

Clustering Data Indeks Pembangunan Manusia di Indonesia menggunakan *Self-Organizing Maps*

Kurniati Asmar^{#1}

^{#1}Program Studi Sistem Informasi PJJ, Universitas Siber Asia
Jl. Harsono RM, Ragunan, Pasar Minggu, Jakarta Selatan

¹kurniatiasmar@lecturer.unsia.ac.id

Abstrak— Data Indeks Pembangunan manusia (IPM) di Indonesia terus mengalami kemajuan tetapi, kenyataannya penyebaran keberhasilannya terbagi ke dalam daerah-daerah tertentu saja, perubahan-perubahan tingkat indeks pembangunan dari tinggi ke rendah atau sebaliknya juga perlu diperhatikan, untuk memahami dan mempertimbangkan tindakan apa yang perlu dilakukan. Maka dari itu perlu dilakukan pengkategorian atau pengelompokan serta analisis terhadap perubahan perkembangan pada daerah-daerah yang di survei. Paper ini akan, menguraikan pengkategorian dan pengelompokannya melalui metode SOM (*Self-Organizing Maps*) berperan untuk menyoroti kesamaan pertumbuhan diseluruh daerah di Indonesia.

Kata kunci— Clustering, SOM, IPM

I. PENDAHULUAN

A. Pendahuluan

Pada bidang ilmu sosial dan ekonomi, pemerataan pembangunan dapat dilihat dari analisis indikator sosial yang menjadi alat penting untuk penilaian kesejahteraan dan kualitas kehidupan masyarakat. Data Indeks Pembangunan manusia (IPM) merupakan salah satu indikator sosial yang telah diakui oleh sebagian besar negara di dunia. Indikator utama dari IPM adalah angka harapan hidup, rata-rata lama sekolah, dan pengeluaran perkapitanya [1].

Data IPM di Indonesia terus mengalami kemajuan, ditandai dengan terus meningkatnya persentase di beberapa sektor penilaian pada data Indeks Pembangunan Manusia, yakni mencapai 70,18%, meningkat 0,63% dibandingkan tahun 2016 [2].

Akan tetapi, kenyataannya penyebaran keberhasilannya terbagi ke dalam daerah-daerah tertentu saja, sementara bagiannya masih rendah, perubahan-perubahan tingkat indeks pembangunan dari tinggi ke rendah atau sebaliknya juga perlu diperhatikan, untuk memahami dan mempertimbangkan tindakan apa yang perlu dilakukan [3].

Dari kasus tersebut, dilakukan pengkategorian atau pengelompokan serta analisis perubahan perkembangan pada daerah-daerah yang telah disurvei, beberapa peneliti telah melakukan penelitian yang terkait dengan yang akan dibahas pada laporan ini. Pengembangan indeks pembangunan komposit manusia dengan metode baru, yang

diterapkan untuk mengukur pembangunan manusia di Desa tepatnya di Kabupaten Seram Barat, Provinsi Papua menggunakan *Principal Component Analysis* (PCA), clustering menggunakan KMeans, dan regresi panel data [4]. Clustering terhadap negara-negara anggota PBB, menggunakan algoritma K-Means untuk membentuk kelompok klaster [5]. Dengan penggunaan data IPM, tanpa memperhatikan tingkat pola perubahan, SOM (*Self-Organizing Maps*) berperan untuk menyoroti kesamaan pertumbuhan diseluruh Negara, menunjukkan bahwa Negaranegara dengan berbagai karakteristik yang sama, melakukan perubahan dari waktu ke waktu, tidak saling tumpang tindih [6].

Dari tiga penelitian tersebut, yang hampir mendekati dengan penelitian yang akan dibahas pada laporan ini yakni penelitian yang dilakukan oleh Rende dan Donduran, dengan menggunakan SOM akan dilakukan pengelompokan daerahdaerah di Indonesia, dari hasil visualisasi akan dilakukan analisis untuk melihat perubahan-perubahan segmentasi, yang kemudian dapat dijadikan alat pertimbangan oleh pihak-pihak terkait.

