PENGELOMPOKAN NEGARA BERDASARKAN POPULASI URBAN DENGAN ALGORITMA EXPECTATION-MAXIMIZATION

Farah Mufidah[1], Irwan Susanto[2], Etik Zukhronah[3]
[1,2,3]Statistika, Universitas Sebelas Maret
e-mail :[1]mufidahfarah30@student.uns.ac.id, [2]irwansusanto@staff.uns.ac.id, [3]etikzukhronah@staff.uns.ac.id,

ABSTRACT

Urbanization is a shift in population from rural to urban areas. Many factors cause people to urbanize. Urbanization is not only related to population, but also related to politics, socio-cultural, and economics It is necessary to group the countries based on the urban population to understand the characteristics of each country so that country can be carried out according to its characteristics. This reserach uses data of world urban population year 2019 with the number of countries is 214 countries. This research, will be carried out modeling of the world on urban population data year 2019 by using a finite mixture model. The data that can be modeled by finite mixture modeling is the data that holds a multimodal pattern or has more than one number of peaks on the histogram as an indication that the data has different clusters. The parameter estimation uses the maximum likelihood estimation through the Expectation-Maximization (EM) algorithm. The distribution of the urban population has a heavy-tailed data type characretistic. The result of parameter estimation represent that the model has a gamma mixture distribution. The model selection process with AIC and BIC measures indicates that the best model is a finite mixture model with two components gamma. The first cluster has 187 countries that have an average urban population of 8,130,222 people and the second cluster has 27 countries that have an average urban population of 97,833,200 people.

Keywords : *EM algorithm, finite mixture model, urban population*

INTISARI

Urbanisasi dapat diartikan sebagai perpindahan penduduk dari pedesaan menuju perkotaan. Banyak faktor yang menyebabkan masyarakat untuk melakukan urbanisasi. Urbanisasi tidak hanya berhubungan dengan kependudukan saja, namun juga dapat berpengaruh dari sisi politik, sosial budaya, hingga ekonomi. Perlu adanya pengelompokan negara berdasarkan populasi urban untuk mengetahui karakteristik masing-masing negara sehingga negara dapat diperlakukan sesuai dengan karakteristiknya. Penelitian ini menggunakan data populasi urban dunia pada tahun 2019 yaitu jumlah negara sebanyak 214 negara. Pada penelitian ini akan dilakukan pemodelan pada data populasi urban di dunia pada tahun 2019 dengan menggunakan model finite mixture. Data yang dapat dimodelkan dengan pemodelan finite mixture adalah data yang memiliki pola multimodal atau memiliki lebih dari satu jumlah puncak pada histogram sebagai indikasi data memiliki klaster yang berbeda. Estimasi parameter menggunakan estimasi maksimum likelihood melalui algortima Expectation-Maximization (EM). Distribusi populasi urban mempunyai karakteristik data bertipe heavy-tailed. Hasil estimasi parameter menunjukkan bahwa model berdistribusi gamma mixture. Proses seleksi model dengan ukuran AIC dan BIC menunjukkan bahwa model yang terbaik adalah model finite mixture dengan dua komponen gamma.. Klaster pertama mempunyai anggota 187 negara yang mempunyai rata-rata populasi urban sebesar 8.130.222 penduduk dan klaster kedua mempunyai anggota 27 negara yang mempunyai rata-rata populasi urban sebesar 97.833.200 penduduk.

Kata kunci : *Algoritma EM , Model Finite Mixture, Populasi Urban*

1. PENDAHULUAN

Secara popular urbanisasi diartikan sebagai perpindahan penduduk dari pedesaan menuju perkotaan. Namun, pengertian yang lebih tepat adalah proporsi penduduk yang tinggal di perkotaan (*urban area*) karena perkotaan (*urban area*) tidak sama artinya dengan kota (*city*) (Tjiptoherijanto, 1999). Banyak faktor yang menyebabkan masyarakat untuk melakukan urbanisasi. Urbanisasi tidak hanya berhubungan dengan kependudukan saja, namun dari sisi politik, sosial budaya, hingga ekonomi dapat berpengaruh pula. Aglomerasi menjadi salah satu sebab karena adanya urbanisasi yang besar pada suatu daerah. Hal tersebut akan memunculkan masalah-masalah baru pada suatu daerah di ranah industri. Perlunya pengelompokan negara berdasarkan populasi urban untuk mengetahui karakteristik masing-masing negara sehingga dapat memperlakukan setiap negara sesuai dengan karakteristiknya. Urbanisasi berkembang pesat di Asia dan Afrika. Diperkirakan bahwa perkembangan urbanisasi 90% di benua tersebut (Widyaningrum, 2018). Pada benua Asia, urbanisasi didominasi oleh negara-negara berkembang sehingga menyebabkan kurangnya pasokan pangan domestik (Mayasari, 2019).

