

Implementasi Algoritma *K-Means Clustering* Menggunakan Aplikasi *Orange* Untuk Mengetahui Pola Indeks Pembangunan Manusia Tahun 2022

Kaylista N.N.K¹, Nabiilah Khoirunnisaa¹, Gading Viewianti E.N.F¹, Ajif Yunizar Pratama Yusuf^{1,*}

¹ Informatika, Fakultas Ilmu Komputer; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan No.81 Margamulya, Kota Bekasi, Jawa Barat, Telp. (021) 7231948; e-mail: 202010225160@mhs.ubharajaya.ac.id, 202010225270@mhs.ubharajaya.ac.id, 202010225165@mhs.ubharajaya.ac.id, ajif.yunizar@dsn.ubharajaya.ac.id

* Korespondensi: e-mail: ajif.yunizar@dsn.ubharajaya.ac.id

Diterima : 20 Juli 2023; Review : 24 Juli 2023; Disetujui : 25 Juli 2023; Diterbitkan : 25 Juli 2023

Abstract

The Human Development Index (HDI) is used as a statistical parameter to evaluate the progress and quality of human life in a country. Human development plans are measured through three basic aspects: a long and healthy life, knowledge, and a decent standard of living. The research focuses on the application of k-means algorithms to identify patterns and group 34 provinces in Indonesia based on indicators that form the 2022 HDI. These indicators include life expectancy (UHH), school life expectancy (HLS), average school time (RLS), and spending. This HDI grouping aims to identify the HDI variables that should be a priority in development. The results of the analysis showed the formation of two groups (clusters) through K-Means Cluster Analysis. Cluster 1 has provincial characteristics with high to very high values on UHH, HLS, RLS, and customized output. Meanwhile, Cluster 2 consists of provinces with medium to high values on UHH, HLS, RLS, and adjusted output.

Keywords: *Clustering, Data Mining, Human Development Index, K-Means, Orange*

Abstrak

Indeks Pembangunan Manusia (IPM) digunakan sebagai parameter statistik untuk mengevaluasi kemajuan serta kualitas hidup manusia dalam suatu negara. Rancangan pembangunan manusia diukur melalui tiga aspek dasar, yakni umur panjang dan sehat, pengetahuan, dan standar hidup yang layak. Penelitian ini fokus pada penerapan algoritma k-means untuk mengidentifikasi pola dan mengelompokkan 34 provinsi di Indonesia berdasarkan indikator yang membentuk IPM tahun 2022. Indikator tersebut mencakup umur harapan hidup (UHH), harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), dan pengeluaran. Pengelompokan IPM ini bertujuan untuk mengetahui variabel IPM yang harus menjadi prioritas dalam pembangunan. Hasil analisis menunjukkan pembentukan 2 kelompok (klaster) melalui Analisis K-Means Cluster. Klaster 1 memiliki karakteristik provinsi dengan nilai tinggi hingga sangat tinggi pada UHH, HLS, RLS, dan pengeluaran yang telah disesuaikan. Sementara itu, Klaster 2 terdiri dari provinsi dengan nilai sedang hingga tinggi pada UHH, HLS, RLS, dan pengeluaran yang telah disesuaikan.

Kata kunci: Clustering, Data Mining, K-Means, Indeks Pembangunan Manusia, Orange

1. Pendahuluan

Dalam pembangunan nasional Indonesia, rakyat ditempatkan sebagai pusat utama dari proses pembangunan. Untuk berpartisipasi dalam proses tersebut, tentu diperlukan masyarakat yang memiliki jumlah dan mutu yang unggul. Indeks Pembangunan Manusia (IPM) sebagai tolok ukur keberhasilan dalam usaha meningkatkan kualitas hidup manusia. IPM merupakan parameter statistik yang dimanfaatkan sebagai penilaian kemajuan dan kualitas hidup manusia di suatu negara. IPM meliputi tiga aspek utama manusia, yaitu usia panjang dan kehidupan yang sehat, pengetahuan, serta standar hidup yang layak. Maksud pokok dari pembangunan manusia adalah meningkatkan kesejahteraan manusia secara merata, menyeluruh, dan berkelanjutan. Konsep ini telah diakui secara internasional dan digunakan oleh berbagai organisasi untuk mengukur kemajuan pembangunan di negara-negara yang berbeda (United Nations Development Programme., n.d.).

Indonesia laksana sebuah hamparan negara yang begitu luas, memiliki wilayah-wilayah yang terbentang di seluruh penjuru tanah air. Setiap wilayah tercatat memiliki keterkaitan dengan Indeks Pembangunan Manusia (IPM), di mana perhitungan dilakukan untuk mengetahui tingkat kesejahteraan dan kemajuan penduduk setiap wilayah. Data mengenai IPM di Indonesia dikumpulkan melalui Badan Pusat Statistik (BPS). Pada tahun 2022, BPS merilis hasil IPM yang menunjukkan peningkatan sebesar 0.86% dibandingkan dengan tahun 2021 (*INDEKS PEMBANGUNAN MANUSIA 2022*, n.d.).

Dalam penelitian ini, algoritma *K-Means Clustering* dimanfaatkan sebagai pengategorian provinsi-provinsi di Indonesia berlandaskan parameter yang membentuk Indeks Pembangunan Manusia tahun 2022. Indikator-indikator tersebut meliputi umur harapan hidup (UHH), harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), serta pengeluaran. Sebagai keseluruhan, hasil penelitian bersifat klusterisasi yang dapat digunakan selaku panduan untuk menggambarkan pengategorian provinsi-provinsi di Indonesia yang termasuk dalam kategori sedang hingga tinggi, bahkan sangat tinggi. Hal ini memungkinkan kita untuk mengetahui jumlah provinsi yang memiliki nilai IPM sedang, sehingga dapat segera diberikan penanganan yang sesuai dan diambil tindakan untuk perbaikan pelayanan seperti pendidikan, layanan kesehatan, serta peningkatan lapangan kerja di wilayah-wilayah yang membutuhkannya. Tujuannya adalah meningkatkan kualitas hidup penduduk.

