

Implementasi *Graph Clustering Algorithm Modification Maximum Standard Deviation Reduction (MMSDR)* dalam *Clustering* Provinsi di Indonesia Menurut Indikator Kesejahteraan Rakyat

Nurfidah Dwitiyanti¹, Septian Wulandari², Noni Selvia³

^{1,2,3}Program Studi Informatika, Universitas Indraprasta PGRI, Indonesia
Jl. Raya Tengah No. 80, Kel. Gedong, Kec. Pasar Rebo, Jakarta Timur 13760

Keywords:

MMSDR
Clustering
Kesejahteraan Rakyat

ABSTRACT

The population of Indonesia from year to year has increased. The increase of population must also be accompanied by the increase of economic growth. The increase of economic growth in Indonesia is marked by the minimum number of poor people. In addition, the increase of economic growth is reflected in the equitable distribution of public income in the country. Even though there are still many Indonesian people who are not yet prosperous in economic terms. To overcome the distribution of public income problem, it is necessary to cluster the characteristics of 34 provinces in Indonesia by implementing the Modification Maximum Standard Deviation Reduction (MMSDR) graph clustering algorithm. The data used are indicators of public welfare in 2017 obtained from the Central Statistics Agency. There are 9 indicators of community welfare used in this research. There are four stages in the MMSDR algorithm namely the "MST", "Subdivide", "Biggest Stepping" and "Create Clusters" processes. The results of this study can be seen from the distance between the nodes or between one province and another province that produced 22 clusters. From the cluster results obtained using the MMSDR algorithm on welfare data, there are many clusters formed with cluster members formed at most two nodes (province).

Copyright © 2020 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Nurfidah Dwitiyanti,
Program Studi Informatika,
Universitas Indraprasta PGRI,
Jl. Nangka No. 58 C, Tanjung Barat, Jagakarsa, Jakarta Selatan.
Email: nurfidah.pulungan@gmail.com

1. PENDAHULUAN

Jumlah penduduk Indonesia dari tahun ke tahun mengalami peningkatan. Menurut data Badan Pusat Statistik Indonesia melalui sensus penduduk pada tahun 2017 diketahui Indonesia memiliki jumlah penduduk sebanyak 237.641.326 jiwa [1]. Peningkatan jumlah penduduk tersebut juga harus disertai dengan meningkatnya pertumbuhan ekonomi di Indonesia. Meningkatnya pertumbuhan ekonomi di Indonesia ditandai dengan berkurangnya jumlah penduduk miskin yang ada di Indonesia. Disamping itu, peningkatan pertumbuhan ekonomi tercermin dari pemerataan pendapatan masyarakat di negaranya.

Fokus pada pertumbuhan ekonomi yang meningkat menjadi penting karena secara empirik berkolerasi positif terhadap penurunan angka kemiskinan, penciptaan lapangan kerja, serta peningkatan indikator kesejahteraan rakyat [2]. Pentingnya sektor ekonomi untuk memberantas kemiskinan telah diakui dan menunjukkan bahwa perkembangan sektor keuangan mengurangi tingkat kemiskinan di Indonesia [3]. Selain

itu, kebijakan pemerintahan dalam menentukan RAPBN dan APBN dalam pelaksanaannya masih terdapat masalah penyalahgunaan kekuasaan yang akhirnya berimbas kepada rakyat. Anggaran dana yang ada masih banyak diperuntukkan oleh pemerintah dibandingkan untuk rakyat. Padahal masih banyak rakyat Indonesia yang belum sejahtera dalam hal ekonomi. Terdapat perbedaan yang signifikan anggaran dana untuk pemerintahan dan rakyat, terlebih lagi untuk masyarakat di desa. Maka tidak heran banyak masyarakat yang berbondong-bondong merantau ke kota untuk mencari pekerjaan di ibukota atau kota yang maju lainnya. Hal ini diperkuat dengan adanya penelitian pada tahun 2011 oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) membuat prioritas kesejahteraan rakyat di suatu daerah yang berlaku di provinsi maupun di Indonesia [4]. Perbedaan mencolok dari kesejahteraan rakyat itulah yang harus diatasi oleh pemerintahan Indonesia Untuk menanggulangi hal tersebut, maka perlu adanya pengelompokkan (*clustering*) dan karakteristik 34 provinsi yang ada di Indonesia.

