

HIERARCHICAL CLUSTERING MULTISCALE BOOTSTRAP **UNTUK PENGELOMPOKAN KEMISKINAN DI JAWA TIMUR**

Gangga Anuraga

Program Studi Statistika Universitas PGRI Adi Buana (UNIPA) Surabaya

Alamat e-mail : ganuraga@gmail.com

ABSTRAK

Kemiskinan di Jawa Timur merupakan problematika hingga sekarang. Berbagai program telah direncanakan guna menanggulangi permasalahan tersebut. Agar program tersebut dapat terlaksana optimal, perlu strategi perencanaan yang mendalam terutama terkait pemetaan daerah dan karakteristik kemiskinannya. Metode hierarki *complete linkage* digunakan untuk mengelompokkan daerah-daerah berdasarkan kemiripan karakteristik. Penggunaan metode hierarki dengan teknik pengukuran jarak hanya memberikan satu solusi dalam penyelesaiannya, yaitu didasarkan pada ukuran kemiripan pada teknik jarak yang digunakan. Sehingga penelitian ini menggunakan pendekatan *multiscale bootstrap* untuk memberikan ukuran ketidaktentuan dalam metode pengelompokan klasik *hierarchical clustering*. Metode ini bekerja dengan pendekatan *bootstrap resampling* untuk setiap kelompok (*cluster*). Pendekatan *multiscale bootstrap* dalam *Hierarchical Cluster Analysis* memiliki kestabilan pada saat iterasi $B = 500$. Hal itu dapat diketahui dari signifikansi nilai AU (*P-values*) yang mendekati 0.95 dengan nilai *standar error bootstrap* yang kecil. Terdapat lima kelompok (*cluster*) yang terbentuk dengan kemiripan satu dengan yang lain dan signifikan dengan nilai AU (*Approximately Unbiased*) $p_value \geq 0,95$. Berdasarkan signifikansi *Approximately unbiased (AU) p-values* $\geq 0,95$, kelompok kelima yang terdiri dari Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan dapat dikatakan sebagai daerah yang masih cukup tinggi persentase kemiskinannya dibandingkan dengan kelompok 1, 2, 3, dan 4.

Kata Kunci : *Approximately Unbiased (AU) p-values, clustering hierarki complete linkage, multiscale bootstrap, kemiskinan*

PENDAHULUAN

Analisis kelompok (*cluster analysis*) adalah salah satu analisis statistik yang bertujuan mengelompokkan obyek-obyek berdasarkan kesamaan karakteristik diantara obyek-obyek tersebut. Obyek tersebut akan diklasifikasikan ke dalam satu atau beberapa kelompok (*cluster*) sehingga obyek-obyek yang berada dalam satu kelompok akan mempunyai kemiripan satu dengan yang lain. Dengan demikian terdapat homogenitas (kesamaan) yang tinggi antar anggota

dalam satu kelompok (*within-cluster*) dan heterogenitas (perbedaan) yang tinggi antar kelompok yang satu dengan kelompok lainnya (*between-cluster*) [7]. Terdapat dua metode umum dalam algoritma *cluster*, yaitu hierarki dan non hierarki. Pada metode non-hierarki (*non hierarchical clustering*), jumlah kelompok ditentukan di awal sedangkan pada metode hierarki (*hierarchical clustering*), jumlah kelompok ditentukan di akhir analisis. Metode hierarki dibagi berdasarkan teknik pengukuran jarak antar kelompok yang digunakan untuk

mengelompokkan objek. Pengukuran jarak dalam metode hierarki antara lain *single linkage* yang menggunakan jarak minimum, *complete linkage* menggunakan jarak maksimum, *average linkage* menggunakan jarak rata-rata, dan *median linkage* yang menggunakan jarak median antar kelompok.

Penggunaan metode hierarki dengan teknik pengukuran jarak hanya memberikan satu solusi dalam penyelesaiannya, yaitu didasarkan pada ukuran kemiripan pada teknik jarak yang digunakan. Sehingga Nilai signifikansi statistik dan ukuran akurasi dalam pengelompokan tidak dapat ditemukan dengan metode klasik ini.