Untuk menilai dan melakukan pemerinkatan perlu dilakukan proses *mining*, melalui salah satu metodenya yakni clustering, sehingga dapat menilai pemerataan pembangunan di suatu wilayah yang dapat dilihat dari pembangunan yang memiliki kesamaan atau kemiripan [7]. Identifikasi dan analisis data IPM yang dibagi kedalam segmentasi kelompok daerah, yang berasal dari data IPM yang terbaru dan tahun-tahun sebelumnya, serta teknik *clustering* menggunakan SOM dan analisis visualisasi data, membantu dalam analisis terhadap segmentasi kelompok-kelompok daerah yang memiliki kemiripan yang tinggi dibandingkan dengan satu atau lebih cluster lain yang sangat berbeda.

B. Dasar Teori Mutakhir

Penelitian [4] Pengembangan indeks pembangunan komposit manusia dengan metode baru, yang diterapkan untuk mengukur pembangunan manusia di Desa tepatnya di Kabupaten Seram Barat, Provinsi Papua menggunakan *principal component analysis*, clustering menggunakan K-Means, dan regresi panel data.

Penelitian [5] melakukan clustering terhadap negara-negara anggota PBB, menggunakan algoritma K-Means untuk membentuk kelompok *cluster*, sehingga variasi setiap *cluster* diminimalkan, pengurangan dimensinya menggunakan *Principal Component Analysis* (PCA), kemudian mengukur *cluster* menggunakan Hopskin statistik, dan sebagai tahap akhir penerapan algoritma K-Means untuk *clustering*.

Penelitian [6] menggunakan data *the global Human Development Index*, Penelitian ini tidak berfokus pada pemeringkatan, SOM berperan untuk menyoroti kesamaan pertumbuhan di seluruh Negara, peta yang disajikan pada dasarnya sebagai pengintai figur peringkat IPM ke dalam kelompok dengan tingkat yang sama dari komponen HDI. Hasil penelitian menunjukkan bahwa para tentangga, yakni negara-negara dengan berbagai karakteristik yang sama, melakukan perubahan dari waktu ke waktu dan tidak selalu tumpang tindih.

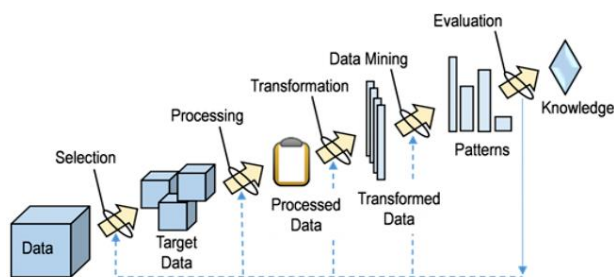
C. Metode Penelitian

1. Data

TABEL I
INFORMASI DATA

Sumber Data	https://data.go.id/dataset/indeks_pembangunan-manusiaipm
Deskripsi Singkat	Dataset ini berisi mengenai Indeks Pembangunan Manusia (IPM) di Indonesia dari tahun 2004 – 2012, dengan level penyajian provinsi, cakupan Nasional.
Format Data	Variabel: 7; Banyaknya Data: 9941; Kelas Data: Data Frame

2. Tahapan kegiatan



Gambar. 1 Tahapan kegiatan

a. Data Selection

Sekumpulan data IPM dari tahun 2004-2012 diseleksi sebelum tahap penggalian informasi kemudian disimpan dalam satu tempat terpisah dari basis data operasional.

b. Pra-proses Data

Pra-proses data yang dilakukan adalah data cleaning dengan menghapus data yang *missing value*. Karena data masuk dalam kategori *data not applicible* atau tidak

berlaku. Sebelumnya dilakukan identifikasi terlebih dahulu, ditemukan bahwa penyebab *missing value* adalah adanya beberapa daerah yang belum terbentuk di tahun 2004 hingga 2008. Untuk menangani hal ini salah satunya cara adalah *remove* karena, data IPM adalah data yang tidak dapat dirata-ratakan, dimoduskan, dan lain sebagainya.

c. Transformasi

Transformasi data yang dilakukan adalah *normalization by decimal scaling* yakni normalisasi yang didasarkan pada skala desimal, dengan rumus sebagai berikut:

$$\frac{\theta'}{10^j} \text{ dimana } j \text{ adalah bilangan bulat terkecil seperti } \text{Max}(|\theta'|)$$

d. Data Mining

Pada tahapan data mining akan digunakan teknik atau metode SOM karena, data yang diolah merupakan data yang bersifat *unsupervised learning*. Yakni data yang tidak memiliki target atau kelas.

e. Evaluasi

Dari pola informasi yang dihasilkan dari proses data mining ditampilkan bentuk yang mudah dipahami oleh pihak yang berkepentingan dengan dilakukan pemeriksaan apakah bertentangan dengan fakta atau tidak.

