

Penerapan Data Mining Untuk Clustering Indeks Pembangunan Manusia Berdasarkan Provinsi Di Indonesia

Hermawan^{1*}, Humisar Hasugian²

^{1,2}Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: ^{1*}1812510442@student.budiluhur.ac.id, ²humisar.hasugian@budiluhur.ac.id
(* : corresponding author)

Abstrak- Pembangunan manusia digunakan sebagai parameter untuk melihat pembangunan dari sisi manusia. Indeks pembangunan manusia atau IPM menjelaskan bagaimana masyarakat mendapatkan penghasilan yang cukup, kesehatan, dan pendidikan yang memadai. Secara geografis Indonesia adalah negara kepulauan dimana setiap provinsi tersebar di berbagai pulau yang dipisahkan oleh laut. Menjadikan disparitas pembangunan manusia antar provinsi menjadi relatif tinggi. Kesenjangan yang terjadi masih menjadi persoalan yang harus segera diselesaikan, karena kesenjangan pada indeks pembangunan manusia yang dapat menghambat tujuan pemerintah untuk pemerataan kesejahteraan manusia di Indonesia. Diperlukan sebuah langkah pengelompokan tiap provinsi sesuai dengan nilai indeks pembangunan manusia sehingga dapat di ambil solusi sesuai dengan kelompoknya. Ilmu yang mempelajari proses penggalian informasi dari sekumpulan data sehingga menghasilkan informasi disebut dengan *data mining*. Teknik yang dapat digunakan dalam menyelesaikan permasalahan pengelompokan data yaitu salah satunya adalah *clustering* dengan algoritma *k-means*. Penelitian ini menghasilkan 6 *cluster* dimana *cluster* dengan nilai IPM baik sekali terdiri dari 2 provinsi, *cluster* dengan nilai IPM baik terdiri dari 3 provinsi, *cluster* dengan nilai IPM lebih dari cukup 17 provinsi, *cluster* dengan nilai IPM cukup 8 provinsi, *cluster* dengan nilai IPM buruk 3 provinsi dan *cluster* dengan nilai IPM sangat buruk 1 provinsi. Dengan pengujian hasil evaluasi kinerja jumlah *cluster* menggunakan *Davies Bouldin Index* (DBI) mendapatkan hasil -0,283.

Kata Kunci: indeks pembangunan manusia, *data mining*, *k-means*

Application of Data Mining for Clustering Human Development Index By Province in Indonesia

Abstract- Human development is used as a parameter to see development from the human side. The human development index or HDI explains how people get sufficient income, adequate health and education. Geographically, Indonesia is an archipelagic country where each province is spread over various islands separated by the sea. Making the disparity in human development between provinces to be relatively high. The gap that occurs is still a problem that must be resolved immediately, because the human development index can hinder the government's goal of equitable distribution of human welfare in Indonesia. A step is needed to group each province according to the value of the human development index so that solutions can be taken according to the group. The science that studies the process of finding information from data so as to produce information is called data mining. The technique that can be used in solving the problem of grouping data is one of them is grouping with the *k-means* algorithm. This study resulted in 6 clusters where the cluster with a very good HDI score consisted of 2 provinces, the cluster with a good HDI score consisted of 3 provinces, the cluster with a HDI score of more than 17 provinces, the cluster with an adequate HDI score of 8 provinces, the cluster with a poor HDI score. 3 provinces and clusters with very poor HDI scores 1 province. By testing the performance evaluation of the number of clusters using the *Davies Bouldin Index* (DBI) the result is -0.283.

Keywords: human development index, *data mining*, *k-means*

1. PENDAHULUAN

Indeks Pembangunan Manusia atau IPM adalah indeks yang digunakan untuk mengukur kinerja pembangunan wilayah, pengukuran tersebut di lihat dari hal harapan hidup dan kesehatan, pengetahuan dan pendidikan, dan pendapatan yang disesuaikan. [1] Secara geografis Indonesia adalah negara kepulauan dimana setiap provinsi tersebar di berbagai pulau yang dipisahkan oleh laut. Menjadikan disparitas pembangunan manusia antar provinsi menjadi relatif tinggi. Hal ini masih menjadi pekerjaan yang harus segera diselesaikan, karena kesenjangan pada indeks pembangunan manusia yang dapat menghambat tujuan pemerintah untuk pemerataan kesejahteraan manusia di Indonesia. Diperlukan sebuah langkah pengelompokan tiap provinsi sesuai dengan nilai indeks pembangunan manusia sehingga dapat di ambil solusi sesuai dengan kelompoknya.. *Data mining* merupakan suatu ilmu atau metode yang digunakan untuk mencari informasi baru dari sekumpulan data yang banyak[2] [3].