Susanto dan Handajani (2020) meneliti pemodelan distribusi pendapatan rumah tangga per kapita di Indonesia dengan model *finite mixture* menggunakan metode estimasi maksimum likelihood melalui algoritma *Expectation-Maximization* (EM) dan menghasilkan model *finite mixture* lognormal dengan empat komponen *mixture*. Penelitian serupa dilakukan oleh Hilda dkk (2020) mengenai penerapan algoritma *Expectation-Maximization* (EM) pada pemodelan normal *mixture* nilai tambah produk domestik bruto di setiap negara dan hasil penelitian tersebut adalah data dapat dimodelkan dengan model normal *mixture* dua komponen.

Distribusi populasi urban mempunyai karakteristik data bertipe *heavy-tailed* sehingga menggunakan distribusi probabilitas yang berkarakteristik *heavy-teailed*, salah satunya yaitu distribusi gamma. Distribusi gamma merupakan salah satu distribusi variabel random kontinu (Harinaldi, 2005). Pengelompokan data populasi urban menggunakan pemodelan *finite mixture*. Data yang dapat dimodelkan dengan pemodelan *finite mixture* adalah data yang memiliki pola multimodal atau memiliki jumlah puncak pada histogram lebih dari satu sebagai indikasi data memiliki klaster yang berbeda (Hilda dkk., 2020). Estimasi parameter dilakukan menggunakan algoritma *Expectation-Maximization* (EM).

2. METODE PENELITIAN

Penelitian ini menggunakan data populasi urban di dunia pada tahun 2019 dengan jumlah 214 negara. Data tersebut adalah data sekunder yang diambil dari *The World Bank* dan diakses di laman http://worldbank.org. Pada penelitian ini dilakukan analisis klaster dengan menggunakan *model based clustering.* Estimasi parameter menggunakan estimasi maksimum likelihood melalui algoritma EM. Komponen yang terbentuk dari model *finite mixture* gamma berdasarkan ukuran AIC dan BIC menjadi dasar pengelompokan negara pada data populasi urban.

Langkah-langkah pemodelan *finite mixture* pada penelitian ini adalah sebagai berikut :

1. Identifikasi distribusi data populasi urban dengan melihat pola distribusi dan melalui uji Anderson Darling.
2. Estimasi model *finite mixture* gamma dengan metode estimasi maksimum likelihood melalui algoritma *Expectation-Maximization* (EM).
3. Uji *bootstrap likelihood ratio statistic* untuk melakukan uji signifikansi model *finite mixture*.
4. Menentukan banyaknya komponen *mixture* dengan melihat nilai AIC dan BIC yang terkecil, kemudian menginterpretasikan hasil pengelompokannya.

Teori analisis pada penilitian ini antara lain, model *finite mixture*, metode estimasi maksimum likelihood pada model *mixture*, uji *goodness of fit*, uji signifikansi model *finite mixture*, dan seleksi model.

2.1. Model *Finite Mixture*

Suatu vektor variabel random $x = [x_1, x_2, \dots, x_n]^T$ yang bertipe diskrit atau kontinu dikatakan berasal dari distribusi *finite mixture* jika memiliki fungsi kepadatan probabilitas didefinisikan dengan Persamaan (1).

$$f(x_i) = w_1 f_1(x_1) + \cdots + w_K f_K(x_i) \tag{1}$$

untuk $x_i, i = 1,2, \dots, n$ dengan k adalah banyaknya komponen *mixture*. $w = [w_1, w_2, \dots, w_K]^T$ adalah vektor parameter *weight* dari distribusi *finite mixture*. Nilai-nilai dalam w harus memenuhi $0 \le w_k \le 1$ dan $\sum_{k=1}^{K} w_k = 1$. Jika θ merupakan vektor parameter dari distribusi probabilitas semua komponen *mixture*, maka fungsi kepadatan *mixture* dinyatakan dalam Persamaan (2).

$$f(x_i; \Psi) = w_1 f_1(x_1; \theta_1) + w_2 f_2(x_2; \theta_2) + \cdots + w_K f_K(x_i; \theta_K) = \sum_{k=1}^{K} w_k f_k(x_i; \theta_k) \tag{2}$$

dengan $\Psi = [w, \theta]^T, \theta = [\theta_1, \theta_2, \dots, \theta_K]^T$ (Frühwirth-Schnatter, 2006).

