

ALGORITMA PARTITIONING AROUND MEDOIDS (PAM) DENGAN METODE DAVIES BOULDIN INDEX UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN KOMPONEN PEMBENTUK INDEKS PEMBANGUNAN GENDER

Vial Virgie Arganata¹, Hardian Oktavianto², Habibatul Azizah Al Faruq, M.Pd.³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Jln. Karimata No. 49 Jember Kode Pos 68121

email: virgiearganata@gmail.com

ABSTRAK

Indeks Pembangunan Gender (IPG) merupakan indikator dari sebuah pencapaian pembangunan manusia. IPG memiliki 4 komponen pembentuk yaitu Angka Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran per Kapita. IPG digunakan untuk mengukur perbedaan pembangunan antara laki-laki dan perempuan, namun permasalahan terjadi karena pada kenyataannya masih terjadi timpang tindih antara status sebagai laki-laki maupun perempuan, hal ini dibuktikan dengan masih adanya provinsi di Indonesia yang memiliki tingkat IPG tinggi dan rendah. Maka dari itu, dibutuhkan pengelompokan provinsi di Indonesia berdasarkan komponen pembentuk IPG. Salah satu algoritma *cluster* yang dapat digunakan adalah *Partitioning Around Medoids (PAM)*, data yang digunakan yaitu data provinsi di Indonesia berdasarkan komponen pembentuk IPG pada tahun 2010 sampai dengan 2019. Dari proses pengujian dimulai dari 2 *cluster* hingga 10 *cluster*, menghasilkan *cluster optimum* yang berada pada 3 *cluster* berdasarkan nilai terendah yang dihitung dengan metode *Davies Bouldin Index (DBI)*. Pada *cluster* 1 diperoleh 2 anggota provinsi, *cluster* 2 diperoleh 1 anggota provinsi, dan *cluster* 3 diperoleh 31 anggota provinsi.

Kata Kunci: Komponen Pembentuk IPG, *clustering*, *Partitioning Around Medoids (PAM)*, *Davies Bouldin Index*.

1. PENDAHULUAN

Indeks Pembangunan Gender (IPG) merupakan indikator dari sebuah capaian pembangunan manusia. Makna dari gender menekankan pada perbedaan laki-laki dan perempuan, perbedaan ini tidak menjadi masalah apabila terdapat kesetaraan pada keduanya. Tetapi kenyataannya terjadi ketidakadilan, dimana pada salah satu gender telah mengalami diskriminasi. Untuk menghapus ketidaksetaraan tersebut maka perlu adanya kesamarataan *gender* dalam bangsa dan negara (Badan Pusat Statistik, 2015).

IPG dipublikasikan pertama kali pada 5 tahun setelah Indeks Pembangunan Manusia (IPM) oleh UNDP (United Nations Development Programme) di tahun 1995 sampai dengan 2009. Pada prosedur lawas IPG semata-mata hanya menghitung kesenjangan pada tiap komponen IPM untuk kedua *gender*. IPG berhenti melakukan perhitungan di tahun 2010 sampai 2013. tetapi kembali dijalankan perhitungan menggunakan metode yang diperbarui yang disesuaikan dengan perubahan dalam IPM. Untuk menyempurnakan metode sebelumnya, setelah diperbarui, IPG mengukur secara langsung ketimpangan antar gender dalam pencapaian IPM (Nazmi & Jamal, 2018).

IPG bisa digunakan untuk mengukur ketimpangan pembangunan diantara laki-laki dan perempuan, permasalahan timbul karena pada kenyataannya masih terjadi timpang tindih antara status sebagai laki-laki atau perempuan, ini diindikasikan atas masih adanya provinsi di Indonesia yang memiliki tingkat IPG tinggi dan rendah. Maka dari itu perlu adanya pengelompokan berdasarkan provinsi di Indonesia sebagai bahan evaluasi. Tujuannya agar indikator tersebut dapat dijadikan sebagai alat pengontrol hasil kesenjangan pembangunan. Pada penelitian sebelumnya oleh Azuri, Zulhanif & Pontoh (2016) telah melakukan penelitian tentang “Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Dengan Menggunakan Algoritma Bisecting K-Means”. Pada penelitian ini menggunakan 6 variabel yaitu angka harapan hidup, rata-rata

