukan faskes-faskes tingkat di suatu kelompok. Analisis hubungan tersebut dilakukan dengan menggunakan pendekatan copula dimana copula dapat menggambarkan keterhubungan dua peubah secara umum (linier maupun nonlinier) melalui distribusi bivariatnya. Penelitian sebelumnya mengenai analisis hubungan dua peubah menggunakan copula antara lain Darwis [3] yang melakukan analisis hubungan Indeks Harga Saham Gabungan (IHSG) dengan faktor makroekonomi melalui pendekatan copula. Sedangkan penelitian tentang keterkaitan pengelompokan data dan copula belum ada.

Tujuan dari penelitian ini adalah untuk memperoleh kelompok-kelompok faskes tingkat I BPJS Kesehatan Surakarta berdasarkan tinggi/rendah laju kunjungan dan rasio rujukan. Setelah itu, memperoleh model copula yang cocok untuk menjelaskan hubungan antara data laju kunjungan dan rasio rujukan tiap kelompok faskes tingkat I BPJS Kesehatan Surakarta yang telah diperoleh.

2. Dasar Teori dan Metode Penelitian

2.1. *K-Means Clustering*

K-Means Clustering merupakan salah satu cara pengelompokan data menggunakan metode non-hierarki yang berbasis jarak (Lesnussa [2]).

Pengelompokan *K-Means* untuk Data Bivariat

Misalkan diketahui sebanyak n data bivariat (titik di dua dimensi) yaitu $X_1, X_2, X_3, \dots, X_n$ dengan $X_i = (x_{i1}, x_{i2})$, $\forall i = 1, 2, 3, \dots, n$. Misalkan pula terdapat k kelompok awal yaitu $C_1, C_2, C_3, \dots, C_k$ dengan titik-titik pusat tiap kelompok tersebut berurutan adalah $\mu_1, \mu_2, \mu_3, \dots, \mu_k$ dengan $\mu_j = (\mu_{j1}, \mu_{j2})$, $\forall j = 1, 2, 3, \dots, k$. Pengelompokan dengan metode *k-means* untuk data bivariat tersebut dilakukan dengan meminimumkan fungsi tujuan

$$D = \sum_{j=1}^{k} \sum_{X_i \in C_j} d(X_i, \mu_j) \tag{1}$$

dengan $d(X_i, \mu_j)$ adalah jarak setiap titik terhadap titik pusat kelompoknya yang dapat dicari menggunakan metode Euclidean. Menurut Lesnussa [2], Jarak Euclidean didefinisikan sebagai akar dari jumlahan kuadrat perbedaan/deviasi antar elemen pada titik yang dinyatakan oleh

$$d(X_i, \mu_j) = \sqrt{\sum_{p=1}^{2} (X_{ik} - \mu_{ik})^2} \ . \tag{2}$$

Untuk Algoritma pengelompokan *k-means* dapat dijabarkan melalui *pseudocode* berikut:

1. Inisialisasi $X_1, X_2, X_3, \dots, X_n$; k;dan $\mu_1, \mu_2, \mu_3, \dots, \mu_k$,
2. Ulangi hingga D konvergen:

   2.a.tentukan tiap titik X_i ke kelompok dengan jarak X_i terhadap μ_j adalah terdekat, sehingga diperoleh $C_1, C_2, C_3, \dots, C_k$,

   2.b. hitung D seperti pada (1),

   2.c.tentukan titik pusat yang baru tiap kelompok yaitu

   $$\mu_j = \frac{1}{|C_j|} \sum_{x \in C_j} X_i$$

   dengan $|C_j|$ menyatakan banyaknya titik pada kelompok ke-j.

2.2. *Goodness of Fit Test*

Goodness of fit test digunakan untuk menjelaskan seberapa baik data empiris mengikuti atau merupakan sampel acak dari suatu distribusi tertentu. Langkah pertama dari *goodness of fit* adalah mengestimasi parameter suatu distribusi berdasarkan data empiris. Estimasi parameter ini menggunakan metode *maximum likelihood estimation.* Selanjutnya dilakukan uji kecocokan distribusi melalui metode teoritis tertentu salah satunya adalah metode Kolmogorov-Smirnov. Dalam paper ini, *goodness of fit test* digunakan untuk mengestimasi distribusi marginal data laju kunjungan dan rasio rujukan sebelum dilakukan uji kecocokan copula.

2.3. Copula

Menurut Sasangko [4], Copula (bivariat) adalah suatu fungsi distribusi bivariat dengan marginal-marginalnya berdistribusi seragam [0,1]. Copula C dinyatakan oleh

$$C(u,v) = Pr[U \leq u, V \leq v]$$

Untuk peubah acak U dan V berdistribusi seragam di [0,1]. Misal peubah-peubah acak baru $U = F(X)$ dan $V = G(Y)$. Menurut Teorema Sklar, jika H adalah fungsi distribusi bivariat dengan fungsi-fungsi distribusi marginal F dan G, maka terdapat suatu copula C untuk semua (x,y) sedemikian hingga

$$H(x,y) = C\big(F(x), G(y)\big)$$

dengan $F(x)$ dan $G(y)$ merupakan peubah acak baru.