2.2. Metode Estimasi Maksimum Likelihood

Metode estimasi maksimum likelihood dapat digunakan untuk mengestimasi parameter model dalam model *mixture*. Jika diberikan observasi x_i , $i = 1, \dots, n$ yang saling independen dan $\Psi = [w, \theta]^T$ maka fungsi likelihood dari model *finite mixture* didefinisikan dengan Persamaan (3).

$$L(\Psi) = \prod_{i=1}^{n} \left[\sum_{k=1}^{K} w_k f_k(x_i; \theta_k) \right] \tag{3}$$

Untuk mengestimasi parameter model *finite mixture* pada pendekatan metode estimasi maksimum likelihood, penduga maksimum likelihood $\widehat{\Psi}$ adalah penyelesaian dari Persamaan (4).

$$\frac{\partial lnL(\Psi)}{\partial(\Psi)} = 0$$

dengan

$$L(\Psi) = lnL(\Psi) = \sum_{i=1}^{n} ln\left[\sum_{k=1}^{K} w_k f_k(x_i;\theta_k)\right] \tag{4}$$

merupakan fungsi log-likehood observasi (McLachlan & Pell, 2000). Dalam penyelesaian persamaan (4) tersebut tidak mudah dilakukan secara analitik maupun pendekatan numerik sehingga diperlukan algoritma EM. Algoritma EM dikemukakan pertama kali oleh Dempster, dkk. (1997).

2.3. Algoritma *Expectation Maximization (EM)*

Pada konsep EM menganggap bahwa x_i adalah data yang tidak lengkap z_i adalah bagian dari pelengkap data yang tidak terobservasi. Jika x_i masuk dalam komponen *mixture k* maka, nilai $z_{ik} = 1$ dan jika x_i tidak masuk dalam komponen *mixture k* maka, nilai $z_{ik} = 0$. Fungsi log likehood untuk data lengkap diberikan oleh Persamaan (5).

$$lnL(\Psi) = \sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik}[\ln w_k + \ln\ f_k(x_i;\theta_k)] \tag{5}$$

Algoritma EM berjalan mengikuti dua tahap yaitu tahap *expectation* (E) dan tahap *maximization* (M) Dempster, dkk. (1997). Dimisalkan nilai awal $\Psi^{(0)}$ maka algoritma EM untuk estimasi model *finite mixture* adalah sebagai berikut,

1) Tahap *Expectation* (E)

Tahap *expectation* dihitung nilai harapan fungsi log-likehood untuk data lengkap pada saat iterasi ke-s adalah sebagai berikut

$$Q\left(\Psi;\Psi^{(k)}\right) = E(lnL(\Psi)|x_i,\Psi^{(s)})$$
$$= \sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik}{}^{(s)}\{\ln\left(w_k^{(s)}\right) + \ln\left(f_k\left(x_i\middle|\theta_k^{(s)}\right)\right)\}$$

Penduga $z_{ik}{}^{(s)}$ yang merupakan probabilitas observasi dari x_i dihitung melalui Persamaan (6).

$$z_{ik}{}^{(s)} = \frac{w_k^{(s)} f_k\left(x_i\middle|\theta_k^{(s)}\right)}{\sum_{k=1}^{K} w_k^{(s)}\left(f_k\left(x_i\middle|\theta_k^{(s)}\right)\right)} \tag{6}$$

2) Tahap *Maximization* (M)

Parameter bobot w_k pada iterasi ke-($s+1$) diduga dengan Persamaan (7)

$$\widehat{w}_k^{(s+1)} = \frac{\sum_{i=1}^{n} \hat{z}_{ik}^{(s)}}{n} \tag{7}$$

Diperkirakan parameter $\hat{\theta}_i^{(s+1)}$ merupakan penyelesaian dari Persamaan (8).

$$\frac{\partial}{\partial\theta}\left\{\sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik}{}^{(s)}\ln\ f_k(x_i|\theta_k)\right\} = 0 \tag{8}$$

Algoritma EM akan diterapkan pada estimasi parameter dari tiga model *finite mixture* yaitu model *finite mixture* lognormal, model *finite mixture* gamma, dan *model finite mixture* Weibull.

