

Pengelompokan Provinsi di Indonesia Menggunakan Algoritma *Partitioning Around Medoids* (Pam) Terhadap Indikator Pembentuk Indeks Pembangunan Manusia (Ipm) Tahun 2020

Reni Anggraini¹, Sri Subanti², Sri Sulistijowati Handajani³

^{1,2,3} Program Studi Statistika, Fakultas Ilmu Pengetahuan Alam, Universitas Sebelas Maret
Email: reniangraini@student.uns.ac.id¹, sri_subanti@yahoo.co.id², rr_ssh@staf.uns.ac.id³

Abstrak. Indeks pembangunan manusia merupakan suatu indikator yang dapat menggambarkan perkembangan pembangunan manusia dan menjadi tolak ukur ekonomi suatu negara. Nilai IPM di provinsi Indonesia pada tahun 2020 mengalami penurunan dibandingkan pada tahun sebelum, hal ini juga akan berpengaruh pada pertumbuhan ekonomi. Pada penelitian ini akan dilakukan pengelompokan provinsi di Indonesia berdasarkan indikator pembentuk indeks pembangunan manusia pada tahun 2020. Algoritma yang akan digunakan pada penelitian ini adalah *partitioning around medoids*. Algoritma *partitioning around medoids* dilakukan dengan memilah-milah data yang dianalisis ke dalam cluster-cluster yang ada. Data yang digunakan dalam penelitian ini merupakan indikator pembentuk IPM, yaitu usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan paritas daya beli yang diperoleh dari web resmi Badan Pusat Statistika Indonesia. Hasil penelitian dari pengelompokan 34 provinsi yang ada di Indonesia berdasarkan indikator pembentuk indeks pembangunan manusia pada tahun 2020 diperoleh 2 cluster sebagai cluster optimal, yaitu cluster pertama dengan provinsi yang memiliki nilai usia harapan hidup rendah dan cluster kedua dengan provinsi yang memiliki nilai harapan lama sekolah rendah. Berdasarkan hasil tersebut maka anggaran pada provinsi yang masuk ke cluster pertama dapat lebih difokuskan untuk meningkatkan usia harapan hidup dan pada cluster kedua dapat lebih difokuskan untuk meningkatkan harapan lama sekolah agar pertumbuhan ekonomi di Indonesia lebih baik sehingga tidak terjadi resesi ekonomi.

Kata Kunci: *ekonomi, indeks pembangunan manusia, partitioning around medoids*

Abstract. *The human development index is an indicator that can convert, human development and become a country's measure. Bank Indonesia the central bank/bi's decision to raise its benchmark interest rate by 25 basis points to 8.25 percent would be lower than the previous quarter of this year, he said. The algoritma to be used on this research is partitioning Around medoids The algorithm partitioning around medoids is done by sifting through data Which is analyzed into the cluster-clusters that exist. The data used in mni suppresses the ipm phasing indicator, which is the biologic age. The rupiah's current exchange rate against rp9,100 per dollar in the Jakarta interbank spot market on Tuesday afternoon strengthened to rp9,310/9,329 per dollar in the Jakarta interbank spot market on Tuesday. The result of a 34 proxies grouping in Indonesia is based on indicators of human development in 2020, at two optimum clusters, In the first half of 2007, the company's total assets in the first quarter of 2007 fell to rp2.1 trillion from rp2.1 trillion in the same period last year*

Keywords : *Economics, development index, partitioning around medoids*

PENDAHULUAN

Pembangunan bertujuan memberikan manfaat bagi kemajuan suatu wilayah. Keberhasilan pembangunan diukur dengan beberapa parameter, dan yang paling sering

digunakan adalah Indeks Pembangunan Manusia (IPM).

Pandemi COVID-19 di Indonesia membawa pengaruh terhadap pembangunan manusia. Hal ini dapat dilihat dari perlambatan pertumbuhan IPM pada tahun

2020 dibandingkan dengan tahun-tahun sebelumnya. IPM di Indonesia pada tahun 2019 tumbuh sebesar 0.74 % sedangkan tahun 2020 tumbuh sebesar 0.03%. Apabila kondisi ini terus berlangsung akan berdampak pada perekonomian di Indonesia dan dapat diambang resesi ekonomi.

Data mining merupakan suatu proses ekstraksi informasi berguna dan potensial dari sekumpulan data yang terdapat secara implisit dalam suatu basis data. Teknik yang paling sering digunakan pada *data mining* merupakan pengelompokan (*clustering*). Secara umum metode pengelompokan dibagi menjadi dua metode, yaitu *hierarchical clustering* dan *partitioned clustering*.

Metode yang paling sering digunakan adalah metode *partitioned* dengan algoritma *partitioning around medoids* (PAM) dan *Clustering Large Application* (CLARA). Perbedaan algoritma PAM dan CLARA terletak pada jumlah data yang digunakan. Algoritma PAM digunakan pada data yang tidak berukuran terlalu besar, sedangkan CLARA untuk data yang berukuran besar. Menurut Tri Binty (2019) algoritma PAM menggunakan objek pada kumpulan objek yang mewakili sebuah cluster, objek tersebut dinamakan medoid. Medoid merupakan objek yang letaknya terpusat di dalam suatu cluster sehingga robust terhadap pencilon. Data yang digunakan pada penelitian ini terjadi masalah *outlier* dan hanya berjumlah 34 provinsi sehingga lebih cocok menggunakan algoritma PAM pada penelitian ini.