lama sekolah, harapan lama sekolah, pengeluaran perkapita, keterwakilan di parlemen, dan profesi sebagai tenaga (manajer, profesional administrasi, dan teknisi). Dari penelitian tersebut menghasilkan 3 *cluster* yang terbentuk baik pada laki-laki ataupun perempuan berdasarkan kabupaten/kota. dengan menggunakan metode Elbow yang digunakan untuk menentukan jumlah *cluster* dan Silhouette Coefficient sebagai validasi jumlah *cluster* untuk mengukur seberapa baik hasil pengelompokan, dengan hasil *cluster* yang terbentuk memiliki nilai silhouette yang masih tergolong lemah.

Penelitian lain yang dilakukan oleh Kamila, Khairunnisa & Mustakim (2019) dengan judul “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau” menggunakan algoritma K-Means dan K-Medoids atau biasa disebut dengan PAM untuk mengelompokkan data transaksi serta menggunakan metode Davies Bouldin Indeks (DBI) dalam menganalisa kualitas hasil *cluster* untuk menentukan *cluster* optimum. Dikarenakan masih adanya ketidaksetaraan gender yang terjadi maka dilakukan penelitian terhadap data IPG yang merupakan hasil pembagian dari IPM Perempuan dan IPM laki-laki, dimana IPG terdapat 4 variabel pembentuk yang meliputi angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, dan pengeluaran perkapita di 34 provinsi Indonesia dari tahun 2010 sampai dengan tahun 2019. Algoritma yang digunakan adalah algoritma PAM untuk mengelompokkan provinsi di Indonesia dan menggunakan metode DBI untuk melakukan validasi pada hasil *clustering*. Maka penelitian ini dibuat dengan judul “Algoritma Partitioning Around Medoids (PAM) Dengan Metode Davies Bouldin Index Untuk mengelompokkan Provinsi di Indonesia Berdasarkan Komponen Pembentuk Indeks Pembangunan Gender”.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining adalah proses dalam mencari pola atau informasi dalam data yang dipilih atau disebut juga suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database dengan menggunakan metode atau teknik tertentu (Kamila, dkk. 2019).

2.2 Clustering

Metisen & Sari (2015) *Clustering* merupakan cara yang dipakai untuk memisah kumpulan data menjadi sejumlah kelompok berlandaskan persamaan yang telah ditetapkan lebih dahulu. Menurutnya, cluster adalah sekumpulan objek data yang serupa satu sama lain dalam satu grup yang sama dan tidak serupa dengan objek cluster yang berbeda. Objek-objek ini dikelompokkan menjadi satu atau lebih cluster sehingga objek-objek dalam satu cluster akan memiliki kemiripan yang tinggi satu sama lain.

2.3 Algoritma Partitioning Around Medoids

Algoritma PAM mengelompokkan sekumpulan objek n menggunakan partisi *clustering* agar menjadi beberapa jumlah k cluster. Algoritma ini memakai sebuah objek dari sekumpulan objek untuk menggantikan sebuah cluster. Objek yang menggantikan sebuah cluster diujarkan dengan medoids. Membangun cluster memakai perhitungan jarak korelasi yang disandang antar objek medoids dengan objek non-medoids.

Tahapan algoritma *Partitioning Around Medoids* (PAM) Menurut Han & Kamber (2006), beberapa tahapan yang dilakukan antara lain:

- 1) Secara acak memilih k objek pada sekumpulan n objek untuk digunakan sebagai medoids.

- 2) Menggunakan *Euclidean Distance* untuk meletakkan objek non-medoids dalam cluster yang terdekat dengan medoids.
- 3) Secara acak menentukan o_{random} : objek non-medoids.
- 4) Menghitung S (selisih), dari total cost pertukaran medoids o_j dengan $o_{(random)}$.
- 5) Bila $S < 0$ ganti o_j dengan $o_{(random)}$ untuk mengasilkan objek baru untuk digunakan sebagai medoid.
- 6) Ulang tahap 3 sampai 5 hingga medoids tidak mengalami perubahan, hingga didapatkan cluster serta anggota pada tiap-tiap cluster.
- 7) Mengulangi tahap 3 sampai 5 hingga tidak ada pergantian medoids, kemudian diperoleh cluster serta tiap-tiap anggota cluster. Pernyataan nilai dengan persamaan:

$$Total\ Cost = \sum_{k=1}^n \sqrt{\sum (x_k - y_k)^2}$$

Dengan:

n = jumlah sebuah data

k = indikator data

x_k = nilai atribut ke- k dari x

y_k = nilai atribut ke- k dari y

Nilai S dinyatakan dalam persamaan:

S = Total Cost baru - Total Cost Lama

Dengan:

S = Selisih

Total Cost Baru = Cost non-medoids

Total Cost Lama = Cost medoids

2.4 Davies Bouldin Indeks

Metode validasi yang digunakan pada penelitian ini memakai metode *Davies Bouldin Index* (DBI). *Davies Bouldin Index* adalah suatu cara yang dipakai untuk menganalisa kualitas *cluster* dari setiap cara *clustering*. pada penelitian tersebut DBI digunakan untuk melakukan analisa data dari setiap *cluster*.

Berikut tahapan dalam menghitung DBI (Sujacka, 2019):

- 1) *Sum of Square Within cluster* (SSW). adalah setaraf yang dipakai untuk mendapat keterikatan matriks pada sebuah *cluster* ke-*i* yang diuraikan sebagai berikut:

$$SSWi = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Dengan:

m_i = total data dalam *cluster* ke-*i*

c_i = *centroid cluster* ke-*i*

$d(x_j, c_i)$ = merupakan jarak setiap data terhadap *centroid*

- 2) Menghitung *Sum of Square Between cluster* (SSB). adalah persamaan yang dimanfaatkan untuk mengetahui pemisahan antar *cluster* yang dijumlah memakai persamaan sebagai berikut:

$$SSB_{i,j} = d(c_i, c_j)$$

- 3) Menghitung jumlah *ratio* (rasio). Sesudah nilai SSW dan SSB didapatkan, kemudian melakukan evaluasi rasio (R_{ij}) untuk memperoleh nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*. Nilai rasio dapat dihitung dengan persamaan sebagai

$$R_{ij} = \frac{SSWi + SSWj}{SSB_{ij}}$$

- 4) Menghitung nilai DBI. Nilai rasio yang didapat kemudian dipakai untuk mencari nilai DBI dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij})$$

Dengan:

k = jumlah *cluster* yang ditentukan.

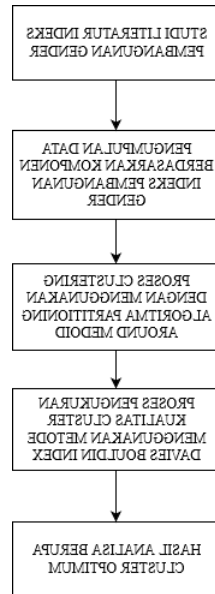
2.5 Rapidminer

Menurut Dennis, dkk (2013) *RapidMiner* adalah *software* yang bersifat *open source*. *RapidMiner* ialah jalan keluar untuk mengerjakan penjabaran pada data mining, *text mining* dan analisa prediksi. *RapidMiner* memakai beragam teknik pemaparan dan estimasi dalam memberikan pandangan kepada pengguna sehingga bisa menciptakan keputusan yang terbaik.

3. METODOLOGI

3.1 Tahapan Penelitian

Pada penelitian ini menggunakan algoritma PAM dengan teknik *performance* menggunakan DBI dalam mengelompokkan provinsi di Indonesia berdasarkan IPG, memiliki tahapan sebagai berikut.