Pada estimasi model *finite mixture* lognormal, $f(x_i;\theta_k)$ merupakan distribusi lognormal, $f(x_i;\theta_k)\sim Logn(\mu_k,\sigma_k^2)$,

$$f(x_i;\theta_k) = f(x_i;\mu_k,\sigma_k^2) = \frac{1}{x_i\sigma_k\sqrt{2\pi}}\left[-\frac{1}{2}\left(\frac{lnx_i-\mu_k}{\sigma_k}\right)^2\right]$$

dengan $\theta_k = [\mu_k,\sigma_k^2]^T$, $\mu_k > 0$, dan $\sigma_k > 0$. Sedangkan pada estimasi model *finite mixture* gamma, $f(x_i;\theta_k)$ merupakan distribusi gamma, $f(x_i;\theta_k)\sim Gam(\alpha_k,\beta_k)$,

$$f(x_i;\theta_k) = f(x_i;\alpha_k,\beta_k) = \beta_k^{\alpha_k}\left(\Gamma(\alpha_k)\right)^{-1} x_i^{\alpha_k-1} e^{-\beta_k x_i}$$

dengan $\theta_k = [\alpha_k,\beta_k]^T$, $\alpha_k > 0$, dan $\beta_k > 0$. Sementara pada estimasi model *finite mixture* Weibull, $f(x_i;\theta_k)$ merupakan distribusiWeibull, $f(x_i;\theta_k)\sim Wei(\delta_k,\lambda_k)$,

$$f(x_i;\theta_k) = f(x_i;\delta_k,\lambda_k) = \frac{\delta_k}{\lambda_k}\left(\frac{x_i}{\lambda_k}\right)^{\delta_k-1} exp(-x_i/\lambda_k)^{\delta_k}$$

dengan $\theta_k = [\delta_k,\lambda_k]^T$, $\delta_k \geq 0$, dan $\lambda_k \geq 0$

2.4. Uji *Goodness of Fit*

Uji *goodness of fit* dilakukan dengan uji Anderson Darling untuk mengidentifikasi data berdistribusi univariat multimodal Uji Anderson Darling memiliki langkah-langkah sebagai berikut:

1. Menentukan uji hipotesis
   H_0: pola data mengikuti distribusi probabilitas *unimodal* tertentu
   H_1: pola data tidak mengikuti distribusi probabilitas *unimodal* tertentu
2. Menentukan tingkat signifikansi α
3. Menghitung statistik uji dengan Persamaan (9).

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}[2i-1][ln(F(x_i)) + ln(1 - F(x_{n+1-i}))] \tag{9}$$

   Modifikasi dari metode Anderson Darling menggunakan rumus Persamaan (10) dan Persamaan (11).

$$A^* = A\left(1 + \frac{0{,}75}{n} + \frac{2{,}25}{n^2}\right) \tag{10}$$

   dan

$$c_\alpha = a_\alpha\left(1 - \frac{b_\alpha}{n} - \frac{d_\alpha}{n^2}\right) \tag{11}$$

   a_α , b_α, d_α ditunjukkan dalam tabel kritis Anderson Darling
4. Menentukan daerah kritis
   H_0 ditolak apabila $A^* > c_\alpha$ atau nilai $p < \alpha$.

2.5. Uji Signifikansi Model *Finite Mixture*

Dilakukan uji signifikansi model *finite mixture* berbasis uji *bootstrap likelihood ratio statistics* untuk melihat pemodelan data yang sesuai pada model *finite mixture* (Feng dan McCulloch, 1996). Uji *bootstrap likelihood ratio statistics* memiliki hipotesis yaitu,

$H_0: K = K_0$ (model *finite mixture* memiliki K_0 komponen *mixture*)

$H_1: K = K_1 = K_0 + 1$ (model *finite mixture* memiliki komponen *mixture* dari banyaknya komponen *mixture* dalam hipotesis *null* ditambahkan satu komponen)

H_0 akan ditolak jika nilai $p < \alpha$, sedangkan nilai p diperoleh dengan menggunakan rumus Persamaan (12), (13) dan (14).