Kendall's *Tau*

Misalkan $\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$ menunjukkan sampel acak n pengamatan dari sebuah vector (X,Y) dari variabel acak kontinu. Terdapat $\binom{n}{2}$ pasangan yang berbeda (x_i,y_i) dan $\big(x_j,y_j\big)$, dan masing-masing pasangan *concordant* atau *discordant*, maka Kendall's *tau* didefinisikan oleh

$$\tau = \frac{c-d}{c+d} = (c-d)/\binom{n}{2}$$

dengan c menunjukkan jumlah pasangan yang *concordant* dan d menunjukkan jumlah pasangan yang *discordant.*

Spearman's Rho

Misalkan X dan Y variabel acak kontinu menurut Nelsen [5] *Spearman's Rho* untuk X dan Y diberikan oleh

$$\rho_s = 1 - \frac{6\sum d_1^2}{n(n^2-1)}$$

dengan d_i merupakan perbedaan ranking dari masing-masing pengamatan dan n merupakan jumlah pengamatan.

Pembangkit Sampel Acak Bivariat Copula (Sasongko [4])

Prosedur untuk membangkitkan sampel acak bivariat $\{(x,y)\}$ dari suatu fungsi distribusi bivariat H menggunakan copula terlebih dahulu pandang fungsi $\frac{\partial C(u,v)}{\partial u}$ adalah fungsi dalam v, misal $u = F(x)$ dan $v = G(y)$,

$$z_u(v) = \frac{\partial C(u,v)}{\partial u}.$$

Prosedur pembangkitan sampel acak bivariat $\{(x,y)\}$ melalui langkah:

- Membangkitkan dua bilangan acak saling bebas u dan t, distribusi seragam di $[0,1]$;
- Memperoleh $v = {z_u}^{(-1)}(t)$, dimana ${z_u}^{(-1)}$ adalah invers fungsi dari z_u;
- Memperoleh sepasang bilangan acak bivariat dari suatu Copula yaitu (u,v);
- Memperoleh sepasang bilangan acak bivariat $(x,y) = \big(F^{-1}(u), G^{-1}(v)\big)$;

***Goodness of Fit for* Copula (Genest dkk [6])**

Uji kecocokaan Copula diperlukan untuk mengetahui seberapa cocok copula dapat memodelkan data. Uji dilakukan berdasarkan proses empirik pada fungsi distribusi bivariat ataupun Copula untuk parameter $\hat{\varphi}$ dinyatakan oleh

$$\mathbb{e}(x_i, y_i) = \sqrt{n}\left[H_e(x_i, y_i) - H_{\hat{\varphi}}(x_i, y_i)\right] = \sqrt{n}\left[C_e\big(F(x_i), G(y_i)\big) - C_{\hat{\varphi}}\big(F(x_i), G(y_i)\big)\right]$$

dengan $H_e(x_i, y_i) = C_e\big(F(x_i), G(y_i)\big) = \frac{\#(x \leq x_i, y \leq y_i)}{n+1}$ fungsi distribusi bivariat empiris atau copula empiris untuk data $\{(x_i, y_i)\}$, $i = 1,2, \dots, n$, fungsi $\#(x \leq x_i, y \leq y_i)$ menyatakan banyaknya data bivariat $\{(x, y)\}$ dengan $x \leq x_i$ dan $y \leq y_i$. Ukuran statistik yang digunakan adalah statistik Cram$\acute{e}$r-von Mises (S_n), yang diperoleh dari

$$S_n = \frac{1}{n}\sum_{i=1}^{n}[\mathbb{e}(x_i, y_i)]^2 = \sum_{i=1}^{n}\left[H_e(x_i, y_i) - H_{\hat{\varphi}}(x_i, y_i)\right]^2$$
$$= \sum_{i=1}^{n}\left[C_e\big(F(x_i), G(y_i)\big) - C_{\hat{\varphi}}\big(F(x_i), G(y_i)\big)\right]^2. \quad (3)$$

Kecocokan data terhadap suatu fungsi distribusi bivariat atau copula bergantung pada nilai S_n terkecil dari beberapa fungsi distribusi bivariat atau copula yang dicocokkan. Setelah diperoleh nilai S_n untuk setiap kelompok, dilakukan uji hipotesis menggunakan *parametric bootstrap*. H_0 ditolak apabila *p-value* kurang dari nilai signifikansi tertentu (0.1, 0.05 atau 0.001).