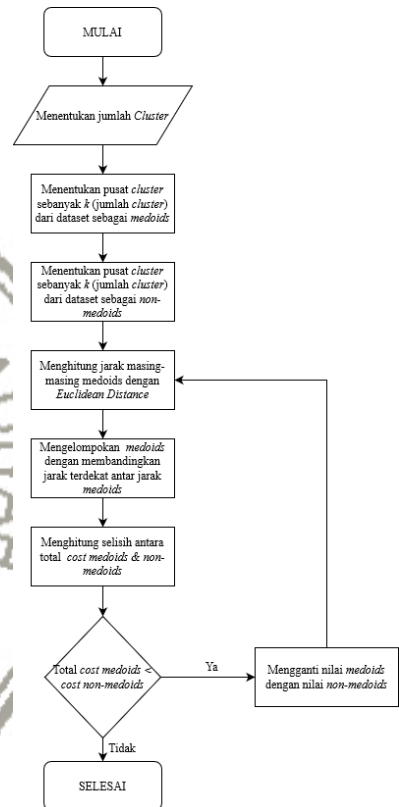
Gambar 1. Diagram Alur Penelitian

3.2 Studi Literatur

Tahapan pertama pada studi ini ialah dengan menentukan persoalan yang hendak dicermati, setelah itu menganalisis dengan menetapkan latar belakang, cakupan, dan menganalisis beberapa literatur yang terkait dengan persoalan yang ada kemudian mencari jalan keluar dari persoalan yang tertera. Untuk menggapai tujuan tersebut, maka penulis harus mempelajari beberapa literatur terkait lalu diseleksi sebagai literatur yang akan digunakan dalam penelitian.

3.3 Metode Analisis Data

Penelitian ini memanfaatkan algoritma PAM merupakan teknik pengelompokan atau *clustering* pada sekumpulan objek yang mewakili (medoid) dalam pusat *cluster* pada setiap *cluster*. Kelebihan dari PAM ialah untuk menanggulangi kekurangan dari algoritma K-Means yang sensitif ketika ada data noise atau pencilan (outliner) data.



Gambar 2. Flowchart Algoritma PAM

3.4 Pengumpulan Data

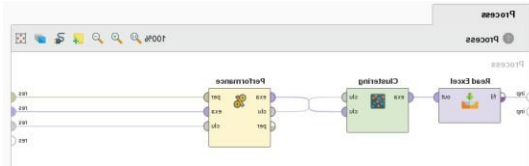
Pengumpulan data yang dibutuhkan berdasar pada studi literatur dari situs resmi Badan Pusat Statistik yaitu <https://www.bps.go.id>. Data yang digunakan pada penelitian ini merupakan data komponen pembentuk IPG menurut provinsi pada tahun 2010 sampai dengan 2019 yang terdiri dari 34 provinsi di Indonesia.

4. HASIL DAN PEMBAHASAN

4.1 Data Pengujian

Menggunakan data yang meupakan data provinsi di Indonesia berdasarkan komponen pembentuk IPG pada tahun 2010 sampai dengan 2019.

4.2 Rapidminer Studio



Gambar 3. Proses *Rapidminer*

Pada gambar 3 terdapat beberapa operator yang digunakan. Berikut fungsi dari operator-operator tersebut::

- Read Excel*: Digunakan untuk memuat data dari sheet pada Microsoft Excel.
- Clustering* : Melakukan pengelompokan menggunakan algoritma PAM. Pada penelitian ini menggunakan jumlah *cluster* sebanyak 2 sampai dengan 10 *cluster*.
- Performance*: Operator performance menggunakan *cluster distance performance* dimana operator ini digunakan untuk mengevaluasi kinerja metode k-medoids. Pada proses performances menampilkan hasil evaluasi kinerja berdasarkan nilai DBI.

4.3 Penentuan Jumlah *Cluster* Optimum

Setelah melakukan proses *cluster* menggunakan *Partitioning Around Medoids*, kemudian melakukan proses penentuan jumlah *cluster* optimum dengan menggunakan teknik *Davies Bouldin*.

Tabel 1. Nilai *Davies Bouldin Indeks*

<i>cluster</i>	Nilai DBI	<i>Cluster</i> yang dihasilkan	Anggota <i>Cluster</i>
2	1,749	2	1 : 31 dan 2 : 3
3	0,361	3	1 : 2, 2 : 1, dan 3 : 31
4	0,405	4	1 : 2, 2 : 1, 3 : 17, 4 : 14
5	1,196	5	1 : 14, 2 : 2, 3 : 1, 4 : 6, dan 5 : 11
6	0,963	6	1 : 7, 2 : 1, 3 : 2, 4 : 1, 5 : 5, dan 6 : 11
7	0,944	7	1 : 1, 2 : 4, 3 : 2, 4 : 1, 5 : 11, 6 : 5, dan 7 : 10
8	0,822	8	1 : 7, 2 : 1, 3 : 4, 4 : 2, 5 : 1, 6 : 5, 7 : 6, dan 8 : 8
9	0,465	9	1 : 1, 2 : 4, 3 : 1, 4 : 2, 5 : 11, 6 : 1, 7 : 3, 8 : 10, dan 9 : 1
10	0,886	10	1 : 1, 2 : 5, 3 : 2, 4 : 1, 5 : 3, 6 : 5, 7 : 5, 8 : 4, 9 : 3, dan 10 : 5