[10] mengenalkan pendekatan *multiscale bootstrap* untuk memberikan ukuran ketidaktentuan dalam metode pengelompokan klasik *hierarchical clustering*. Metode ini bekerja dengan pendekatan *bootstrap resampling* untuk setiap kelompok (*cluster*). Sehingga didapatkan ukuran validasi yaitu nilai *p-values* dalam membentuk kelompok (*cluster*) yang mempunyai kemiripan satu dengan yang lain. Terdapat dua nilai *p-values* dalam pendekatan *bootstrap resampling* yaitu *probability bootstrap (BP) value* dan *approximately unbiased (AU) p-values*. *Multiscale bootstrap resampling* digunakan untuk menghitung nilai *Approximately unbiased (AU) p-values*, yang mana memberikan taksiran yang lebih baik dalam mengatasi bias. Beberapa penelitian tentang klasifikasi dengan pendekatan *bootstrap* antara lain adalah [9] memperkenalkan *Approximately unbiased (AU) p-values* pada klasifikasi gen. [8] menggunakan pendekatan *bootstrap* untuk klasifikasi atau pengelompokan pada data *microarray*, yang digunakan untuk mengukur reliabilitas pengelompokan.

Kemiskinan merupakan masalah multidimensi dan lintas sektor yang dipengaruhi oleh berbagai faktor yang saling berkaitan, antara lain tingkat

pendapatan, kesehatan, pendidikan, akses terhadap barang dan jasa, lokasi, geografis, gender, dan kondisi lingkungan (BAPPENAS, 2010). Menurut [2], pada bulan Maret 2012 penduduk miskin Jawa Timur sebanyak 5,071 juta (13,40 persen) atau turun 5,53 persen dibandingkan dengan jumlah penduduk miskin pada bulan Maret 2011 yang mencapai 5,356 juta jiwa (14,23 persen). Dengan wilayah yang luas sekitar $\pm 47.922 \text{ km}^2$, Provinsi Jawa Timur terbagi atas 29 kabupaten dan 9 kota atau secara administratif terdapat 38 Kabupaten/Kota, sehingga dapat dikatakan berkembang dalam bidang ekonomi dan dapat menjadi potensi pembangunan. Namun demikian, dalam realitanya masih banyak masyarakat yang hidup di bawah persentase garis kemiskinan. Pada tahun 2011 persentase kemiskinan di Provinsi Jawa Timur masih berada diatas persentase garis kemiskinan tingkat nasional sebesar 13,33%. Berbagai program telah dicanangkan pemerintah baik pusat maupun daerah untuk menang-gulangi masalah kemiskinan mulai dari bidang pendidikan, pangan, kesehatan, perluasan kesempatan kerja, bantuan sarana dan prasarana pertanian, bantuan kredit usaha untuk masyarakat miskin dan lain sebagainya. Banyaknya program yang dicanangkan tentu saja perlu didukung oleh strategi perencanaan yang matang. Strategi perencanaan program bisa saja berbeda untuk tiap-tiap daerah.

Oleh karena itu, untuk mencapai terselenggaranya program pengentasan kemiskinan yang optimal, perlu dilakukan pengelompokan karakteristik kemiskinan di Kabupaten/Kota di Jawa Timur. Sehingga dilakukan penelitian dengan menggunakan *Hierarchical clustering* dengan pendekatan *multiscale bootstrap* untuk pengelompokan kemiskinan di Jawa Timur tahun 2011, dimana variabel yang digunakan antara

lain indikator-indikator kemiskinan di sektor ekonomi, kesehatan, dan SDM.