Misal diketahui data bivariat sebanyak *n* pasang yaitu $\{(x_a, y_a)\}, a = 0,1,2,3, \dots, n$. Untuk *N* bilangan bulat positif sangat besar, algoritma *parametric bootstrap* adalah sebagai berikut:

1. Membangkitkan *n* sampel acak bivariat $\{(x_i, y_i)\}$ dengan $i = 0,1,2, \dots, n$ dari suatu distribusi bivariat $H_\theta(x, y)$atau Copula $C_\theta\big(F(x), G(y)\big)$,
2. Menghitung $H_e(x_i, y_i) = C_e\big(F(x_i), G(y_i)\big) = \frac{\#(x_a \leq x_i, y_a \leq y_i)}{n+1}$ dengan $\#(x_a \leq x_i, y_a \leq y_i)$ adalah banyak data bivariat $\{(x_a, y_a)\}$ dengan $x_a \leq x_i$ dan $y_a \leq y_i$,
3. Jika $j = 1$, hitung
$$s_{n,j}^* = \sum\nolimits_{i=1}^{n}[H_e(x_i, y_i) - H_\theta(x_i, y_i)]^2$$
$$= \sum\nolimits_{i=1}^{n}\left[C_e\big(F(x_i), G(y_i)\big) - C_\theta\big(F(x_i), G(y_i)\big)\right]^2$$
4. Jika $j = j + 1$, ulangi poin 1 sampai poin 3
5. Jika $j = N + 1$, hitung *p-value*, $\frac{\#\left(s_{n,j}^* > s_n\right)}{N}$ atau $\sum_{j=1}^{N}\left(\frac{I\left(s_{n,j}^* > s_n\right)}{N}\right)$, yang mana $I\left(s_{n,j}^* > s_n\right)$ adalah fungsi bernilai 1 untuk nilai $s_{n,j}^* > s_n$.

2.4. Copula Archimedean

Menurut Nelsen [5], copula Archimedean memiliki kekhasan bahwa suatu copula yang memiliki satu parameter kebergantungan (θ) dapat dibentuk dari suatu fungsi pembangkit copula φ. Copula Archimedean untuk peubah acak bivariat dapat dituliskan sebagai

$$\varphi\big(C(u, v)\big) = \varphi(u) + \varphi(v).$$

Misalkan X dan Y adalah variabel acak dengan Copula Archimedean C yang dihasilkan oleh φ pada W. Estimasi parameter θ dari Copula Archimedean diperoleh dari menyelesaikan solusi berikut

$$\tau_C = 1 + 4\int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)}dt. \tag{4}$$

Copula Clayton

Bentuk umum dari copula Clayton:

$$C_{C,\theta}(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$$

dengan $\theta \in (0,\infty)$. Fungsi generator copula Clayton yaitu

$$\varphi_\theta(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right). \tag{5}$$

Dengan menyelesaikan persamaan (4) dan (5) diperoleh parameter θ untuk copula Clayton:

$$\theta = \frac{2\tau}{1-\tau}. \tag{6}$$

Copula Gumbel

Bentuk umum dari copula Gumbel:

$$C_{G,\theta}(u,v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right)$$

dengan $\theta \in [1,\infty)$. Fungsi generator copula Gumbel yaitu

$$\varphi_\theta(t) = (-\ln t)^\theta. \tag{7}$$

Dengan menyelesaikan persamaan (4) dan (7) diperoleh parameter θ untuk copula Gumbel:

$$\theta = \frac{1}{1-\tau}. \tag{8}$$

Copula Ali-Mikhail-Haq (AMH)

Bentuk umum dari copula AMH:

$$C_{A,\theta} = \frac{uv}{1-\theta(1-u)(1-v)}$$

dengan $\theta \in [-1,1)$. Fungsi generator copula AMH adalah

$$\varphi_\theta(t) = ln\left[\frac{1-\theta(1-t)}{t}\right]. \tag{9}$$

Dengan menyelesaikan persamaan (4) dan (9) diperoleh parameter θ untuk copula AMH:

$$\tau = \frac{3\theta - 2}{3\theta} - \frac{2(1-\theta)^2 ln(1-\theta)}{3\theta^2}. \tag{10}$$

Copula Frank

Bentuk umum dari copula Frank:

$$C_{F,\theta}(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right)$$

dengan $\theta \in \mathbb{R}\backslash\{0\}$. Fungsi generator copula Frank adalah

$$\varphi_\theta(t) = -\ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right). \tag{11}$$

Dengan menyelesaikan persamaan (4) dan (11) diperoleh parameter θ untuk copula Frank:

$$\tau = 1 - \frac{4\left(1 - \theta^{-1}\int_0^\theta \frac{t}{e^t - 1}dt\right)}{\theta}. \tag{12}$$

2.5. Copula Plackett

Copula Plackett mempunyai bentuk umum sebagai berikut:

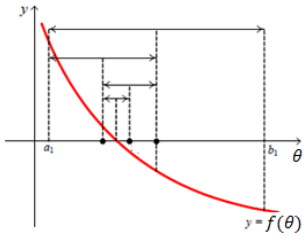
$$C_{\mathrm{P},\theta}(u,v)=\frac{[1+(\theta-1)(u+v)]}{2(\theta-1)}-\frac{\sqrt{[1+(\theta-1)(u+v)]^2-4uv\theta(\theta-1)}}{2(\theta-1)}$$

untuk $0<\theta<\infty$ dan $\theta\neq 1$. Saat $\theta=1$, $C_{\mathrm{P},1}(u,v)=uv$. Estimasi parameter θ dapat dilakukan melalui persamaan Spearman's *rho* dengan mencari solusi dari

$$\rho=\frac{\theta+1}{\theta-1}-\frac{2\theta}{(\theta-1)^2}ln(\theta)\,. \tag{13}$$