Pada penelitian serupa [4] membahas tentang penerapan teknik *data mining clustering* dengan dengan *k-means* dapat digunakan untuk menyelesaikan permasalahan kepadatan penduduk di kota Medan. Pada penelitian [5] penerapan *data mining* dengan menggunakan *k-means* digunakan untuk mengelompokan potensi desa berdasarkan

jumlah industri menurut provinsi di Indonesia. Pada penelitian [6] penerapan *data mining* dengan menggunakan *k-means* digunakan untuk *clustering* kecamatan di kota Bandung berdasarkan indikator penduduk.

Berdasarkan penelitian yang sudah ada, sehingga dipilih metode *data mining* untuk menyelesaikan persoalan dalam mengelompokkan data indeks pembangunan manusia berdasarkan provinsi di Indonesia dengan teknik *clustering* menggunakan algoritma *k-means*. Diharapkan penelitian ini dapat memberikan informasi dari data yang ada sebagai penunjang dalam pengambilan keputusan oleh pemerintah untuk dapat meningkatkan Provinsi dengan Indeks Pembangunan Manusia yang tergolong rendah. Sehingga ketimpangan pada indeks pembangunan manusia pada provinsi di Indonesia semakin berkurang dan pemerataan kesejahteraan manusia dapat diwujudkan.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Metode yang digunakan dalam penelitian ini adalah metode eksperimen. Tahap eksperimen yang digunakan adalah *Cross Industry Standard Process Model for Data Mining* (CRISP-DM).

2.2 Teknik Pengumpulan Data

Tahapan ini ditentukan metode pengumpulan data, pemahaman data, dan persiapan data. Pada penelitian ini menggunakan data sekunder. Data sekunder adalah data yang diterbitkan oleh instansi/perusahaan dan sudah di publikasi ke pihak lain, dimana data dapat digunakan oleh siapapun yang ingin menggunakannya.[7]. Data yang diperoleh yaitu data Nilai Indeks Pembangunan Manusia di Indonesia yang berjumlah 34 data. Data yang digunakan yaitu data nilai indeks pembangunan manusia 6 tahun terakhir yaitu 2016 sampai dengan 2021.

2.3 K-Means

Pemodelan yang digunakan penulis dalam penelitian ini yaitu *clustering* dengan menggunakan algoritma *k-means*. *K-Means* merupakan salah satu metode yang dipopulerkan oleh James B MacQueen. Metode *k-means* merupakan metode yang bertujuan untuk mengelompokkan objek ke dalam *k cluster* ($k < n$) dimana nilai *k* telah ditentukan sebelumnya.[8]

Proses *clustering* dimulai dengan mengidentifikasi data yang akan dilakukan *clustering* X_{ij} ($i = 1, \dots, n$), $j = 1, \dots, m$) dengan n adalah jumlah data yang akan dilakukan *clustering*, dan m adalah atribut (variable). Pada awal iterasi, pusat pada setiap *cluster* ditentukan secara sembarang. c_{kj} ($k = 1, \dots, k$); $j = 1, \dots, m$.

Kemudian dilakukan perhitungan jarak antar setiap data dengan setiap *cluster* disebut centroid. Perhitungan jarak dilakukan dengan Euclidian distance yaitu:[9]

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

dengan :

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

m = banyaknya atribut (dimensi data)

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

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

Data tersebut yang menjadi anggota dari *cluster* ke- k jika jarak antar data ke *centroid* ke- k bernilai paling kecil jika dibandingkan dengan jarak ke *centroid* lainnya yang dapat di hitung dengan persamaan:

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

Sedangkan nilai centroid yang baru dapat dihitung dengan mencari rata-rata dari data yang menjadi anggota pada setiap *cluster* dengan persamaan:

$$c_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

dengan :

$x_{ij} \in$ *cluster* ke- k

p = banyaknya anggota *cluster* ke- k

Pengelompokan dengan *k-means* dapat berfungsi dengan baik ketika dilakukan pada kumpulan data kecil. Pada kumpulan data besar harus dikelompokkan sehingga setiap entitas atau titik data lain dalam *cluster* sama dengan entitas lainnya di *cluster* yang sama[10].