Analisis kelompok (*cluster analysis*) digunakan untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan pengukuran variabel-variabel yang diamati, sehingga objek dalam kelompok yang sama memiliki kemiripan karakteristik dan antarkelompok memiliki ketakmiripan karakteristik. Metode pengelompokan hierarki (*Hierarchical clustering*) memulai pembentukan kelompok dengan jumlah kelompok sama dengan jumlah obyek (n) kemudian dua kelompok yang mempunyai jarak terdekat digabung menjadi satu kelompok sehingga jumlah kelompok menjadi $(n-1)$. Pada penelitian ini digunakan metode *complete linkage* karena metode ini menghasilkan perbedaan faktor-faktor yang mempengaruhi karakteristik kemiskinan pada masing-masing kelompok. Selanjutnya jarak-jarak antar kelompok yang baru dihitung kembali. Prosedur ini diulang sampai terbentuk satu kelompok dengan anggota sebanyak n . Kriteria yang digunakan untuk menentukan jarak antar dua kelompok dalam penelitian ini adalah metode pautan lengkap (*complete linkage*) yaitu dengan memaksimumkan jarak antara kelompok. Jarak maksimum antara kelompok satu yang merupakan gabungan dari kelompok u dan v dengan kelompok w adalah sebagai berikut [7] :

$$d_{(uv)w} = \max \{d_{uw}, d_{vw}\} \quad (1)$$

Adapun rumus untuk menghitung jarak korelasi obyek i dan j pada k variabel adalah sebagai berikut

$$d_{ij} = 1 - \frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\left\{ \sum_{k=1}^p (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^p (x_{jk} - \bar{x}_j)^2 \right\}^{1/2}} \quad (2)$$

dimana \bar{x}_i adalah rata-rata dari variabel (x_{i1}, \dots, x_{ip})

Metode bootstrap pertama kali dikembangkan oleh Efron pada tahun

1979. *Bootstrap* adalah teknik yang dikembangkan untuk membuat inferensia statistik tertentu yang lebih sederhana dengan memerlukan komputer modern. Menurut Efron, *bootstrap* berbeda dengan inferensia statistika, metodenya sangat sederhana dan berdasar pada prosedur *resampling*. Secara singkat algoritma bootstrap dapat dinyatakan sebagai berikut [4] :

1. Sampel data D didefinisikan sebagai data sampel berukuran n yang terdiri dari $D_i = D_1, D_2, \dots, D_n$ dimana D_i merupakan vektor data pengamatan.
2. Sampel data diambil secara acak dengan pengembalian sebanyak n kali. Diperoleh data sampel baru yang didefinisikan sebagai D^* .
3. Langkah (2) dilakukan secara berulang sebanyak B kali sehingga didapatkan himpunan data *bootstrap* ($D^{*1}, D^{*2}, \dots, D^{*B}$) dimana setiap sampel *bootstrap* merupakan sampel acak yang saling independen.

Multiscale bootstrap resampling digunakan untuk menghitung nilai *probability bootstrap* (BP) *value* dan *Approximately unbiased* (AU) *p-values*. Berikut ilustrasi penerapan *Multiscale bootstrap resampling* pada *hierarchical clustering* [10]:

1. Sampel data D didefinisikan sebagai data sampel berukuran n yang terdiri dari $D_i = D_1, D_2, \dots, D_n$ dimana D_i merupakan vektor data pengamatan.
2. Sampel data diambil secara acak dengan pengembalian sebanyak n kali. Diperoleh data sampel baru yang didefinisikan sebagai D^* .
3. Langkah (2) dilakukan secara berulang sebanyak B kali sehingga didapatkan himpunan data *Bootstrap* ($D^{*1}, D^{*2}, \dots, D^{*B}$) dimana setiap sampel *bootstrap* merupakan sampel acak yang saling independen.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini merupakan publikasi Data dan Informasi Kemiskinan Kab/Kota Tahun 2011 hasil olah data triwulanan Survei Sosial Ekonomi Nasional (SUSENAS). Informasi yang didapatkan antara lain adalah 16 indikator kemiskinan di sektor ekonomi, kesehatan, dan SDM.