2.6. Metode Bagi Dua

Metode bagi dua digunakan untuk mencari akar persamaan dari suatu fungsi. Diasumsikan fungsi $f(\theta)$ kontinu pada interval $[a_1,b_1]$ dan $f(a_1)\cdot f(b_1)<0$ sehingga terdapat minimal satu akar pada interval tersebut. Ilustrasi dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi grafis untuk akar hampiran dalam metode bagi dua

Menurut Nugroho [7], Algoritma untuk metode bagi dua sebagai berikut:

1. Menghitung $\theta_n=\frac{a_n+b_n}{2}$.
2. Menentukan subinterval mana yang akan mengurung akar:
   a. Jika $f(a_1)\cdot f(b_1)<0$, maka $a_{n+1}=a_n$, $b_{n+1}=\theta_n$.
   b. Jika $f(a_1)\cdot f(b_1)>0$, maka $a_{n+1}=\theta_n$, $b_{n+1}=b_n$.
   c. Jika $f(a_1)\cdot f(b_1)=0$, maka diperoleh akar sama dengan θ_n.
3. Menghitung $\varepsilon_n=\frac{\theta_n-\theta_{n-1}}{\theta_n}\times 100\%$ dimana $n\geq 1$
4. Mengulangi langkah 1 hingga langkah 3 sedemikian sehingga $f(\theta_n)\approx 0$.
5. Mendapatkan akar persamaan yaitu θ_n.

Dalam penelitian ini, metode bagi dua digunakan untuk mencari parameter θ pada copula AMH, Frank, dan Plackett.

2.7. Metode Penelitian

Dilakukan pengolahan data yang dimiliki melalui langkah-langkah berikut:

1. Melakukan algoritma *k-means clustering* untuk memperoleh kelompok-kelompok data dimana nilai *D* pada persamaan (1) minimum yaitu dengan langkah-langkah:
   a. Menentukan kelompok awal berdasarkan median dari data.
   b. Menentukan titik pusat awal kelompok-kelompok berdasarkan median dari setiap kelompok.
   c. Melakukan algoritma *k-means clustering*
   d. Memperoleh dan menginterpretasikan kelompok-kelompok data yg telah diperoleh.
2. Estimasi distribusi data tiap kelompok melalui pendekatan copula yang meliputi
   a. Estimasi distribusi marginal data melalui:
      - estimasi parameter distribusi data menggunakan metode *maximum likelihood estimation*,

- uji kecocokan distribusi menggunakan uji Kolmogorov-Smirnov.

b. Estimasi Kendall's *tau* dan *Spearman's rho.*

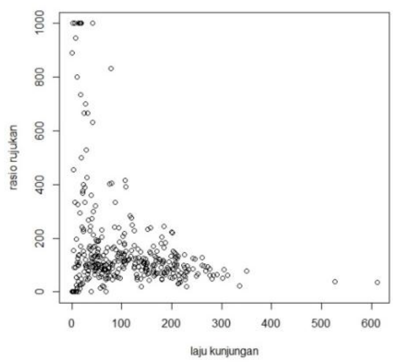
c. Estimasi parameter copula Archimedean berdasarkan Kendall's *tau* dan copula Plackett berdasarkaan *Spearman's rho* dimana:
   - estimasi parameter θ dari copula Clayton berdasarkan persamaan (6),
   - estimasi parameter θ dari copula Gumbel berdasarkan persamaan (8),
   - estimasi parameter θ dari copula Ali-Mikhail-Haq berdasarkan persamaan (10) dan mencari solusi dengan menggunakan metode bagi dua,
   - estimasi parameter θ dari copula Frank berdasarkan persamaan (12) dan mencari solusi dengan menggunakan metode bagi dua,
   - estimasi parameter θ dari copula Plackett berdasarkan persamaan (13) dan mencari solusi dengan menggunakan metode bagi dua.

d. Uji kecocokan copula menggunakan uji statistik Cramér-von Mises seperti pada persamaan (3) dengan bantuan *parametric bootstrap*.

3. Hasil dan Pembahasan

3. 1. Data

Data yang digunakan dalam penelitian ini terdiri dari data laju kunjungan dan rasio rujukan bulan April 2017 di faskes-faskes tingkat I yang tercatat di BPJS Kesehatan Surakarta. *Scatterplot* dari data dapat dilihat pada Gambar 2.



Gambar 2. *Scaterplot* Data Laju Kunjungan dan Rasio Rujukan

3. 2. Kelompok Data

Pengelompokan data dalam penelitian ini menggunakan metode *k-means clustering,* data akan dikelompokan menjadi empat kelompok. Untuk langkah pengerjaan dan hasil adalah sebagai berikut:

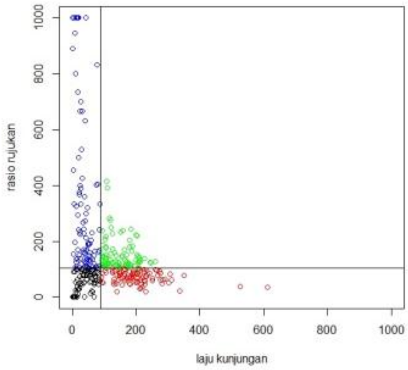
Menentukan Kelompok Awal dan Titik Pusat Awal Kelompok

Langkah pertama untuk menentukan kelompok awal terlebih dahulu dicari nilai median dari data. Nilai median berguna untuk membagi data menjadi 4 kelompok yaitu:

- data dengan laju kunjungan dan rasio rujukan di bawah nilai median,
- data dengan laju kunjungan di atas nilai median sedangkan rasio rujukan berada di bawah nilai median,
- data dengan laju kunjungan di bawah nilai median sedangkan rasio rujukan berada di atas nilai median,

- dan data dengan laju kunjungan dan rasio rujukan di atas nilai median.