Kelebihan pada metode *k-means* yaitu:

- 1) Prosesnya yang mudah di implementasikan.
- 2) Algoritma dapat dijalankan dengan waktu singkat.
- 3) Mampu memecah kumpulan pola menjadi kelompok-kelompok yang memiliki kemiripan[11].

2.4 Evaluasi

Pada tahapan ini dilakukan evaluasi dengan menggunakan Davies Bouldin Index (DBI) untuk mengetahui jumlah *cluster* yang paling optimal (dengan nilai DBI yang paling kecil). Dimana semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering*[12]. *Davies Bouldin Index* (DBI) adalah metric untuk mengevaluasi hasil algoritma *clustering*. Evaluasi menggunakan Davies Bouldin Index ini memiliki skema evaluasi internal *cluster*, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster*[13].

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian adalah data nilai indeks pembangunan manusia yang diperoleh secara online melalui *website* Badan Pusat Statistik Indonesia sebanyak 34 data. Data yang digunakan yaitu data 6 nilai indeks pembangunan manusia 6 tahun terakhir yaitu 2016 sampai dengan 2021 yang dapat dilihat pada Tabel 1.

Tabel 1 Data Nilai Indeks Pembangunan Manusia 2016-2021

Provinsi	[Metode Baru] Indeks Pembangunan Manusia menurut Provinsi					
	2021	2020	2019	2018	2017	2016
Aceh	72.18	71.99	71.90	71.19	70.60	70
Sumatera Utara	72	71.77	71.74	71.18	70.57	70
Sumatera Barat	72.65	72.38	72.39	71.73	71.24	70.73
Riau	72.94	72.71	73	72.44	71.79	71.20
Jambi	71.63	71.29	71.26	70.65	69.99	69.62
Sumatera Selatan	70.24	70.01	70.02	69.39	68.86	68.24
Bengkulu	71.64	71.40	71.21	70.64	69.95	69.33
Lampung	69.90	69.69	69.57	69.02	68.25	67.65
Kep. Bangka Belitung	71.69	71.47	71.30	70.67	69.99	69.55
Kep. Riau	75.79	75.59	75.48	74.84	74.45	73.99
DKI Jakarta	81.11	80.77	80.76	80.47	80.06	79.60
Jawa Barat	72.45	72.09	72.03	71.30	70.69	70.05
Jawa Tengah	72.16	71.87	71.73	71.12	70.52	69.98
DI Yogyakarta	80.22	79.97	79.99	79.53	78.89	78.38
Jawa Timur	72.14	71.71	71.50	70.77	70.27	69.74
Banten	72.72	72.45	72.44	71.95	71.42	70.96
Bali	75.69	75.50	75.38	74.77	74.30	73.65
Nusa Tenggara Barat	68.65	68.25	68.14	67.30	66.58	65.81
Nusa Tenggara Timur	65.28	65.19	65.23	64.39	63.73	63.13
Kalimantan Barat	67.90	67.66	67.65	66.98	66.26	65.88
Kalimantan Tengah	71.25	71.05	70.91	70.42	69.79	69.13
Kalimantan Selatan	71.28	70.91	70.72	70.17	69.65	69.05
Kalimantan Timur	76.88	76.24	76.61	75.83	75.12	74.59
Kalimantan Utara	71.19	70.63	71.15	70.56	69.84	69.20
Sulawesi Utara	73.30	72.93	72.99	72.20	71.66	71.05
Sulawesi Tengah	69.79	69.55	69.50	68.88	68.11	67.47
Sulawesi Selatan	72.24	71.93	71.66	70.90	70.34	69.76
Sulawesi Tenggara	71.66	71.45	71.20	70.61	69.86	69.31
Gorontalo	69	68.68	68.49	67.71	67.01	66.29
Sulawesi Barat	66.36	66.11	65.73	65.10	64.30	63.60
Maluku	69.71	69.49	69.45	68.87	68.19	67.60
Maluku Utara	68.76	68.49	68.70	67.76	67.20	66.63
Papua Barat	65.26	65.09	64.70	63.74	62.99	62.21