3. Lingkungan Pengembangan

TABEL II
INFORMASI LINGKUNGAN PENGEMBANGAN

Perangkat keras	Laptop Asus AMD Quad Core A8
Perangkat lunak	Windows 10 Microsoft Excel 2010 R Studio 3.5.0

D. Hasil dan Analisa

1. Hasil pra-proses

Setelah dilakukan Pra-Proses data yang digunakan menjadi:

TABEL II
HASIL PRA-PROSES

Banyak data	49x7
Variabel	7

2. Hasil transformasi

Berikut adalah gambaran 6 data terpilih dari dataset yang diolah sebelum ditransformasi:

TABEL II
DATASET SEBELUM DITRANSFORMASI

	Harapan hidup	Pendidikan	Pengeluaran perkapita
1	62.91	8.30	617.10

2	64.69	7.74	608.22
3	66.82	8.28	604.59
4	69.19	9.34	596.01
5	69.63	8.49	586.29
6	69.53	9.44	615.52

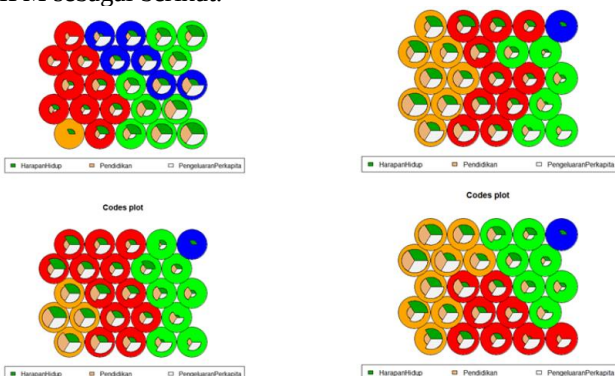
Mengapa dilakukan transformasi, transformasi dilakukan untuk membantu algoritma dapat berjalan dengan benar, sementara dataset yang diolah memiliki data satuan, puluhan, dan ratusan. Dengan menggunakan perintah “scale” di R maka ditemukan transformasi data sebagai berikut:

TABEL II
HASIL TRANSFORMASI DATA

	Harapan hidup	Pendidikan	Pengeluaran perkapita
1	-1.9330751	80.39799334	-0.2301965
2	-1.2947322	0.04979296	-0.6713532
3	-0.5308725	0.38555761	-0.8516910
4	0.3190560	1.04465120	-1.2779438
5	0.4768486	0.51613276	-1.7608316
6	0.4409867	1.10682984	-0.3091874

3. Hasil mining data

Dengan menggunakan library kohonen yakni library milik SOM di R, sesuai dengan nama penciptanya yakni Teuvo Kohonen, dapat dijalankan perintah-perintah SOM di R Studio dengan mudah, dengan syarat R Studio versi 3.4.5 ke atas, karena jika tidak beberapa fungsi SOM di R tidak akan berfungsi dengan baik. Pada penelitian ini, data di kelompokkan ke dalam dimensi 5x5. Alasannya tidak lain adalah untuk memudahkan visualisasi, dan analisis daerah-daerah mana saja yang masuk dalam cluster tersebut. Dengan menggunakan hierarchycal clustering, maka perhitungan distance pada penelitian ini menggunakan Euclidean. Sehingga dengan penggunaan plot type code dapat dilihat segmentasi map dari dataset IPM sebagai berikut:



Gambar. 2 Segmentasi Map dari dataset IPM

4. Analisis data transition

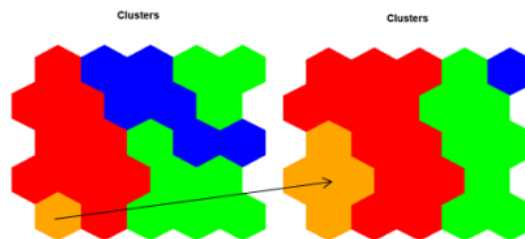
Dengan deskripsi warna, bahwa hijau menunjukkan cluster 1, biru sebagai cluster 2, merah sebagai cluster 3, dan oranye sebagai cluster 4. Mengikuti aturan baru pengelompokan dari Badan Pusat Statistik yang membagi

kelompok segmentasi wilayah di Indonesia menjadi empat bagian, segmen pertama sebagai segmen yang sangat rendah, segmen ke dua rendah, segmen ke tiga menengah, dan ke empat tinggi.