Dari data, diperoleh nilai median (88.13967, 105.0509). Dari median yang telah diperoleh terbentuk kelompok awal yang dapat dilihat pada Gambar 3.



Gambar 3. Kelompok Awal Faskes Tingkat I

Penentuan titik pusat awal kelompok berdasarkan median dari masing-masing kelompok awal yang terbentuk. Diperoleh titik pusat awal kelompok seperti pada Tabel 1.

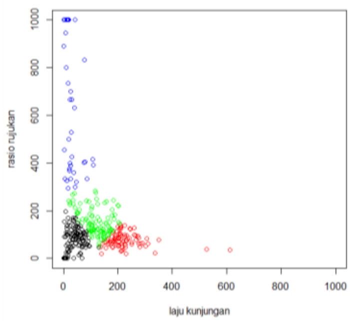
Tabel 1. Titik Pusat Awal Kelompok

	μ_1	μ_2	μ_3	μ_4
Laju kunjungan (x_i)	43.185	191.238	35.649	143.865
Rasio Rujukan (x_j)	66.667	72.072	228.045	133.333

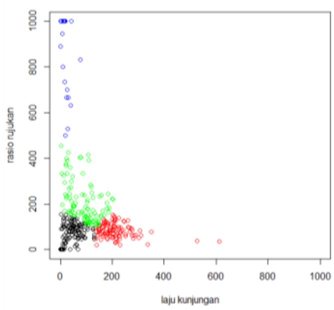
Menentukan tiap titik X_i ke kelompok dengan jarak X_i terhadap μ_j adalah terdekat, sehingga diperoleh C_1, C_2, C_3, C_4.

Setelah diperoleh titik pusat awal kelompok $(\mu_1, \mu_2, \mu_3, \mu_4)$, selanjutnya menghitung jarak setiap titik X_i terhadap μ_j. Dalam penelitian ini perhitungan jarak menggunakan metode Euclidean seperti pada persamaan (2). Kelompok yang baru dipilih berdasarkan jarak terdekat dari titik data terhadap titik pusat kelompok.

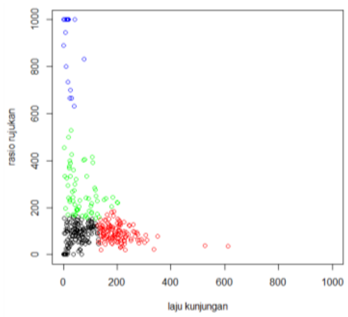
Langkah-langkah tersebut bersifat iteratif hingga diperoleh nilai *D* pada persamaan (1) yang minimum atau dijumpai titik-titik pusat kelompok yang sama pada iterasi selanjutnya.Gambar 4 sampai dengan Gambar 7 menujukan beberapa iterasi yang dilakukan.



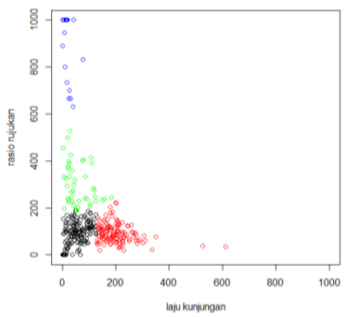
Gambar 4. *Scatterplot* Hasil Akhir Kelompok pada Iterasi 1



Gambar 5. *Scatterplot* Hasil Akhir Kelompok pada Iterasi 3



Gambar 6. *Scatterplot* Hasil Akhir Kelompok pada Iterasi 5



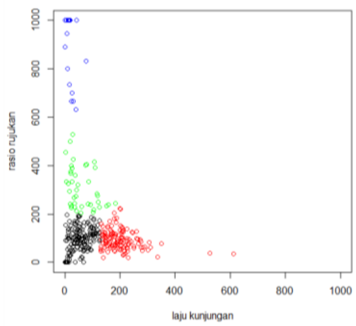
Gambar 7 *Scatterplot* Hasil Akhir Kelompok pada Iterasi 7

Iterasi berhenti pada iterasi ke 11 karena dijumpai titik pusat kelompok yang sama seperti iterasi sebelumnya (iterasi ke 10). Diperoleh titik pusat kelompok terakhir seperti Tabel 2.