Provinsi	[Metode Baru] Indeks Pembangunan Manusia menurut Provinsi					
	2021	2020	2019	2018	2017	2016
Papua	60.62	60.44	60.84	60.06	59.09	58.05
Indonesia	72.29	71.94	71.92	71.39	70.81	70.18

3.1 Pra Pemrosesan Data

Pra pemrosesan data dilakukan untuk memudahkan proses pencarian informasi dari hasil pengumpulan data. Pra pemrosesan data yang dilakukan dalam penelitian anatara lain:

a. Pembersihan Data

Pada tahap ini dilakukan penghapusan pada data yang tidak dibutuhkan. Yaitu dengan menghapus data Indonesia pada dataset, karena data Indonesia merupakan data rata rata nilai Indeks Pembangunan Manusia pada tiap provinsi.

b. Transformasi Data

Pada tahap ini agar memudahkan pada tahap perhitungan dengan algoritma *k-means* dataset yang sebelumnya sebanyak 6 atribut. Dilakukan perhitungan rata-rata nilai IPM dari tahun 2016-2021. Sehingga atribut menjadi satu dengan nilai rata-rata. Hasil pra pemrosesan data dapat dilihat pada tabel 2.

Tabel Dataset setelah pra pemrosesan data

Provinsi	AVG
Aceh	71,31
Bali	74,88
Banten	71,99
Bengkulu	70,70
DI Yogyakarta	79,50
DKI Jakarta	80,46
Gorontalo	67,86
Jambi	70,74
Jawa Barat	71,44
Jawa Tengah	71,23
Jawa Timur	71,02
Kalimantan Barat	67,06
Kalimantan Selatan	70,30
Kalimantan Tengah	70,43
Kalimantan Timur	75,88
Kalimantan Utara	70,43
Kep. Bangka Belitung	70,78
Kep. Riau	75,02
Lampung	69,01
Maluku	68,89
Maluku Utara	67,92
Nusa Tenggara Barat	67,46
Nusa Tenggara Timur	64,49
Papua	59,85
Papua Barat	64,00
Riau	72,35
Sulawesi Barat	65,20
Sulawesi Selatan	71,14
Sulawesi Tengah	68,88
Sulawesi Tenggara	70,68
Sulawesi Utara	72,36
Sumatera Barat	71,85
Sumatera Selatan	69,46
Sumatera Utara	71,21

3.2 Pemodelan

Setelah data set pada table 2 siap maka dilakukan pemodelan *clustering* dengan menggunakan algoritma *k-means*. Berdasarkan langkah-langkah pengolahan *data mining* dengan algoritma *k-means*, akan dilakukan tahapan sebagai berikut:

- Tahapan pertama yang perlu dilakukan yaitu menentukan jumlah *cluster*. Dimana ditentukan 6 *cluster*, yang terdiri dari *cluster* pertama (C0) dengan nilai baik sekali, *cluster* kedua (C1) dengan nilai baik, *cluster* ketiga (C2) dengan lebih dari cukup, *cluster* keempat (C3) dengan nilai cukup, *cluster* kelima (C4) dengan nilai buruk, dan *cluster* keenam (C5) dengan nilai sangat buruk.
- Memilih centroid awal, pada tahapan ini akan dipilih secara acak dari dataset pada tabel 2. Pusat *cluster* awal dapat dilihat pada tabel 3.

Tabel 3 *Cluster* pusat awal

No	Cluster	Provinsi	IPM
1	C0	DKI Jakarta	80,46
2	C1	Bali	74,88
3	C2	Kalimantan Selatan	70,30
4	C3	Lampung	69,01
5	C4	Sulawesi Barat	65,20
6	C5	Papua	59,85