Variabel penelitian yang digunakan dalam penelitian ini adalah sebagai berikut :

Tabel 1. Variabel Penelitian

Ekonomi
Persentase penduduk miskin usia 15 tahun keatas yang tidak bekerja (X_1)
Persentase penduduk miskin usia 15 tahun keatas yang bekerja di sektor pertanian (X_2)
Persentase rumah tangga yang pernah membeli beras raskin (X_3)
Persentase pengeluaran per kapita untuk non makanan (X_4)
SDM
Persentase penduduk miskin usia 15 tahun keatas yang tidak tamat SD (X_5)
Angka Melek Huruf penduduk miskin usia 15-55 tahun (X_6)
Angka Partisipasi Sekolah penduduk miskin usia 13-15 tahun (X_7)
Rata-rata lama sekolah (X_8)
Kesehatan
Persentase Perempuan pengguna alat KB di rumah tangga miskin (X_9)
Persentase Balita di rumah tangga miskin yang proses kelahirannya ditolong oleh tenaga kesehatan (X_{10})
Persentase Balita di rumah tangga miskin yang telah diimunisasi (X_{11})
Persentase rumah tangga miskin dengan luas lantai perkapita $\leq 8 \text{ m}^2$ (X_{12})
Persentase rumah tangga miskin yang menggunakan air bersih (X_{13})
Persentase rumah tangga miskin yang menggunakan jamban sendiri/bersama (X_{14})
Persentase rumah tangga miskin yang mendapatkan pelayanan jamkesmas (X_{15})
Angka harapan hidup (X_{16})

Langkah-langkah Penelitian

Metode dan tahapan analisis yang akan digunakan dalam mencapai tujuan penelitian adalah sebagai berikut :

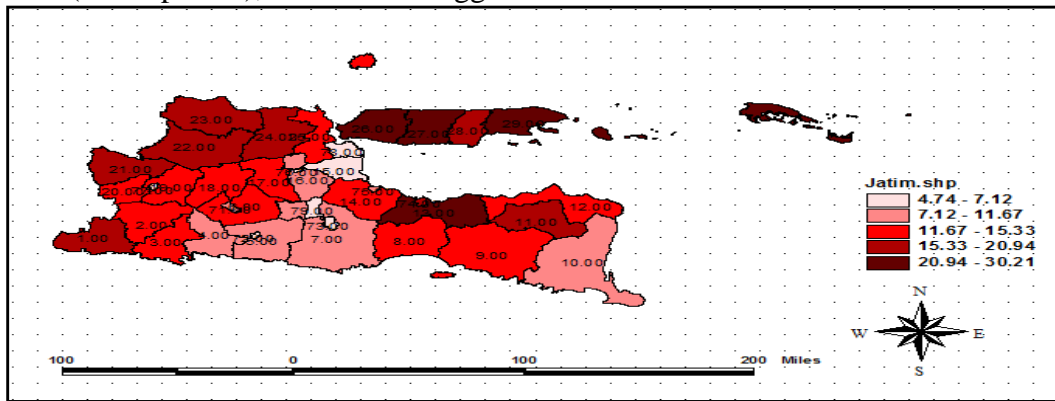
- a. Sampel data D didefinisikan sebagai data sampel berukuran n yang terdiri dari $D_i = D_1, D_2, \dots, D_n$ dimana D_i merupakan vektor data pengamatan.
- b. Sampel data diambil secara acak dengan pengembalian sebanyak n kali. Diperoleh data sampel baru yang didefinisikan sebagai D^* .
- c. Langkah (2) dilakukan secara berulang sebanyak B kali sehingga didapatkan himpunan data *Bootstrap* ($D^{*1}, D^{*2}, \dots, D^{*B}$) dimana setiap sampel *bootstrap* merupakan sampel acak yang saling independen.
- d. Untuk setiap kelompok (*cluster*) dari sampel *bootstrap* D^{*B} , nilai *probabilitybootstrap* (BP) *value* adalah $z = -\Phi^{-1}(BP)$, dimana $\Phi^{-1}(\cdot)$ adalah invers dari fungsi normal standar.
- e. *Approximately unbiased (AU) p-values* adalah $AU = -\Phi^{-1}(-v + c)$ dimana v dan c adalah estimasi parameter dari setiap kelompok (*cluster*). *Approximately unbiased (AU) p-values* memiliki bias yang lebih kecil dibandingkan dengan *probabilitybootstrap* (BP) *value*.