Tabel 2. Titik Pusat Kelompok Terakhir

	μ_1	μ_2	μ_3	μ_4
x_1	56.02486	201.66174	20.01157	63.37815
x_2	95.49006	93.45363	893.24561	307.73572

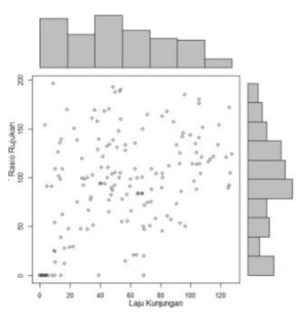
Dengan nilai fungsi tujuan $D = 22710.42$. Scatterplot hasil akhir kelompok dapat dilihat pada Gambar 8.

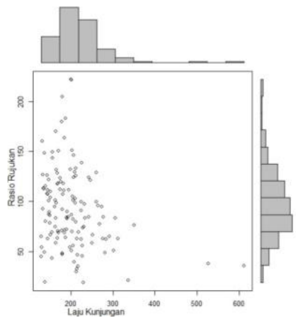


Gambar 8. *Scatterplot* Hasil Akhir Kelompok

3. 3. Model Copula untuk Keterhubungan tiap Kelompok

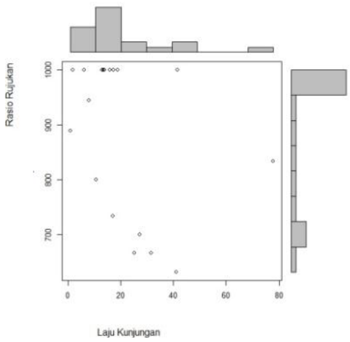
Dari hasil pengelompokan data menggunakan metode *k-means clustering* diperoleh 4 kelompok dimana *scatterplot* dan histogram dari ke empat kelompok tersebut dapat dilihat pada Gambar 9 sampai dengan Gambar 12.



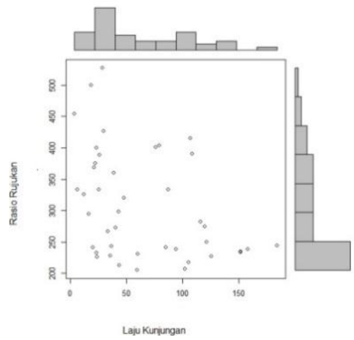


Gambar 9. *Scatterplot* dan Histogram Data Kelompok 1

Gambar 10. *Scatterplot* dan Histogram Data Kelompok 2



Gambar 11. *Scatterplot* dan Histogram Data Kelompok 3



Gambar 12. *Scatterplot* dan Histogram Data Kelompok 4

Estimasi Distribusi Marginal Menggunakan Uji *Kolmogorov-Smirnov*

Dalam penelitian ini, penentuan distribusi marginal data menggunakan bantuan program *Easyfit.* Tabel 3 menampilkan taksiran parameter-parameter (MLE) dan tes uji kecocokan dari uji Kolmogorov Smirnov (*p-value*). Dari Tabel 3 dapat dilihat bahwa nilai dari *p-value* masing-masing tes uji kecocokan setiap marginal data lebih besar dari 0.05, sehingga penaksiran awal data marginal masing-masing distribusi yang telah dipilih diterima. Distribusi marginal yang dipakai merujuk pada lampiran 1.

Tabel 3. Parameter dan Tes Uji Kecocokan Distribusi Marginal Data

Kel.	Marg.	Dist.	Taksiran Parameter		Kolmogorov-Smirnov (*p-value*)
1	X_1	Normal	$\hat{\mu} = 56$	$\hat{\sigma} = 34.8$	0.52199
	X_2	Normal	$\hat{\mu} = 94.9$	$\hat{\sigma} = 50.7$	0.46452
2	X_1	Log-normal	$\hat{\mu} = 5.3$	$\hat{\sigma} = 0.3$	0.4718
	X_2	Weibull	$\alpha = 2.6$	$\beta = 103.1$	0.97765
3	X_1	Gamma	$\alpha = 1.2$	$\beta = 16$	0.72513
	X_2	Weibull	$\alpha = 5.9$	$\beta = 957.2$	0.07818
4	X_1	Weibull	$\alpha = 1.2$	$\beta = 66.7$	0.78198
	X_2	Gamma	$\alpha = 13.2$	$\beta = 23.3$	0.19768

Ukuran Keterhubungan

Untuk mengetahui keterhubungan antar peubah data dilakukan estimasi ukuran keterhubungan menggunakan Kendall's *Tau* dan *Spearman's Rho*. Hasil ukuran keterhubungan setiap kelompok dapat dilihat pada Tabel 4.

Tabel 4. Ukuran keterhubungan Kendall's *Tau* dan Sperman's *Rho* Setiap Kelompok

Kelompok	Kendall's *Tau*	*Spearman's Rho*
1	0.2429901	0.3489590
2	-0.1739341	-0.2580892
3	-0.3133580	-0.4182848
4	-0.2544459	-0.3855276

Estimasi Parameter Copula Archimedian dan Copula Plackett

Estimasi parameter copula Archimedian yang terdiri dari copula Cayton, Gumbel, Ali-Mikhail-Haq (AMH), dan Frank menggunakan ukuran keterhubungan Kendall's *tau* bedasarkan persamaan (4) sedangkan estimasi copula Plackett menggunakan ukuran keterhubungan *Spearman's Rho* berdasarkan persamaan (9). Tabel 5 menunjukan nilai parameter θ dari setiap kelompok untuk setiap copula. Untuk mengestimasi parameter copula gumbel, nilai ukuran keterhubungan tidak boleh negatif. Oleh karena itu, tidak dilakukan estimasi parameter Copula Gumbel untuk kelompok 2 hingga kelompok 4.