Terdapat beberapa jenis *clustering*, salah satunya adalah *graph clustering*. *Graph clustering* adalah jika data direpresentasikan sebagai graf, dimana objek menjadi simpul atau titik dan busur menyatakan koneksi diantara objek, maka cluster dapat didefinisikan sebagai komponen yang terhubung (grup objek yang terkoneksi satu sama lain, tetapi tidak memiliki koneksi dengan objek di luar grup) [5]. Algoritma *graph clustering* yang sering digunakan adalah algoritma yang menggunakan *Minimum Spanning Tree* (MST). Algoritma ini mampu mendeteksi *cluster* dengan batas-batas yang tidak teratur.

K-spanning Tree merupakan salah satu *graph clustering* yang berbasis MST. Pada algoritma *k-spanning Tree*, terdapat sebanyak $k - 1$ busur dengan bobot tertinggi pada MST dihapus, untuk menghasilkan sebanyak k cluster. [5]. Namun, terdapat kelemahan pada algoritma ini yaitu banyaknya cluster ditentukan oleh pengguna (*supervised learning*). Sehingga, pada penelitian ini digunakan algoritma *MST graph clustering* lainnya, dimana banyaknya *cluster* ditentukan oleh algoritma untuk menghasilkan suatu *cluster* terbaik menurut algoritma tersebut. Algoritma lainnya tersebut adalah *Modification Maximum Standard Deviation Reduction* (MMSDR) yang merupakan algoritma *graph clustering* yang berbasis MST dan termasuk ke dalam *unsupervised learning*.

Untuk mendapatkan gambaran tentang masalah kesejahteraan rakyat yang ada di provinsi Indonesia, pada penelitian ini dilakukan *clustering* provinsi menurut indikator kesejahteraan rakyat dengan mengimplementasikan *graph clustering algorithm Modification Maximum Standard Deviation Reduction* (MMSDR). Terdapat beberapa penelitian sejenis yang berkaitan dengan *clustering* dengan menggunakan MMSDR adalah penelitian yang dilakukan oleh Kamal, K., Sugeng, K.A, dan Siswantining, T pada tahun 2018 dengan melakukan *clustering* rute pesawat disalah satu perusahaan penerbangan di Indonesia untuk mengetahui kemiripan beberapa bandara serta mengelompokkannya ke dalam cluster berdasarkan jarak rute 60 bandara di Indonesia dan memperoleh 4 cluster [6]. Hal serupa juga dilakukan oleh Prastiwi, D., Sugeng, K.A., dan Siswantining, T. pada tahun 2017 dengan melakukan *clustering* rute domestik pesawat disalah satu perusahaan penerbangan di Indonesia dan memperoleh 2 cluster [5]. Namun, pada penelitian ini pengukuran kesejahteraan yang digunakan adalah Indikator Kesejahteraan Rakyat (Inkestra) pada tahun 2017 yang diperoleh dari Badan Pusat Statistik. Indikator kesejahteraan masyarakat tersebut meliputi kepadatan penduduk (KP), presentase penduduk miskin (PPM), laju pertumbuhan (LP), angka kematian bayi (AKB), angka harapan hidup (AHP), angka partisipasi sekolah (APS), tingkat partisipasi angkatan kerja (TPAK), tingkat pengangguran terbuka (TPT), serta rata-rata pengeluaran perkapita (RPP). Kesembilan indikator tersebut akan dibuat cluster yang berisi 34 provinsi di Indonesia berdasarkan kemiripan karakteristik kesejahteraan rakyat dengan menggunakan algoritma MMSDR.

2. METODE

Pendekatan penelitian yang digunakan dalam penelitian ini adalah pendekatan studi literatur deskriptif kuantitatif. Pada studi literatur, dilakukan pengumpulan bahan-bahan pustaka yang mendukung dalam penyelesaian topik penelitian ini. Sedangkan untuk pendekatan deskriptif kuantitatif mencakup mengolah, menganalisa dan menginterpretasikan data sesuai dengan kebutuhan peneliti.

Pada penelitian, data yang digunakan adalah data sekunder dengan populasi unit pengamatan 34 Provinsi di Indonesia dengan sampel penduduk Indonesia yang masuk dalam pendataan berdasarkan kriteria variabel yang digunakan pada penelitian tahun 2017. Data sekunder pada penelitian ini bersumber dari Badan Pusat Statistik Indonesia.