HASIL PENELITIAN

Kemiskinan di Jawa Timur

Pada Gambar 1 diketahui bahwa daerah kantong kemiskinan masih seputar Sampang (30.21%), Bangkalan (26.22 persen), Sumenep (23.10 persen), Pamekasan (20.94 persen), Ngawi (16.74

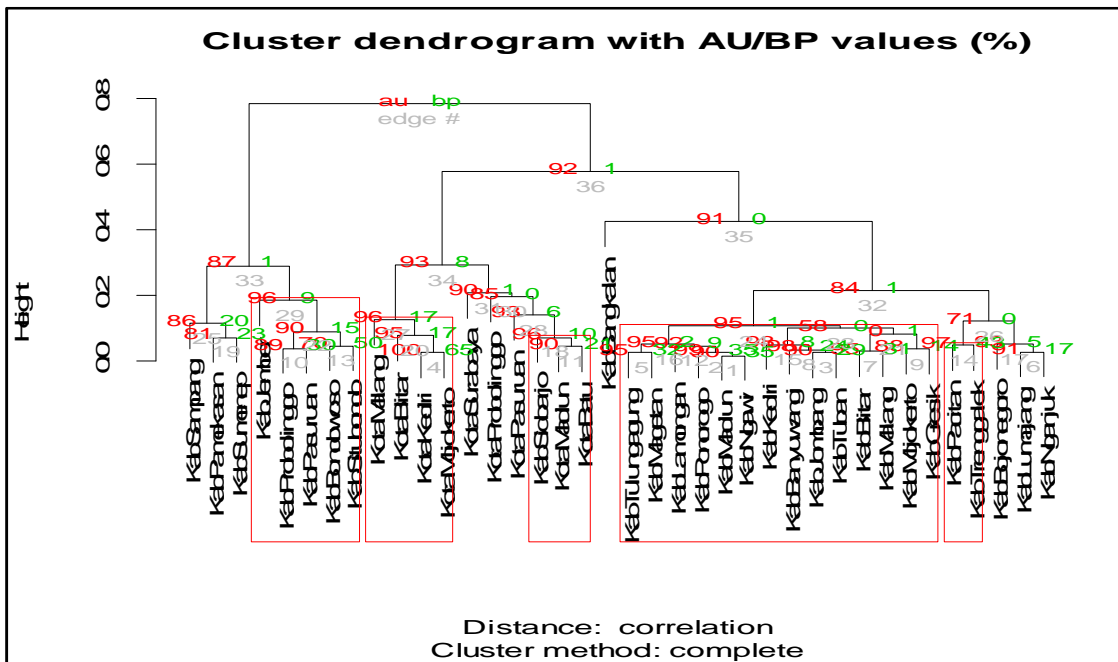
persen), Bojonegoro (17.47 persen), Tuban (18.78 persen), Lamongan (17.41), Pacitan (18.13 persen), Kota Probolinggo (17.74 persen), Probolinggo (23.48 persen), dan Bondowoso (16.66 persen).



Gambar 1. Persebaran persentase penduduk miskin di Jatim

Hierarchical Clustering Multiscale Bootstrap

Berikut disajikan analisis dan pembahasan metode hierarki *complete linkage* pendekatan *multiscale bootstrap*.

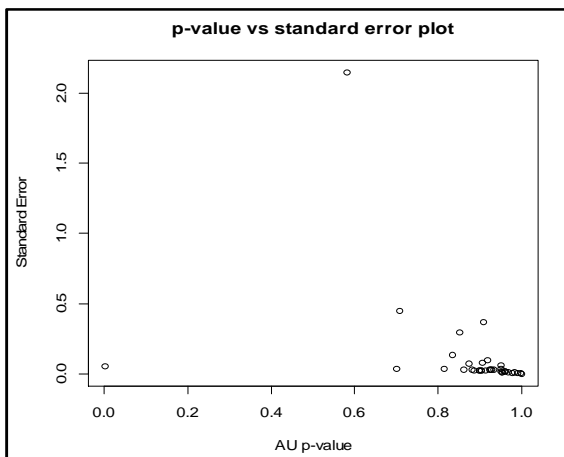


Gambar 2. Dendrogram dengan nilai AU/BP (B=500) pada masing-masing Kabupaten / Kotamadya di Provinsi Jawa Timur