- Pada tahap ini dilakukan perhitungan dengan rumus *eucclidean distance* untuk mencari jarak data pada nilai centroid pada IPM Papua sebagai berikut:

$$C0 = \sqrt{(59,85 - 80,46)^2} = 424,77$$

$$C1 = \sqrt{(59,85 - 74,88)^2} = 15,03$$

$$C2 = \sqrt{(59,85 - 70,30)^2} = 10,45$$

$$C3 = \sqrt{(59,85 - 69,01)^2} = 9,16$$

$$C4 = \sqrt{(59,85 - 65,20)^2} = 5,35$$

$$C5 = \sqrt{(59,85 - 59,85)^2} = 0,00$$

Pehitungan jarak nilai centroid dilakukan terus-menerus sampai data yang terakhir. Dimana pusat *cluster* ditentukan dari setiap *cluster* yang memiliki jarak terdekat dari pusat *clusternya*. Perhitungan terus dilakukan sampai nilai pusat *cluster* tidak berubah lagi.

- Proses perhitungan dihentikan pada iterasi ketiga diperoleh nilai *cluster* akhir pada tabel 4.

Tabel 4 Hasil dari perhitungan iterasi ke-3

AVG	C0	C1	C2	C3	C4	C5	JARAK	CLUSTER
71,31	75,15	3,95	0,14	2,81	6,12	11,46	0,14	C2
74,88	25,98	0,38	3,71	6,38	9,7	15,03	0,38	C1
71,99	63,83	3,27	0,82	3,49	6,8	12,14	0,82	C2
70,7	86,2	4,57	0,48	2,2	5,51	10,85	0,48	C2
79,5	0,23	4,24	8,32	11	14,31	19,65	0,23	C0
80,46	0,23	5,2	9,29	11,96	15,28	20,61	0,23	C0
67,86	146,79	7,4	3,31	0,63	2,68	8,01	0,63	C3
70,74	85,36	4,52	0,43	2,24	5,55	10,89	0,43	C2
71,44	73	3,83	0,26	2,94	6,25	11,59	0,26	C2
71,23	76,55	4,03	0,06	2,73	6,04	11,38	0,06	C2
71,02	80,24	4,24	0,15	2,52	5,84	11,17	0,15	C2
67,06	167,03	8,21	4,12	1,44	1,87	7,2	1,44	C3
70,3	93,75	4,96	0,88	1,8	5,11	10,45	0,88	C2
70,43	91,28	4,84	0,75	1,93	5,24	10,58	0,75	C2
75,88	16,82	0,62	4,71	7,38	10,69	16,03	0,62	C1
70,43	91,22	4,83	0,74	1,93	5,24	10,58	0,74	C2
70,78	84,66	4,48	0,39	2,28	5,59	10,93	0,39	C2
75,02	24,56	0,24	3,85	6,53	9,84	15,17	0,24	C1
69,01	120,25	6,25	2,16	0,52	3,83	9,16	0,52	C3
68,89	123,08	6,38	2,29	0,39	3,7	9,03	0,39	C3
67,92	145,34	7,34	3,25	0,57	2,74	8,07	0,57	C3
67,46	156,85	7,81	3,72	1,04	2,27	7,61	1,04	C3
64,49	239,86	10,77	6,68	4,01	0,69	4,64	0,69	C4
59,85	405,18	15,41	11,32	8,65	5,34	0	0	C5
64	255,39	11,26	7,17	4,5	1,19	4,15	1,19	C4

AVG	C0	C1	C2	C3	C4	C5	JARAK	CLUSTER
72,35	58,26	2,91	1,17	3,85	7,16	12,5	1,17	C2
65,2	218,42	10,06	5,97	3,3	0,01	5,35	0,01	C4
71,14	78,16	4,12	0,03	2,64	5,95	11,29	0,03	C2
68,88	123,12	6,38	2,29	0,39	3,7	9,03	0,39	C3
70,68	86,44	4,58	0,49	2,18	5,5	10,83	0,49	C2
72,36	58,13	2,91	1,18	3,86	7,17	12,51	1,18	C2
71,85	66,03	3,41	0,68	3,36	6,67	12	0,68	C2
69,46	110,65	5,8	1,71	0,96	4,27	9,61	0,96	C3
71,21	76,9	4,05	0,04	2,71	6,02	11,36	0,04	C2

e. Titik pusat *cluster* akhir perhitungan dapat di lihat pada tabel 5.