Variabel indikator kesejahteraan rakyat tahun 2017 yang digunakan pada penelitian ini adalah sebagai berikut [1]:

- a. Variabel kepadatan penduduk (KP)
- b. Variabel presentase penduduk miskin (PPM)
- c. Variabel laju pertumbuhan (LP)

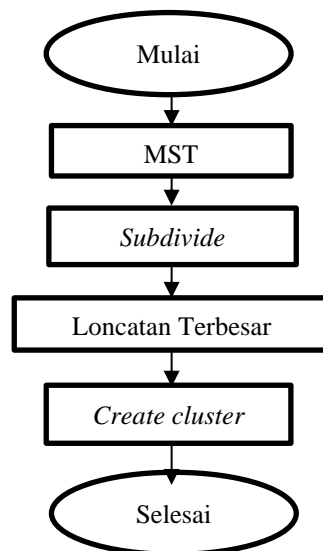
- d. Variabel angka harapan hidup (AHP)
- e. Variabel angka partisipasi sekolah (APS)
- f. Variabel tingkat partisipasi angkatan kerja (TPAK)
- g. Variabel tingkat pengangguran terbuka (TPT)
- h. Variabel rata-rata pengeluaran perkapita (RPP)

Metode yang digunakan dalam analisis kluster pada penelitian ini adalah implementasi *Graph Clustering* dengan menggunakan Algoritma *Modification Maximum Standard Deviation Reduction* (MMSDR). Metode MMSDR memiliki kelebihan yaitu banyak *cluster* yang terbentuk dapat ditentukan sendiri dan dapat mendeteksi cluster yang tidak beraturan berdasarkan berat pada tepi grafik. Pada algoritma MMSDR terbagi menjadi empat proses yang berurutan.

2.1 Algoritma MMSDR

Algoritma *Modification Maximum Standard Deviation Reduction* (MMSDR) digunakan untuk menyelesaikan masalah *graph clustering* dalam berbagai proses. Kelebihan dari penggunaan algoritma ini adalah banyak cluster yang terbentuk dapat ditentukan sendiri dan dapat mendeteksi cluster yang tidak beraturan berdasarkan berat pada tepi grafik. MMSDR dapat terbagi menjadi empat proses berurutan. Berikut ini adalah penjelasan informal mengenai apa yang dialami oleh input sampai akhirnya menjadi output [5]:

1. Misalkan G adalah input yang berupa graf berbobot sederhana, terhubung dan tak berarah.
2. Pada proses “MST” dihasilkan pohon perentang minimumnya sebut saja T .
3. Proses “Subdivide” terdiri atas beberapa iterasi. Didefinisikan $T^{(0)} = T$. Dihapus salah satu busur dari $T^{(0)}$ untuk dihasilkan $T^{(1)}$, lalu dihapus salah satu busur dari $T^{(1)}$ untuk dihasilkan $T^{(2)}$. Dan seterusnya sampai stopping condition terpenuhi. Pada akhirnya dihasilkan graf-graf hutan $T^{(0)}, T^{(1)}, \dots, T^{(r)}$ sebagai kandidat-kandidat. Pada proses ini, dihasilkan cluster-cluster.
4. Pada proses “PolyReg” dipilih salah satu kandidat, sebut saja $T^{(k)}$. Pada proses ini, dihasilkan *cluster* yang terbaik dari cluster-cluster lainnya, pada MMSDR proses “PolyReg” diganti dengan proses baru yang disebut sebagai “Loncatan Terbesar” menurut algoritma MMSDR.
5. Pada proses “Create Clusters”, dihasilkan output yang berupa himpunan simpul pada masing-masing pohon di graf hutan $T^{(k)}$. Pada proses ini, dihasilkan himpunan simpul dari cluster yang terbentuk pada proses “Loncatan Terbesar”.



Gambar 1. Bagan Alir MMSDR

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dari bulan September 2019 sampai dengan Januari 2020. Penelitian bertujuan untuk mengimplementasikan *graph clustering* dengan algoritma *Maximum Standar Deviaion Reduction* (MSDR) dalam *clustering* provinsi di Indonesia menurut indikator kesejahteraan masyarakat tahun 2017.