$$p = \frac{1}{B}\sum_{b=1}^{B} I(\,Irs_1^{(b)} > Irs_0) \tag{12}$$

$$Irs_0 = -2\left(\ell(\hat{\psi}_0) - \ell(\hat{\psi}_1)\right) \tag{13}$$

$$Irs_1 = -2\left(\ell(\hat{\psi}_0; x^*) - \ell(\hat{\psi}_1; x^*)\right) \tag{14}$$

dan B adalah banyaknya proses *bootstrap* yang dilakukan untuk membentuk vektor $Irs_1^{(1)}, \dots, Irs_1^{(B)}$

2.6. Seleksi Model

Seleksi model dilakukan untuk mengetahui banyaknya komponen *mixture* yang sesuai berdasarkan metode berbasis kriteria informasi yaitu *Akaike Information Criterion* (AIC) dan *Bayesian Information Criteria* (BIC). Perhitungan AIC menggunakan Persamaan (15), sedangkan BIC menggunakan Persamaan (16).

$$AIC = -2\ln L(\hat{\psi}) + 2p \tag{15}$$

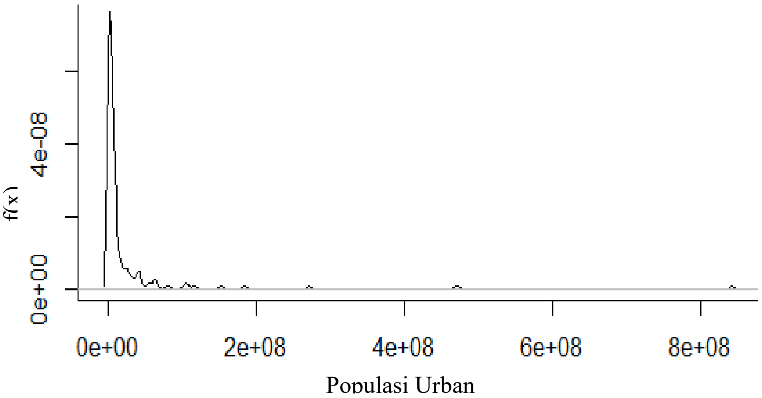
$$BIC = -2\ln L(\hat{\psi}) + p\ln(n) \tag{16}$$

dengan $L(\hat{\psi})$ adalah fungsi likelihood dari penduga maksimum likelihood $\hat{\psi}$ dan p adalah banyaknya parameter dalam model *finite mixture* serta n adalah banyaknya data.

3. HASIL DAN PEMBAHASAN

3.1. Identifikasi Pola Data Populasi Urban

Diberikan observasi $\boldsymbol{x_i}$ sebagai populasi urban di dunia tahun 2019. Dilakukan identifikasi pola distribusi data pada data populasi urban di dunia tahun 2019 untuk melihat adanya pola multimodal. Identifikasi pola data dilakukan melalui plot kernel histrogram pada Gambar 1. dan *uji goodness of fit*.



Gambar 1. Plot Kernel Histrogram Populasi Urban di Dunia

Gambar 1 menunjukkan bahwa grafik memiliki beberapa puncak distribusi atau bersifat multimodal. Data populasi urban memiliki karakteristik *heavy tailed* dan bernilai positif, maka distribusi yang sesuai adalah distribusi lognormal, distribusi gamma, atau distribusi Weibull. Selanjutnya, untuk melihat terpenuhi atau tidak pola multimodal dilakukan uji signifikansi *goodness of fit*.

Uji signifikansi *goodness of fit* menggunakan uji Anderson Darling yang memiliki hipotesis sebagai berikut :

H_0: pola data mengikuti distribusi probabilitas *unimodal* tertentu

H_1: pola data tidak mengikuti distribusi probabilitas *unimodal* tertentu

Tabel 1. Hasil Uji Anderson Darling

Distribusi Probabilitas pada H_0	AD	Nilai p	Kesimpulan
lognormal	2,938	<0,005	H_0 ditolak
gamma	4,828	<0,005	H_0 ditolak
Weibull	0,974	0,015	H_0 ditolak