Gambar 2 menunjukkan dendrogram beserta nilai *Approximately unbiased (AU) p-values* dan *probabability bootstrap (BP) value* untuk 38 Kabupaten/Kota di Propinsi Jawa Timur. Kelompok dengan AU (*approximately unbiased*) $\geq 0,95$ ditunjukkan dalam satu kelompok (*cluster*) dan dimana dapat

diinterpretasikan sebagai kelompok yang signifikan dengan tingkat signifikansi 0.05. Berdasarkan signifikansi *Approximately unbiased (AU) p-values* kelompok pertama terdiri dari Kabupaten Pacitan dan Trenggalek. Kelompok kedua adalah Kabupaten Sidoarjo, Kota. Madiun, Kota. Batu. Kelompok ketiga

terdiri dari Kabupaten Ponorogo, Kabupaten Tulungagung, Kabupaten Blitar, Kabupaten Kediri, Kabupaten Malang, Kabupaten Banyuwangi, Kabupaten Mojokerto, Kabupaten Jombang, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Tuban, Kabupaten Lamongan, Kabupaten Gresik. Kelompok keempat terdiri dari Kota Kediri, Kota Blitar, Kota Malang, Kota Mojokerto. Dan kelompok kelima yaitu terdiri dari Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan.



Gambar 3. AU P-values dengan *standard error* bootstrap 500

Kestabilan proses dalam membentuk kelompok (*cluster*) yang mempunyai kemiripan satu dengan yang lain dan signifikan didapatkan pada replikasi $B = 500$, dimana nilai standar error yang kecil dengan nilai AU/BP *value* yang mendekati nilai lebih besar dari 0,95.

KESIMPULAN

Dari hasil penelitian dapat disimpulkan bahwa terdapat lima kelompok (*cluster*) yang terbentuk dengan kemiripan satu dengan yang lain dan signifikan dengan nilai AU (*Approximately Unbiased*) $p_value \geq 0,95$. Kestabilan proses dalam membentuk

kelompok (*cluster*) tercapai pada replikasi $B = 500$. Berdasarkan signifikansi *Approximately unbiased (AU) p-values* $\geq 0,95$, kelompok kelima yang terdiri dari Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan dapat dikatakan sebagai daerah yang masih cukup tinggi persentase kemiskinannya dibandingkan dengan kelompok 1, 2, 3, dan 4.

DAFTAR PUSTAKA

[1] Anuraga, G., & Otok, B. W., 2014, *Spasial Structural Equation Modeling-Partial Least Square* untuk Pemodelan Kemiskinan di Jawa Timur, *Statistika*

[2] BPS, 2012, *Data dan Informasi Kemiskinan Kabupaten/Kota 2011*, Badan Pusat Statistik. Jakarta:

[3] Badan Pusat Statistik, 2012, *Perhitungan dan Analisis Kemiskinan Makro Indonesia 2012*

[4] Efron, B. dan Tibshirani, R.J., 1993, *An Introduction to the Bootstrap*. Chapman Hall, New York.

[5] Hair, Jr. F., Black, W.C., Babin, B.J., dan Anderson, R.E., 2010, *Multivariate Data Analysis*, Pearson Prentice Hall.

[6] Hastie, T., Tibshirani, R., dan Friedman, J., 2001, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York.

[7] Johnson, R.A. dan Wichern, D.W., 2007, *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey

[8] Kerr, M.K. and Churchill, G.A., 2001, *Bootstrapping cluster analysis: assessing the reliability of*

conclusions from microarray experiments, USA.

- [9] Shimodaira, H., 2004, *Approximately unbiased tests of regions using multistep multiscale bootstrap resampling*. Ann. Stat., 32, 2616–2641.
- [10] Suzuki, R. dan Shimodaira, H., 2004, *An application of multiscale bootstrap resampling to hierarchical clustering of microarray: How accurate are these clusters*, In proceedings by the fifteenth International Conference on Genome Informatics (GIW 2004), hal. 34