Penyelesaian implementasi algoritma *Maximum Standar Deviaion Reduction* (MSDR) dilakukan dengan menggunakan RStudio *programming* dengan langkah-langkah sebagai berikut:

1. Memasukkan data kesejahteraan masyarakat tahun 2017 yang diubah kedalam bentuk matrix sebut saja M dan dilakukan normalisasi serta dilakukan perhitungan dengan menggunakan *Euclidean Distance* dengan rumus:

Perhitungan jarak dilakukan dengan *Euclidian distance* yaitu [7]:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

Dengan:

d_{ik} = Jarak objek i ke centroid ke- k

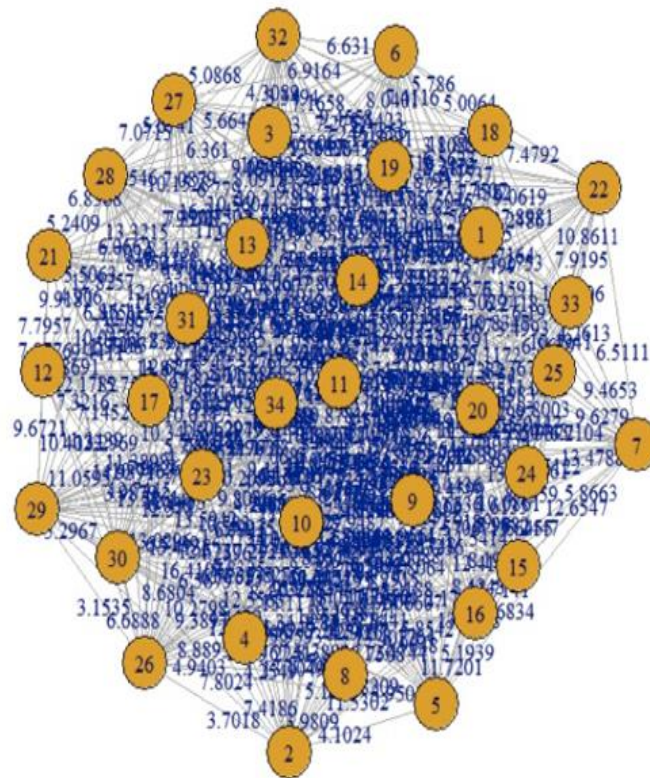
m = Banyaknya atribut (dimensi data)

x_{ij} = Koordinat dari objek ke- i pada dimensi ke- j

c_{kj} = Koordinat centroid ke- k pada dimensi ke- j .

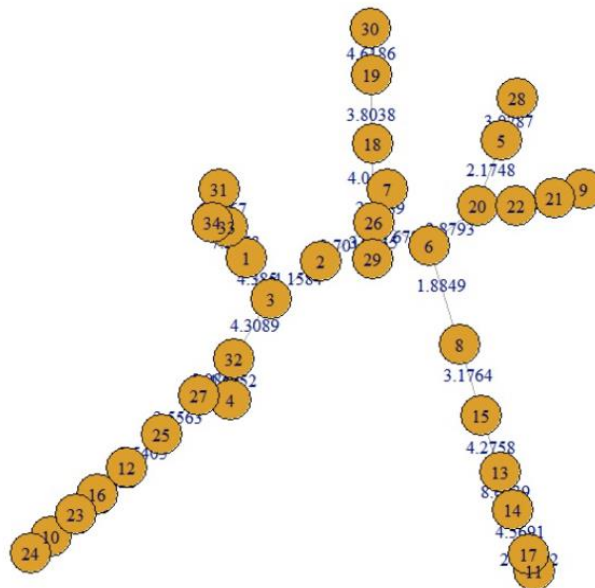
Euclidean distance dipilih karena memiliki tingkat akurasi yang lebih baik dibandingkan dengan metode lainnya [8]. Disamping itu, metode perhitungan ini mampu membandingkan jarak yang tidak diperlukan untuk mengkalkulasi akar dua sebab jarak selalu merupakan angka-angka positif [9]. Perhitungan *Euclidean distance* merupakan perhitungan jarak dari 2 buah titik dalam *Euclidean space*. Pada penelitian ini, jarak dihitung berdasarkan indikator-indikator kesejahteraan untuk masing-masing provinsi. Pada Gambar 2 menampilkan himpunan simpul pada graf lengkap *Euclidean* yang terdiri dari 34 simpul. Simpul pada graf ini menyatakan banyaknya provinsi yang ada di Indonesia. Simpul-simpul tersebut dihubungkan oleh busur-busur yang memiliki bobot, dimana bobot tersebut merupakan jarak antar satu simpul ke simpul lainnya. Sedangkan busur-busur tersebut merupakan indikator kesejahteraan untuk setiap provinsi yang ada di Indonesia. Untuk selanjutnya, matriks *Euclidean* yang merupakan informasi dari indikator-indikator kesejahteraan untuk masing-masing provinsi tersebut disimpan dalam bentuk matriks ketetanggaan. Matriks ketetanggaan menyatakan hubungan ketetanggaan antara simpul satu dengan simpul lainnya. Rincian matriks *euclidean* dapat dilihat pada Lampiran.