Tabel 5 Titik pusat *cluster* akhir

Cluster Pusat	Nilai
C0	79,98
C1	75,26
C2	71,17
C3	68,50
C4	65,19
C5	59,85

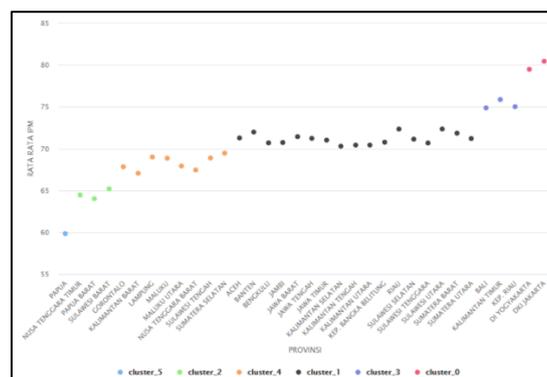
3.2.1 Hasil Pemodelan

Di akhir tahapan dapat disimpulkan bahwa *cluster* terhadap nilai Indeks Pembangunan Manusia di Indonesia berdasarkan provinsi sebagai berikut:

- Cluster* 0 dengan nilai Indeks Pembangunan Manusia baik sekali adalah Provinsi DKI Jakarta dan DI Yogyakarta
- Cluster* 1 dengan nilai Indeks Pembangunan Manusia baik adalah Provinsi Bali, Kep. Riau, dan Kalimantan Timur
- Cluster* 2 dengan nilai Indeks Pembangunan lebih dari cukup adalah Provinsi Aceh, Bangka Belitung, Bengkulu, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Utara, Riau, Sulawesi Tenggara, Jambi, Kep, Sulawesi Selatan, Sumatera Utara, Sumatera Barat, Banten, dan Sulawesi Utara.
- Cluster* 3 dengan nilai Indeks Pembangunan cukup adalah Provinsi Gorontalo, Lampung, Kalimantan Barat, Nusa Tenggara Barat, Maluku Utara, Sulawesi Tengah, Maluku, dan Sumatera Selatan.
- Cluster* 4 dengan nilai Indeks Pembangunan Manusia buruk adalah Provinsi Papua Barat, Nusa Tenggara Timur, dan Sulawesi Barat.
- Cluster* 5 dengan nilai Indeks Pembangunan Manusia sangat buruk adalah Provinsi Papua.

3.3 Implementasi Pada RapidMiner

Pada tahapan ini akan dilakukan pengujian menggunakan *tools rapidminer*, *RapidMiner* adalah perangkat lunak yang dibuat menggunakan bahasa java dan bersifat terbuka yang digunakan sebagai sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. [14]. untuk pengujian dataset Indeks Pembangunan Manusia. Dengan jumlah *cluster* yang telah ditentukan adalah sejumlah 6 *cluster*. Dimana dataset yang diproses sebanyak 34 data. Plot hasil proses data dapat di lihat pada gambar 1.



Gambar 1 Plot hasil Implementasi pada *rapidminer*

Hasil implementasi ditampilkan pot *clustering* dimana *cluster* 0 masuk kategori dengan nilai IPM baik sekali, *cluster* 3 baik, *cluster* 1 lebih dari cukup, *cluster* 4 cukup, *cluster* 2 buruk, dan *cluster* 5 sangat buruk.

3.3.1 Uji Validasi

Dilakukan perhitungan evaluasi DBI pada software *rapiminer* dengan pemodelan jumlah *cluster* ($k=2$ sampai dengan $k=8$) dimana hasil nilai DBI dapat dilihat pada tabel 6.

Tabel 6 Hasil DBI

Cluster Nilai (k)	Nilai DBI
2	-0,733
3	-0,485
4	-0,428
5	-0,338
6	-0,283
7	-0,355
8	-0,284

Dapat dilihat pada Tabel 6 dimana *cluster* terbaik terdapat pada $k=6$ dengan nilai DBI -0,283. Dimana semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering*[12].