Tabel 5. Nilai Parameter Copula Setiap Kelompok

Kel.	Clayton	Gumbel	AMH	Frank	Plackett
1	0.6419	1.3210	0.8222	2.2984	2.9086
2	-0.2963	-	-0.9490	-1.6050	0.4537
3	-0.4772	-	-1	-3.0704	0.2649
4	-0.4057	-	-1	-2.4190	0.2973

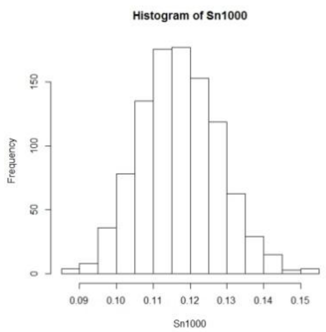
Uji Kecocokan Copula

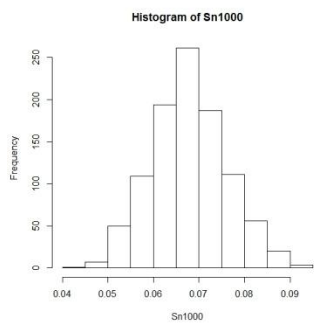
Uji kecocokan copula dilakukan melalui statistik Cram$\grave{e}$r-von Mises (S_n) seperti pada persamaan (3). Tabel 6 menunjukan nilai S_n yang telah diperoleh. Copula yang dipilih adalah copula dengan nilai S_n terkecil dibandingkan dengan copula lainnya.

Tabel 6. Statistik Crame`r-von Mises (S_n)

Kel.	Clayton	Gumbel	AMH	Frank	Plackett
1	0.12923	0.21988	0.15755	0.19540	0.19275
2	0.08162	-	0.08296	0.08220	0.08196
3	0.34164	-	0.32927	0.39166	0.37335
4	0.10930	-	0.08573	0.10297	0.10729

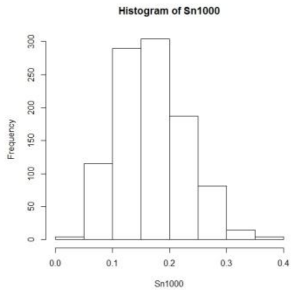
Setelah diperoleh nilai S_n untuk setiap kelompok, dilakukan uji hipotesis menggunakan *parametric bootstrap*. Melalui simulasi *parametric bootstrap* akan diperoleh 1000 nilai statistik uji S_n untuk kecocokan distribusi setiap kelompok seperti pada histogram Gambar 13 sampai Gambar 16.

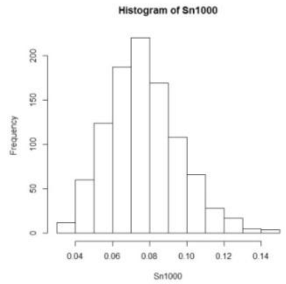




Gambar 13. Histogram S_n (1000 nilai)
Kelompok 1 Copula Clayton Distribusi Marginal Normal-Normal

Gambar 14. Histogram S_n (1000 nilai)
Kelompok 2 Copula Ali-Mikhail-Haq Distribusi Marginal Lognormal-Weibull





Gambar 15. Histogram S_n (1000 nilai)
Kelompok 3 Copula Ali-Mikhail-Haq Distribusi Marginal Gamma-Weibull

Gambar 16. Histogram S_n (1000 nilai)
Kelompok 4 Copula Ali-Mikhail-Haq Distribusi Marginal Weibull-Gamma

Setelah melakukan simulasi, langkah selanjutnya adalah mengulangi simulasi sebanyak 1000 kali. Tabel IV.13 menunjukkan banyaknya H_0 yang diterima, rata-rata *p-value* dan standar deviasi *p-value* setelah dilakukan pengulangan sebanyak 1000 kali. Terlihat bahwa semakin besar nilai S_n membawa kesimpulan ditolaknya H_0. H_0 ditolak apabila *p-value* kurang dari nilai signifikansi tertentu (0.1, 0.05 atau 0.001).

Tabel 7. Hasil Simulasi 100 Kali *Parametric Bootstrap*

Kel.		Clayton	Gumbel	AMH	Frank	Plackett
1	H_0 acc	100	100	100	100	100
	$\hat{p}_{value}$	0.1337	0.3868	0.0886	0.2437	0.2142
	Sd $\hat{p}_{value}$	0.0122	0.0139	0.0088	0.0124	0.0139
2	H_0 acc	0	-	4	0	0
	$\hat{p}_{value}$	0.0001	-	0.0377	0.0129	0.0139
	Sd $\hat{p}_{value}$	0.0004	-	0.0064	0.0039	0.0040
3	H_0 acc	0	-	0	0	0
	$\hat{p}_{value}$	0.0009	-	0.0073	1e-05	9e-05
	Sd $\hat{p}_{value}$	0.0010	-	0.0028	1e-04	0.0004
4	H_0 acc	99	-	100	82	7
	$\hat{p}_{value}$	0.0687	-	0.2821	0.0569	0.0389
	Sd $\hat{p}_{value}$	0.0082	-	0.0139	0.0076	0.0062

Setelah dilakukan pengulangan simulasi terlihat bahwa untuk tingkat signifikansi 0.05 (5%) pada copula terpilih (*p-value* terbesar) memberikan kesimpulan yang sama untuk kelompok 1 dan kelompok 4. Sedangkan untuk kelompok 2 dan kelompok 3 uji kecocokan ditolak yang berarti tidak ada copula yang cocok untuk menggambarkan distribusi pada kelompok 2 dan kelompok 3.