Pada matriks *Euclidean*, perlu diperhatikan bahwa meskipun setiap simpul saling terhubung satu dan lainnya. Namun dapat dilihat bahwa terdapat simpul yang memiliki jarak terdekat dengan simpul lainnya diantara keseluruhan simpul yang bertetangga. Misal simpul 1 dalam hal ini adalah provinsi Aceh yang memiliki jarak terdekat dengan simpul 3 yaitu provinsi Sumatera Barat (4.38503). Hal ini dapat diartikan bahwa dua provinsi tersebut memiliki kemiripan dalam indikator kesejahteraan wilayahnya. Sedangkan visualisasi dari matriks *Euclidean* yang dinyatakan dengan graf $G_{kesejahteraan}$, ditampilkan pada Gambar 2.



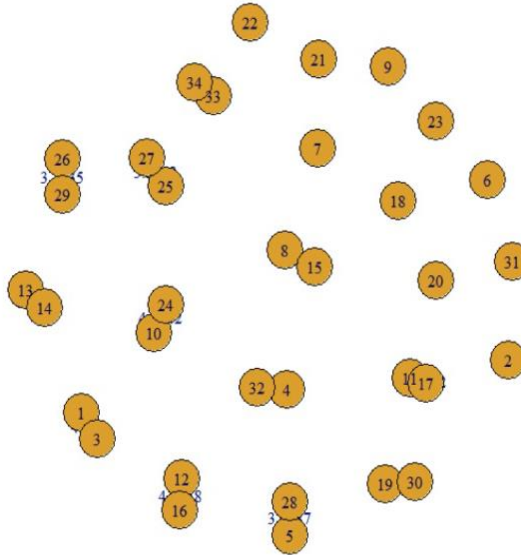
Gambar 2. Visualisasi Graf $G_{kesejahteraan}$ Provinsi di Indonesia

2. Selanjutnya algoritma MMSDR diterapkan pada Graf $G_{kesejahteraan}$, dan akan dilihat hasilnya. Secara teknis yang digunakan adalah matriks *adjacency*. Kemudian dimasukkan kedalam *script* MMSDR lalu di-*run* pada program RStudio. Hasilnya adalah sebagai berikut. Seperti dipaparkan pada bab 4 sebelumnya, MMSDR terdiri dari 4 proses yaitu proses “MST”, “Subdivide”, “Loncatan Terbesar” dan “Create Clusters”. Proses “MST” menghasilkan pohon merentang minimum, sebut saja $T_{kesejahteraan}$. Secara visualisasi ditampilkan pada Gambar 3.



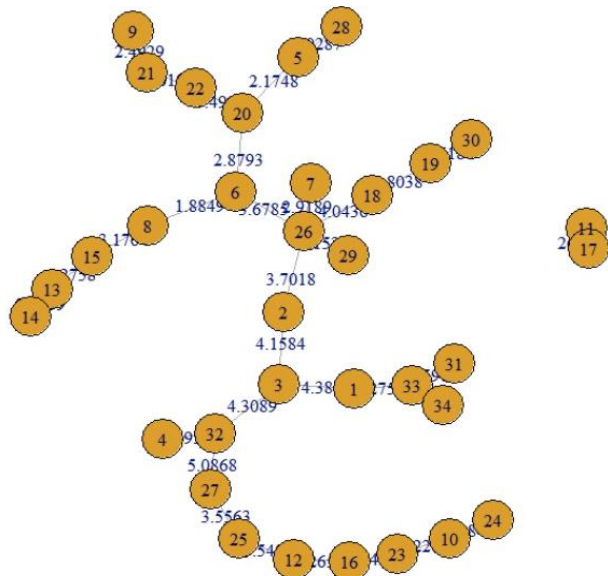
Gambar 3. Pohon merentang minimum $G_{kesejahteraan}$