Catatan : tingkat signifikansi α = 0,05

Tabel 1 menunjukkan bahwa data pada semua distribusi memiliki nilai $p < 0{,}05$ yang berarti pola data tidak mengikuti distribusi probabilitas *unimodal* tertentu. Selanjutnya, akan dilakukan uji signifikansi berbasis *boostrap likelihood ratio statistics* untuk melihat lebih lanjut distribusi yang cocok untuk distribusi *finite mixture.*

3.2. Uji Signifikansi Model

Uji signifikansi model berbasis *boostrap likelihood ratio statistics* dilakukan untuk mengetahui sesuai atau tidaknya data jika dimodelkan dengan *finite mixture*. Hipotesis null pada uji *boostrap likelihood ratio statistics* adalah bahwa model *finite mixture* memiliki K_0 komponen *mixture*, sedangkan hipotesis alternative pada uji *boostrap likelihood ratio statistics* adalah bahwa model *finite mixture* memiliki komponen *mixture* dari banyaknya komponen *mixture* dalam hipotesis *null* ditambahkan satu komponen. Berikut hipotesis dari uji *boostrap likelihood ratio statistic* :

$H_0: K = K_0$ (model *finite mixture* memiliki K_0 komponen *mixture*)

$H_1: K = K_1 = K_0 + 1$ (model *finite mixture* memiliki komponen *mixture* dari banyaknya komponen *mixture* dalam hipotesis *null* ditambahkan satu komponen)

Pengujian dilakukan untuk setiap model *finite mixture* lognormal, gamma, dan Weibull. Hasil dari uji *boostrap likelihood ratio statistic* ditunjukkan pada Tabel 2 Jika nilai $p < \alpha$ maka, data sesuai untuk dimodelkan dengan model *finite mixture* dengan komponen *mixture* yang berdistribusi lognormal, gamma, maupun Weibull. Namun, pada prosesnya hasil dari uji *boostrap likelihood ratio statistic* distribusi lognormal tidak memberikan hasil perhitungan, maka yang digunakan hanya distribusi gamma dan Weibull.

Tabel 2. Hasil Uji *Boostrap Likelihood Ratio Statistic*

Distribusi Komponen *Mixture*	Nilai p	Keputusan	Kesimpulan
lognormal	*NaN*	H_0 diterima	Tunggal
gamma	0,00	H_0 ditolak	*Mixture*
Weibull	0,15	H_0 diterima	Tunggal

Catatan : tingkat signifikansi α = 0,05

Tabel 2 mengetahui bahwa yang memenuhi signifikansi adalah distribusi gamma, sehingga data populasi urban di dunia tahun 2019 lebih sesuai dimodelkan dengan model *finite mixture* yaitu komponen *mixture* berdistribusi gamma.

3.3. Seleksi Model

Dilakukan seleksi model melalui ukuran AIC dan BIC untuk mengetahui banyaknya komponen *mixture* yang sesuai untuk model. Hasil perhitungan AIC dan BIC untuk setiap komponen dituliskan dalam Tabel 3. Banyaknya komponen *mixture* yang dihitung hanya sampai dua komponen saja karena komponen yang lebih dari dua tidak memberikan hasil perhitungan.

Tabel 3. Nilai AIC dan BIC dari Estimasi Model *Finite Mixture*

Distribusi	Banyak Komponen	AIC	BIC
gamma	2	7269,7239522	7286,5538323

Tabel 3 menunjukkan bahwa estimasi model *finite mixture* yang sesuai untuk data populasi urban di dunia tahun 2019 adalah model *finite mixture* gamma dengan dua komponen. Kedua komponen menggambarkan dua klaster berdasarkan jumlah populasi urban di masing-masing negara. Hasil komputasi estimasi model *finite mixture* diberikan oleh Persamaan (17).

$$f(x_i|\Psi) = \hat{w}_1 Gam(\hat{\alpha}_1, \hat{\beta}_1) + \hat{w}_2 Gam(\hat{\alpha}_2, \hat{\beta}_2) \tag{17}$$

dengan parameter bobot $\hat{w}_1$= 0,869 dan $\hat{w}_2$= 0,131. Parameter-parameter dari distribusi gamma pada komponen pertama yaitu $\hat{\alpha}_1$= 0,419 dan $\hat{\beta}_1$= 0,0000001, sedangkan parameter-parameter dari distribusi gamma pada komponen kedua yaitu $\hat{\alpha}_2$= 0,385 dan $\hat{\beta}_2$= 0,0000000039. Masing-masing komponen memiliki *mean* dan standar deviasi yaitu pada komponen pertama μ_1= 8.130.222 dan σ_1= 12.559.640, sedangkan pada komponen kedua μ_2= 97.833.200 dan σ_2= 157.669.200