4. Kesimpulan

Penelitian ini mempelajari bagaimana mengelompokan data laju kunjungan dan rasio rujukan dari faskes tingkat I BPJS Kesehatan dan kemudian menganalisis hubungan antar data pada setiap klaster yang telah terbentuk. Beberapa hal yang dapat ditarik menjadi kesimpulan pada penelitian ini antara lain :

1. Melalui metode *K-Means Clustering* diperoleh kelompok-kelompok faskes tingkat I yang dapat dilihat pada Gambar 8 dimana:
   - kelompok 1 menunjukkan faskes dengan laju kunjungan rendah dan rasio rujukan juga rendah,
   - kelompok 2 menunjukkan faskes dengan laju kunjungan tinggi dan rasio rujukan rendah,
   - kelompok 3 menunjukkan faskes dengan laju kunjungan rendah dan rasio rujukan tinggi,
   - kelompok 4 menunjukkan faskes dengan laju kunjungan rendah dan rasio rujukan sedang.
2. Hubungan antar laju kunjungan dan rasio rujukan setiap kelompok dinyatakan oleh Kendall's *tau* dan Spearman *rho* pada Tabel 4 dimana setiap kelompok memiliki ukuran keterhubungan yang berbeda, dari:
   - kelompok 1 memiliki ukuran keterhubungan yang positif dan diperoleh copula gumbel dengan marginal Normal-Normal,
   - kelompok 2 memiliki ukuran keterhubungan yang negatif dan tidak diperoleh copula yang cocok dengan marginal Lognormal-Weibull,
   - kelompok 3 memiliki ukuran keterhubungan yang negatif dan tidak diperoleh copula yang cocok dengan marginal Gamma-Weibull,
   - kelompok 4 memiliki ukuran keterhubungan yang negatif dan diperoleh copula AMH dengan marginal Weibull-Gamma.

5. Daftar Pustaka

[1] W. M. P. Dhuhita, "C*lustering* dengan metode *k-means* untuk menentukan status gizi balita" Jurnal Informatika, Vol. 15, No 2, Desember 2016.

[2] T. P. Lesnussa, "Analisis klaster Indeks Pembangunan Manusia tahun 2004-2010 provinsi Maluku Utara menggunakan analisis klaster k-means", Skripsi, Fakultas Ilmu Alam dan Teknologi Rekayasa Universitas Halmahera, Tobelo, 2013.

[3] Darwis "Analisis Hubungan dan Prediksi Indeks Harga Saham Gabungan dengan Faktor Makroekonomi Melalui Pendekatan Copula", Disertasi, Sekolah Pascasarjana Institut Pertanian Bogor, Bogor, 2016.

[4] L. R. Sasongko, "Copula untuk Memodelkan Kegagalan Dua Dimensi pada Produk Bergaransi dengan Strategi Penggantian," M.Si. tesis, Program Pascasarjana Magister Aktuaria, Institut Teknologi Bandung, Bandung, 2014.

[5] R. B. Nelsen, *An Introduction to Copulas*, New York: Springer Series in Statistics, USA, 2006.

[6] C. Genest, B. Remillard, dan D. Beaudoin, "Goodness-of-fit Tests for Copulas: a review and a power study, *Insurance : Mathematics and Economics"*, **44**, 199-214, 2009.

[7] D. B. Nugroho, "Metode Numerik", unpublished.

Lampiran 1. Fungsi Distribusi Univariat

Pada bagian ini diberikan fungsi distribusi univariat yang digunakan dan dituliskan dalam paper ini. Misalkan X adalah peubah acak kontinu dengan fungsi distribusi F_X.

1. Normal, $X \sim N(\mu, \sigma^2)$.

$$F_X = \frac{1}{x\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} dt, \qquad x \in [0, \infty)$$

2. Lognormal, $X \sim Lognormal(\mu, \sigma^2)$.

$$F_X = \frac{1}{x\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} dt, \qquad x \in [0, \infty)$$

3. Weibull dua parameter, $X \sim Weibull(\alpha, \beta)$.

$$F_X = 1 - \exp\left\{-\left(\frac{x}{\alpha}\right)^{\beta}\right\}, \qquad x \in [0, \infty)$$

4. Gamma, $X \sim Gamma(\alpha, \beta)$

$$F_X = \frac{\Gamma_{X/\beta}(\alpha)}{\Gamma(\alpha)}, \qquad x \in [0, \infty)$$