3. Proses “*Subdivide*” terdiri atas beberapa iterasi. Didefinisikan $T^{(0)} = T$. Dihapus salah satu busur dari $T^{(0)}$ untuk dihasilkan $T^{(1)}$, lalu dihapus salah satu busur dari $T^{(1)}$ untuk dihasilkan $T^{(2)}$. Dan seterusnya sampai *stopping condition* terpenuhi. Pada akhirnya dihasilkan graf-graf hutan $T^{(0)}, T^{(1)}, \dots, T^{(r)}$ sebagai kandidat-kandidat. Pada proses ini, dihasilkan *cluster-cluster*. Pada proses “PolyReg” dipilih salah satu kandidat, sebut saja $T^{(k)}$. Pada proses ini, dihasilkan cluster yang terbaik dari cluster-cluster lainnya menurut algoritma MMSDR. Dalam penelitian ini proses “PolyReg” diubah menjadi proses “Loncatan Terbesar”.



Gambar 4. Visualisasi $T^{(1)}$ sebagai MMSDR pada graf $G_{kesejahteraan}$

4. Kemudian pada tahap akhir, yaitu proses “*Create Clusters*”, diperoleh *output* berupa himpunan simpul pada masing-masing pohon di graf hutan $T^{(k)}$. Pada proses ini, dihasilkan himpunan simpul dari *cluster* yang terbentuk pada proses “Loncatan Terbesar” dan dihasilkan sebanyak 22 *cluster*. Berikut ini adalah himpunan simpul (provinsi) di masing-masing *cluster*.



Gambar 5. Hasil Akhir MMSDR pada graf $G_{kesejahteraan}$

Tabel 1. Hasil Cluster

Cluster	Jumlah Anggota Cluster	Anggota Cluster
1	2	Aceh, Sumatera Barat
2	1	Sumatera Utara
3	2	Riau, Maluku Utara
4	2	Jambi, Sulawesi Tenggara
5	1	Sumatera Selatan
6	1	Bengkulu
7	2	Lampung, Jawa Timur
8	1	Kep. Bangka Belitung
9	2	Kep. Riau, Kalimantan Utara
10	2	DKI Jakarta, Bali
11	2	Jawa Barat, Banten
12	2	Jawa Tengah, DI Yogyakarta
13	1	Nusa Tenggara Barat
14	2	Nusa Tenggara Timur, Sulawesi Barat
15	1	Kalimantan Barat
16	1	Kalimantan Tengah
17	1	Kalimantan Selatan
18	1	Kalimantan Timur
19	2	Sulawesi Utara, Sulawesi Selatan
20	2	Sulawesi Tengah, Gorontalo
21	1	Maluku
22	2	Papua Barat, Papua

Penelitian sebelumnya yang telah dilakukan yaitu dengan mengelompokkan provinsi berdasarkan indikator kesejahteraan rakyat menggunakan algoritma *Self Organizing Maps* (SOM) menghasilkan 3 cluster yaitu cluster 1 terdiri dari 9 provinsi, cluster 2 terdiri dari 1 provinsi, dan cluster 3 terdiri dari 24 provinsi [10]. Sedangkan, pada penelitian ini dapat terlihat pada Tabel 1. Hasil cluster dilihat dari jarak kedekatan antar simpul atau antara provinsi satu dan provinsi lainnya. Dari hasil cluster yang diperoleh dengan menggunakan algoritma MMSDR pada data kesejahteraan terbentuk banyak cluster dengan anggota cluster yang terbentuk paling banyak dua simpul (provinsi) dan menghasilkan 22 cluster. Hal ini menunjukkan bahwa clustering provinsi berdasarkan indikator kesejahteraan rakyat dengan menggunakan algoritma MMSDR menghasilkan cluster yang lebih banyak dibandingkan dengan algoritma SOM yang hanya menghasilkan 3 cluster. Oleh karena itu, clustering provinsi berdasarkan indikator kesejahteraan rakyat dengan menggunakan algoritma MMSDR memberikan hasil yang lebih homogen di setiap cluster dibandingkan dengan algoritma SOM.