Berdasarkan model *finite mixture* yang didapat dan memperhatikan parameter dari setiap komponen, maka diketahui deskripsi dari masing-masing klaster populasi urban di dunia. Klaster pertama mempunyai anggota 187 negara yang mempunyai rata-rata populasi urban sebesar 8.130.222 penduduk dan klaster kedua mempunyai anggota 27 negara yang mempunyai rata-rata populasi urban sebesar 97.833.200 penduduk. Klaster pertama didominasi oleh negara-negara yang mempunyai wilayah kecil seperti San Marino, Monako, dan Grenada sedangkan klaster kedua didominasi oleh negara yang mempunyai populasi penduduk tinggi seperti China, India, dan Indonesia serta negara yang mempelopori perekonomian dan teknologi dunia seperti Amerika Serikat, Jepang, dan Korea Selatan.

3.4. KESIMPULAN

Data populasi urban di dunia tahun 2019 memiliki pola distribusi multimodal dan *heavy-tailed* sehingga lebih tepat dimodelkan dengan distribusi *finite mixture*. Estimasi model melalui metode EM menunjukka bahwa model *finite mixture* dengan komponen *mixture* yang berdistribusi gamma. Proses seleksi model dengan ukuran AIC dan BIC menunjukkan bahwa model yang terbaik adalah model *finite mixture* gamma dengan dua komponen. Kedua komponen menggambarkan dua klaster berdasarkan jumlah populasi urban di masing-masing negara. Klaster pertama mempunyai anggota 187 negara yang mempunyai rata-rata populasi urban sebesar 8.130.222 penduduk dan klaster kedua mempunyai anggota 27 negara yang mempunyai rata-rata populasi urban sebesar 97.833.200 penduduk. Klaster pertama didominasi oleh negara-negara yang mempunyai wilayah kecil seperti San Marino, Monako, dan Grenada sedangkan klaster kedua didominasi oleh negara yang mempunyai populasi penduduk tinggi seperti China, India, dan Indonesia serta negara yang mempelopori perekonomian dan teknologi dunia seperti Amerika Serikat, Jepang, dan Korea Selatan.

UCAPAN TERIMA KASIH

Artikel ini dibuat sebagai salah satu rangkaian pelaksanaan tugas akhir di program studi statistika UNS. Penulis mengucapkan terima kasih kepada seluruh dosen pembimbing yang telah membimbing dan mendukung keberlangsungan kegiatan penelitian ini.

DAFTAR PUSTAKA

Dempster, A., Laird, N., dan Rubin, D. (1997). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 39, No. 3, hal 609-617.

Feng , Z., dan McCulloch, C. (1996). Using Bootstrap Likelihood Ratios in Finite Mixture Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 3, hal 609-617.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models.* Springer, New York

Harinaldi. (2005). *Prinsip-prinsip Statistik untuk Teknik dan Sains.* Erlangga, Jakarta

Hilda, M., Susanto, I., dan Handajani, S. (2020). Penerapan Algoritma Expectation-Maximization pada Pemodelan Normal Mixture Nilai Tambah Produk Domestik Bruto di Setiap Negara. *Sendika*, Vol. 6 No. 1.

Mayasari, D. (2019). *Pertumbuhan Populasi, Urbanisasi dan Melambungnya Kebutuhan Pangan, Sementara Stok Pangan Kian Terbatas akan Pengaruhi Pasokan Pangan Masa Depan di Asia*. Diakses dari Cargill: https://www.cargill.co.id/id/

McLachlan, G., dan Pell, D. (2000). *Finite MIxture Models.* Wiley, New York

Susanto, I., dan Handajani, S. S. (2020). Pengelompokan Rumah Tangga di Indonesia Berdasarkan Pendapatan Per Kapita dengan Model Finite Mixture. *Media Statistika*, *13*(1), 13-24.

Tjiptoherijanto, P. (1999). Urbanisasi dan Pengembangan Kota di Indonesia. *Populasi*, 10(2).

Widyaningrum, G. L. (2018). Diakses 18 Mei 2020, dari National Geographic Indonesia: https://nationalgeographic.grid.id/