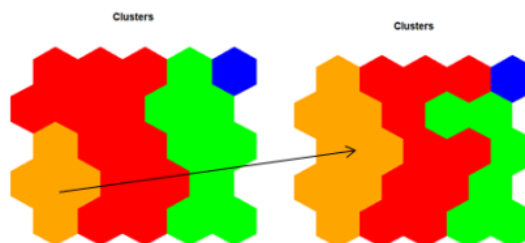
a. 2009-2010

Dari hasil visualisasi pada gambar 3, dapat diperhatikan perubahan bahwa cluster 4 memperlihatkan transisi yang sangat mencolok dari sebelumnya menempati 1 dimensi atau node, menjadi 4 node. Menunjukkan bahwa peningkatan segmen kategori sangat rendah meningkat di tahun 2010 dibanding tahun 2009.

b. 2010-2011



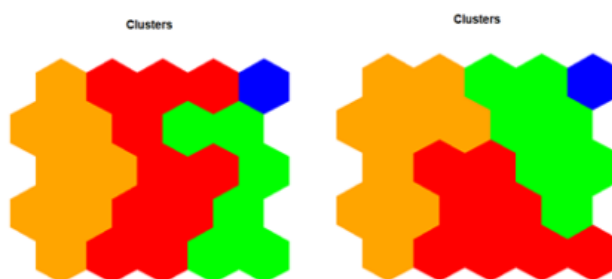
Gambar. 3 Visualisasi 2009-2010



Gambar. 4 Visualisasi 2010-2011

Sama seperti di tahun 2009-2010, Gambar 4. menunjukkan transisi yang sangat mencolok adalah segmen cluster oranye yang menunjukkan peningkatan segmen kelas rendah meningkat di tahun 2010, sementara lainnya dengan tingkat presentase yang sama.

c. 2011-2012



Gambar. 5 Visualisasi 2011-2012

Dari hasil visualisasi pada gambar 5, dapat diperhatikan bahwa presentase perpindahan segmentasi kelompok wilayah tidak berubah cukup signifikan.

E. Kesimpulan

Dari penelitian pengolahan data IPM menggunakan SOM didapatkan pemodelan untuk melihat segmentasi map dari data IPM, dalam pengerjaan laporan ini ditemukan banyak kendala, seperti penentuan tabel segmentasi per cluster, diakibatkan banyaknya data dan terbatasnya waktu yang dimiliki oleh peneliti. Untuk kedepannya diharapkan pemodelan ini dapat diterapkan di Shyni R. Studio sehingga dapat berbasis web, yang kemudian dapat dengan mudah dimanfaatkan oleh pihakpihak terkait dalam menjalankan tugasnya.

REFERENSI

- [1] A. Otoi, E. Titan, and R. Dumitrescu, "Are the variables used in building composite indicators of well-being relevant? validating composite indexes of well-being," *Ecological indicators*, vol. 46, pp. 575–585, 2014.4
- [2] B. P. Statistik, "Berita resmi statistik tentang indeks pembangunan manusia (ipm)," BPS, 2017.
- [3] J. A. F. Costa, A. P. V. Pinto, J. R. de Andrade, and M. G. de Medeiros, "Clustering of regional hdi data using self-organizing maps," in *Computational Intelligence (LA-CCI), 2017 IEEE Latin American Conference on*. IEEE, 2017, pp. 1–6.
- [4] A. Amaluddin, R. W. Payapo, A. A. Laitupa, M. R. Serang et al., "A modified human development index and poverty in the villages of west seram regency, maluku province, indonesia," *International Journal of Economics and Financial Issues*, vol. 8, no. 2, pp. 325–330, 2018.
- [5] S. Mylevaganam, "The analysis of human development index (hdi) for categorizing the member states of the united nations (un)," *Open Journal of Applied Sciences*, vol. 7, no. 12, p. 661, 2017.
- [6] S. Rende and M. Donduran, "Neighborhoods in development: human development index and self-organizing maps," *Social Indicators Research*, vol. 110, no. 2, pp. 721–734, 2013.
- [7] D. Septiafani and A. Alamsyah, "Clustering data indeks pembangunan manusia dan produk domestik bruto untuk indentifikasi pemerataan pembangunan di indonesia," *eProceedings of Management*, vol. 3, no. 3, 2016.