4. PENUTUP

4.1 Simpulan

Pada penelitian ini, algoritma MMSDR dengan menggunakan “Loncatan Terbesar” pada proses PolyReg-nya diterapkan pada pengelompokkan data 34 provinsi di Indonesia berdasarkan indikator kesejahteraan rakyat. Algoritma MMSDR dilakukan dengan menggunakan bantuan software RStudio. Melalui bantuan RStudio, pada proses “Create Clusters” diperoleh output berupa simpul pada masing-masing pohon di graf hutan $T^{(k)}$. Pada proses ini, dihasilkan himpunan simpul (provinsi) dari cluster yang terbentuk dengan menggunakan “Loncatan Besar” adalah sebanyak 22 cluster. Clustering provinsi berdasarkan indikator kesejahteraan rakyat dengan menggunakan algoritma MMSDR memberikan hasil yang lebih homogen di setiap cluster dibandingkan dengan algoritma SOM. Hasil cluster dilihat dari jarak kedekatan antara simpul atau antara provinsi satu dan lainnya. Pada data kesejahteraan, anggota cluster yang terbentuk paling banyak dua simpul (provinsi).

4.2 Saran

Penelitian ini hanya dilakukan analisis pengelompokkan dengan menggunakan algoritma MMSDR untuk graf Euclidean. Sedangkan untuk pengembangan lebih lanjut pada penelitian berikutnya, penulis menyarankan untuk membandingkan hasil MMSDR ini dengan metode clustering lainnya.

DAFTAR PUSTAKA

- [1] I. A. Riyadi, Sri Hartini, Zulhan R, Chairul A, Adwi H, Nia S, Putri L, “Indikator Kesejahteraan Rakyat Welfare Indicators 2017,” *Badan Pus. Stat.*, pp. 170–175, 2018.
- [2] E. C. Sugiarto, “Pemerataan Kesejahteraan Rakyat,” *Pemerataan Kesejahteraan Rakyat*, no. Sekretariat Kabinet Republik Indonesia, Jakarta, Indonesia, 2017.
- [3] M. S. A. M. Sovia Dewi, Aliasuddin, “Peran Sektor Keuangan dalam Pengentasan Kemiskinan Di Indonesia,” vol. 2, no. 1, pp. 80–93, 2014.
- [4] J. Gaussian, “JURNAL GAUSSIAN Vol . 4 , No . 4 , Tahun 2015 Halaman,” vol. 4, no. 2006, pp. 875–884, 2015.
- [5] D. Prastiwi, K. A. Sugeng, and T. Siswantining, “Modification of MSDR algorithm and ITS implementation on graph clustering,” *AIP Conf. Proc.*, vol. 1862, no. July, 2017.
- [6] K. Kamal, K. A. Sugeng, and T. Siswantining, “Implementation algorithm modification maximum standard deviation reduction in graph clustering using matrix complement as input,” *AIP Conf. Proc.*, vol. 2023, 2018.
- [7] S. Wulandari, “Clustering Kecamatan di Kota Bandung Berdasarkan Indikator Jumlah Penduduk dengan Menggunakan Algoritma K-Means,” in *Seminar Nasional Riset dan Teknologi (SEMNAS RISTEK) 2020*, 2020, pp. 128–132.
- [8] R. Rizaldi, A. Kurniawati, C. V. Angkoso, T. Informatika, and U. T. Madura, “Implementasi Metode Euclidian Distance untuk Rekomendasi Recomender System of Appareal Size in Virtual Fitting Room Based,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, pp. 129–137, 2018.
- [9] R. Fitriani and N. Rosmawanti, “Penerapan Algoritma Euclidean Distance Untuk Pemilihan Paket Internet Berdasarkan Wilayah,” *Progresif*, vol. 13, no. 1, pp. 1651–1662, 2017.
- [10] S. Wulandari and N. Dwitiyanti, “Analyze Clustering Province Based on Indicator of Social Welfare Using Self Organizing Maps (SOM) Clustering Algorithm,” in *Brawijaya International Conference on Multidisciplinary Sciences and Technology 2020*, 2020, pp. 118–124.