Berdasarkan Tabel 1 menunjukkan nilai *Davies Bouldin index* dengan pengujian skenario *cluster 2* sampai dengan 10 *cluster*, maka *cluster* yang paling optimum berdasarkan nilai *Davies Bouldin* indeks terkecil yaitu 3 *cluster* dengan nilai *Davies Bouldin* yaitu 0,361.

4.4 Hasil Profiling Cluster

Metode *clustering* dengan memanfaatkan algoritma PAM dan proses menemukan *cluster* optimum memakai metode DBI, didapat *cluster* optimum ada pada *cluster 3*. Perlu memahami keunikan pada tiap-tiap *cluster*, maka dijalankan *profiling cluster* pada total data IPG di tiap provinsi yang ada pada kelompok *cluster 1*, *cluster 2*, dan *cluster 3*. Berlandaskan perhitungan sebelumnya. Hasil 3 *cluster* mencakup *cluster 1* terdapat 2 provinsi, *cluster 2* terdapat 1 provinsi, dan *cluster 3* terdapat 31 provinsi.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

- 1) Pemanfaatan algoritma PAM untuk proses *clustering* data Indeks Pembangunan Gender di 34 provinsi berdasarkan 4 komponen pembentuk dan hasil dari validasi *cluster* menggunakan metode DBI didapat nilai *cluster* yang paling optimum adalah *cluster 3* yang memiliki nilai terendah pada DBI yaitu 0,361.
- 2) *Cluster 1* beranggotakan 2 provinsi yaitu Papua Barat dan Papua. *Cluster 2* beranggotakan 1 provinsi yaitu provinsi Kalimantan Utara. *Cluster 3* beranggotakan 31 Provinsi yaitu Aceh,

Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, dan provinsi Maluku Utara.

5.2 Saran

- 1) Dalam proses validasi *cluster* untuk mencari *cluster* optimum dapat menggunakan alternatif lain seperti metode *Silhouette Coefficient*, *elbow* dan *Calinski-Harabasz index*.
- 2) Perhitungan manual *cluster* pada algoritma *Partitioning Around Medoids* (PAM) pada *Microsoft Excel* dapat digunakan sebagai alternatif pembanding hasil *cluster* pada tools *RapidMiner*.

6. DAFTAR PUSTAKA

- [1] Dennis, A., Baskoro, D.A., Ambarwati L., Wicaksana I.W.S. 2012. Belajar Data Mining dengan RapidMiner. Jakarta: Gramdia Pustaka Utama.
- [2] Han, J., & Kamber, M. 2006. Data Mining : Concept and Techniques, 2nd ed. San Fransisco: Morgan Kauffman.
- [3] Kaur, N. U., & Singh, D. 2014. K-Medoids *Clustering* Algorithm – A Review. [pdf] International Journal of Computer Application and Technology (IJCAT). ISSN. 2349-1841 Vol. 1, Issue 1. April 2014.
- [4] Kemenkes Kesehatan RI, 2010. Penyebab perkembangan virus HIV. Jakarta.

- [5] Larose, D.T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.
- [6] Larose, D.T. 2006. DATA MINING METHODS AND MODELS, New Jersey: John Wiley & Sons, Inc.
- [7] RuiXu and Donald, C.W. 2009. *Clustering*, A John Wiley & Sons, Inc., Publication
- [8] Retno, S. 2019. Peningkatan Akurasi Algoritma K-Means dengan *Clustering Purity* sebagai Titik Pusat *Cluster* Awal (Centroid).[pdf] repository.usu.ac.id. [diakses 9 Desember 2019].
- [9] Turban, E., Ramesh, S., & Dursun, D. 2005. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset.