4. KESIMPULAN

Penerapan *data mining* menggunakan teknik *clustering* dengan algoritma *k-means* untuk pengelompokan provinsi di Indonesia berdasarkan nilai Indeks Pembangunan Manusia dapat dilakukan. Dengan menetapkan nilai rata-rata IPM dari tahun 2016-2021 sebagai *attribute*. Dimana dalam proses penelitian dengan perhitungan manual maupun menggunakan *rapidminer* didapatkan hasil penelitian yang sama yaitu sejumlah 6 *cluster* dengan *cluster* dengan nilai IPM baik sekali terdiri dari 2 provinsi, *cluster* dengan nilai IPM baik terdiri dari 3 provinsi, *cluster* dengan nilai IPM lebih dari cukup 17 provinsi, *cluster* dengan nilai IPM cukup 8 provinsi, *cluster* dengan nilai IPM buruk 3 provinsi dan *cluster* dengan nilai IPM sangat buruk 1 provinsi. Dengan pengujian hasil evaluasi kinerja jumlah *cluster* menggunakan *Davies-Bouldin Index* (DBI) didapatkan nilai terbaik pada jumlah *cluster* 6 yaitu -0,283.

DAFTAR PUSTAKA

- [1] Yoyo Karyono, et al., *Indeks Pembangunan Manusia 2020*, Direktorat Analisis dan Pengembangan Statistik, Jakarta, Badan Pusat Statistik, 2021, .
- [2] J. Hutagalung dan F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, hlm. 1187, Jul 2021, doi: 10.30865/mib.v5i3.3113.
- [3] F. Nurdiansyah, I. Akbar, R. Artikel, K. Kunci, dan B. Corresponding Author Firman Nurdiansyah, "Jurnal Teknologi dan Manajemen Informatika Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop Info Artikel ABSTRAK," vol. 7, no. 2, hlm. 86–94, 2021, [Daring]. Available: <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- [4] P. Kepadatan, P. Berdasarkan, J. Penduduk, K. Medan, P. Marpaung, dan R. F. Siahaan, "Penerapan Algoritma K-Means Clustering Untuk," 2021.
- [5] E. Muningsih, I. Maryani, dan V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, 2021, [Daring]. Available: www.bps.go.id
- [6] S. Wulandari, "Clustering Kecamatan Di Kota Bandung Berdasarkan Indikator Jumlah Penduduk Dengan Menggunakan Algoritma K-Means," 2020.
- [7] R. Mustofa dan A. Syahza, "Food Carrying Capacity as an Indicator of Sustainability of Smallholder Oil Palm Plantations in Riau Province Analysis of the Effect of Climate Variability Risk on Rice Farming Productivity Using Robust Regression," *Review of International Geographical Education (RIGEO)*, vol. 11, no. 8, hlm. 111–121, 2021, doi: 10.48047/rigeo.11.08.11.
- [8] K. R. Adzima, A. Bustamam, dan D. Aldila, "The implementation of k-means partitioning algorithm in HOPACH clustering method," dalam *IOP Conference Series: Earth and Environmental Science*, Apr 2019, vol. 243, no. 1. doi: 10.1088/1755-1315/243/1/012073.
- [9] A. Bustamam, H. Tasman, N. Yuniarti, Frisca, dan I. Mursidah, "Application of k-means clustering algorithm in grouping the DNA sequences of hepatitis B virus (HBV)," dalam *AIP Conference Proceedings*, Jul 2017, vol. 1862. doi: 10.1063/1.4991238.
- [10] V. H. Pham dan B. R. Lee, "An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm," *Vietnam Journal of Computer Science*, vol. 2, no. 1, hlm. 25–33, Feb 2015, doi: 10.1007/s40595-014-0028-3.

- [11] X. Wang dan Y. Bai, “The global Minmax k-means algorithm,” *Springerplus*, vol. 5, no. 1, Des 2016, doi: 10.1186/s40064-016-3329-4.
- [12] A. Bates dan J. Kalita, “Counting clusters in twitter posts,” dalam *ACM International Conference Proceeding Series*, Mar 2016, vol. 04-05-March-2016. doi: 10.1145/2905055.2905295.[13]N. K. Zuhail, “Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering,” 2022.
- [13] L. Elvitaria, “Memprediksi Tingkat Peminat Ekstrakurikuler Pada Siswa Smk Analisis Kesehatan Abdurrah Menggunakan Algoritma C4.5 (Studi Kasus: Smk Analis Kesehatan Abdurrah),” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 2, no. 2, hlm. 220–233, Agu 2017, doi: 10.36341/rabit.v2i2.212.