Penelitian yang relevan dengan penelitian yang diambil adalah penelitian dari Garibaldi dan Hidayat (2013) . Dalam penelitiannya telah dianalisis hubungan IPM dan pertumbuhan ekonomi. Berdasarkan *scatter plot* menunjukkan bahwa terdapat hubungan positif antara IPM dan pertumbuhan ekonomi. Kemudian berdasarkan penelitian yang sudah dilakukan oleh Garibaldi dan Hidayat maka dilakukan pengembangan penelitian untuk menganalisis keadaan pertumbuhan ekonomi di Indonesia dengan melihat data indikator pembentuk IPM. Sangga (2018) pada penelitiannya melakukan perbandingan

metode *K-Means* dan *Partitioned*. Hasil yang diperoleh bahwa metode *Partitioned* lebih baik dibandingkan dengan *K-Means*. Berdasarkan hasil dari penelitian Sangga maka pada penelitian ini digunakan metode *partitioned*. Berdasarkan uraian dan penelitian sebelumnya, penelitian ini menerapkan metode *Partitioned* dengan mengelompokkan provinsi di Indonesia berdasarkan indikator pembentuk IPM, yaitu UHH, HLS, RLS dan paritas daya beli pada tahun 2020 menggunakan algoritma PAM. Dari penelitian ini diharapkan dapat menjadi acuan pemerintah dalam mengambil kebijakan di suatu wilayah berdasarkan karakteristik yang dimiliki wilayah tersebut agar pertumbuhan ekonomi di Indonesia membaik dan tidak terjadi resesi ekonomi.

METODE

Data yang digunakan merupakan data sekunder dari web resmi Badan Pusat Statistika (BPS) Indonesia. Variabel yang digunakan adalah indikator pembentuk IPM, yaitu UHH, HLS, RLS, dan paritas daya beli di Indonesia pada tahun 2020. Data yang diperoleh berjumlah 34 provinsi yang akan dilakukan pengelompokan menggunakan algoritma PAM. Berikut adalah langkah-langkah yang dilakukan dalam penelitian ini:

1. Mengunduh data pada alamat *website* <https://www.bps.go.id>;
2. Melakukan standarisasi data untuk menyamakan satuan pada masing-masing variabel;
3. Mendeteksi adanya *outlier* menggunakan *boxplot*;
4. Mencari jumlah cluster optimal dengan metode *silhouette*;
5. Melakukan pengelompokan provinsi di Indonesia berdasarkan indikator pembentuk IPM pada tahun 2020 menggunakan algoritma PAM.
6. Melakukan profiling *cluster*.

HASIL DAN PEMBAHASAN

Statistika Deskriptif

Analisis statistika deskriptif dari variabel indikator pembentuk IPM disajikan pada Tabel 1. dan Tabel 2.

Tabel 1. Statistika Deskriptif Data Indikator Pembentuk IPM di Indonesia pada tahun 2020

	UHH	HLS	RLS	Paritas daya beli
Min	65.06	11.08	6.69	6954000
Kuartil 1	68.67	12.72	7.93	9332000
Median	69.96	13.03	8.71	10536000
Rata-rata	70.04	13.14	8.649	10686000
Kuartil 3	71.53	13.57	9.115	11489000
Max	74.99	15.59	11.13	18227000

Tabel 1. menunjukkan bahwa variabel paritas daya beli memiliki rentang maksimum dan minimum yang cukup jauh karena satuan pada variabel tersebut berbeda dengan satuan variabel yang lain. Oleh karena itu, perlu dilakukan standarisasi data untuk menyamakan satuan pada keempat variabel tersebut sebelum melakukan pengelompokan.

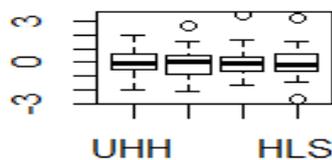
Tabel 2. Statistika Deskriptif Data Standarisasi Indikator Pembentuk IPM di Indonesia pada tahun 2020

	UHH	HLS	RLS	Paritas daya beli
Min	-1.97	2.75	2.11	-1.71
Kuartil 1	-0.54	0.56	0.77	-0.62
Median	-0.03	0.14	0.06	-0.07
Rata-rata	0	0	0	0
Kuartil 3	0.59	0.58	0.5	0.37
Max	1.96	3.29	2.67	3.45

Tabel 2. menunjukkan statistika deskriptif untuk data yang sudah distandarisasi dan didapatkan bahwa rata-rata data tersebut menjadi nol. Hal ini menunjukkan bahwa keempat variabel tersebut telah memiliki skala data yang sama.

Deteksi Outlier

Salah satu cara yang dapat dilakukan untuk mendeteksi outlier adalah dengan melihat *boxplot*



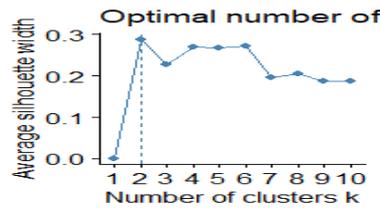
Gambar 1. *Boxplot* Data yang Sudah Distandarisasi

Gambar 1. menunjukkan bahwa terdapat data yang berada di luar plot, yang berarti bahwa data yang digunakan terdapat outlier. Pada kasus ini outlier pada data dibiarkan.

Penentuan Jumlah Cluster Optimum

Metode yang digunakan dalam penentuan jumlah cluster optimum pada analisis ini

adalah metode *silhouette*. Nilai koefisien *silhouette* yang digunakan untuk penentuan jumlah cluster optimum diperoleh dari *output software RStudio* dan ditunjukkan pada **Gambar 2.**



Gambar 2. Jumlah Cluster Optimum

Gambar 2. menunjukkan bahwa berdasarkan nilai koefisien *silhouette* diperoleh jumlah *cluster* optimum sebanyak 2 *cluster*.

Pengelompokan menggunakan Algoritma Partitioning Around Medoids

Hasil pengelompokan provinsi di Indonesia berdasarkan indikator pembentuk IPM pada tahun 2020 terdapat 27 provinsi yang masuk ke dalam *cluster* pertama, yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep.Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, D I Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi

Tenggara, dan Maluku Utara. Provinsi pada *cluster* kedua terdiri dari 7 provinsi lainnya, yaitu Provinsi Nusa Tenggara Barat, Nusa Tenggara Timur, Gorontalo, Sulawesi Barat, Maluku, Papua Barat, dan Papua.

Profiling Hasil Cluster

Profiling merupakan interpretasi hasil pengelompokan. Untuk menginterpretasi profil hasil pengelompokan dapat dilakukan menggunakan rata-rata setiap klaster pada setiap variabel. Tabel 4. menunjukkan nilai rata-rata untuk masing-masing variabel. Pada *cluster* pertama variabel HLS mempunyai nilai rata-rata tertinggi dan nilai UHH mempunyai nilai rata-rata terendah dibandingkan dengan variabel lain. Pada *cluster* kedua variabel UHH mempunyai nilai rata-rata tertinggi dan HLS mempunyai rata-rata terendah.

Tabel 4. Nilai rata-rata untuk masing-masing klaster

Cluster	UHH	HLS	RLS	Paritas daya beli
1	-1.46335	-0.24299	-0.87314	-0.90767
2	0.379387	0.062998	0.226368	0.235321

KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa pengelompokan provinsi di Indonesia berdasarkan data indikator pembentuk IPM pada tahun 2020 dikelompokkan menjadi 2 *cluster*. Kemudian dilihat dari nilai rata-rata *cluster* untuk masing-masing variabel didapatkan bahwa provinsi pada *cluster* pertama memiliki UHH yang rendah sehingga pada provinsi ini anggaran pemerintah lebih baik difokuskan untuk meningkatkan UHH. Sedangkan provinsi pada *cluster* kedua memiliki nilai HLS yang rendah sehingga

pada anggaran pemerintah pada provinsi ini lebih baik difokuskan untuk meningkatkan HLS. Diharapkan dengan pemberian anggaran yang tepat pembangunan manusia di Indonesia akan lebih merata dan tepat sasaran sehingga akan berdampak pada pertumbuhan ekonomi yang lebih baik dan Indonesia dapat jauh dari ambang resesi ekonomi. Untuk penelitian selanjutnya diharapkan dapat

dilakukan pengelompokan dengan algoritma pengelompokan lainnya dengan data yang sama untuk dibandingkan dengan hasil pengelompokan menggunakan algoritma PAM yang sudah dilakukan pada penelitian ini.

REFERENSI

- Badan Pusat Statistika. 2020. Indeks Pembangunan Manusia. Indonesia : Badan Pusat Statistika
- Binty, T. (2018, Juli 8). *Algoritma PAM*. Retrieved from Medium.com: <https://medium.com/@tribinty/k-medoids-partitioning-around-medoids-pam-non-hierarchical-clustering-with-r-9d0af590bbc0>
- Muhammad Garibaldi, Paidi Hidayat. (2013). Analisis Hubungan IPM dan Pertumbuhan Ekonomi Kabupaten/Kota di Provinsi Sumatra Utara. <https://media.neliti.com/media/publications/14820-ID-analisis-hubungan-ipm-dan-pertumbuhan-ekonomi-kabupatenkota-di-provinsi-sumatera.pdf>.
- Riadi, M. (2017, September 21). *Pengertian, Fungsi Proses dan Tahapan Data Mining*. Retrieved from [kajianpustaka.com](https://www.kajianpustaka.com/2017/09/data-mining.html): <https://www.kajianpustaka.com/2017/09/data-mining.html>
- Sangga, V. A. (2018). Perbandingan Algoritma K-Means dan Algoritma K-Medoids dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah Tahun 2015. Yogyakarta: Universitas Islam Indonesia