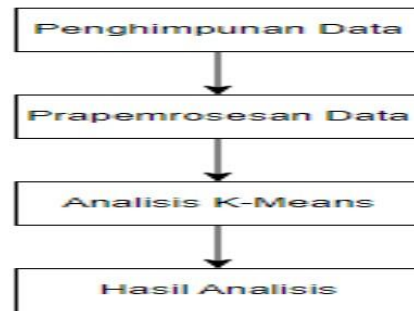
2. Metode Penelitian

2.1. Data Mining

Penambangan data merupakan konsep yang dimanfaatkan untuk menganalisis dan mengeksplorasi pemahaman yang terselubung di dalam basis data. Penambangan data ialah suatu teknologi yang menggabungkan pendekatan statistika, matematik, *artificial intelligence*, dan pembelajaran mesin untuk berkorelasi melalui kumpulan data yang besar dan menyingkap fakta yang memiliki nilai dan dapat dimanfaatkan (Widiastari, 2021). Data mining dapat berupa klasifikasi, clustering, sequence, forecasting, dan lain-lain.

Sebutan "*Data Mining*" menyanggah sejumlah interpretasi, seperti "*knowledge discovery*" dan "*pattern recognition*". Penggunaan istilah "*knowledge discovery*" atau penemuan pemahaman akurat digunakan karena sasaran utama dari data mining adalah untuk mengungkap pemahaman yang potensial internal data yang ada. Sementara itu, istilah "*pattern recognition*" atau pengenalan pola cocok digunakan karena data mining bermaksud untuk mendapatkan pola yang tersembunyi dalam himpunan data tersebut (Nabila et al., 2021).

Rincian tahapan tersebut dapat ditemukan dalam Gambar 1.



Sumber: Hasil Penelitian (2023)

Gambar 1. Tahapan Penelitian

Pada gambar 1 menunjukkan beberapa tahapan yang dimulai dengan pengumpulan data, dilanjutkan dengan pengolahan awal, dan diikuti oleh proses klusterisasi yang kemudian akan dianalisis dan divisualisasikan.

1) Penghimpunan Data

Dalam fase berjalan, seorang pengkaji menggarap penghimpunan data yang akan menjadi bukti konkret dalam penelitiannya (Sibarani et al., 2022). Pada tahap ini data dikumpulkan, kemudian data tersebut dipelajari dengan maksud untuk memahami informasi tersebut, mengidentifikasi dan mengetahui kualitasnya, dan menemukan subset informasi menarik yang dapat dijadikan hipotesis untuk informasi yang tersembunyi.

2) Prapemrosesan Data

a. *Data Cleaning*

Dilangkah ini, dilangsungkan untuk menghilangkan redundansi atau duplikasi data dan untuk menghilangkan *missing value* serta untuk menghilangkan atribut-atribut yang tidak difungsikan dalam penelitian.

b. *Data Integration*

Pada tahap integrasi data, yang melibatkan konversi data nominal menjadi nilai numerik untuk pengolahan memanfaatkan algoritma *K-Means clustering*, proses ini tidak diperlukan karena semua atribut data yang ada telah berupa data numerik.

c. *Data Selection*

Dalam tahap ini, dilakukan pemilihan terhadap himpunan data yang sebelumnya diperoleh lalu mengubahnya menjadi format yang mungkin diproses dalam model yang telah didefinisikan lebih lanjut (Akhmad Budi, 2019).

d. *Data Transformation*

Data diubah menjadi format .csv agar dapat diproses memanfaatkan tools data mining untuk mempergunakan algoritma K-Means.

2.2. K-Means Clustering

K-Means sebagai bagian dalam algoritma operasi penambangan data yang termasuk pada metode clustering non-hierarki. Pada algoritma ini objek-objek yang ada akan dibagi ke dalam satu atau lebih kelompok (*cluster*) berlandaskan spesifikasinya. Dengan demikian, kelompok objek-objek yang memiliki kekhususan yang serupa berada atas sebuah kelompok *cluster* yang serupa, sedangkan objek-objek dengan kekhususan yang bertentangan akan digabungkan di dalam *cluster* yang berbeda pula (Syukron Nawawi et al., 2021).

K-Means adalah tata cara kajian klaster yang bertujuan untuk membagi N objek pengamatan menjadi K kelompok (*cluster*), di mana setiap objek akan menjadi bagian dari klaster dengan rata-rata tertentu. Metode ini berusaha untuk mendapatkan pusat (*centroid*) dari setiap klaster dalam data melalui iterasi pembaruan. Tahapan dari algoritma K-Means ialah seperti berikut ini (Utami et al., 2021):

- 1) Menentukan total *cluster* k
- 2) Inisialisasi k *center cluster (centroid)* melalui cara acak
- 3) Mengestimasi jarak antara data dan *centroid* dilakukan dalam penelitian ini. Metode yang difungsikan adalah *Euclidean Distance*, yang menjadi metode yang paling populer untuk mengetahui jarak terdekat antara data dan *centroid*. Rumus yang digunakan adalah seperti berikut ini:

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2}$$

Keterangan :

$D_{(i,j)}$ = Jarak data ke i ke pusat klaster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

- 4) Langkah berikutnya ialah menghitung ulang senter klaster menggunakan komponen klaster yang telah diperbarui. *Center* klaster dihitung sebagai rerata dari seluruh data atau entitas dalam himpunan klaster tersebut.
- 5) Langkah sekarang diulang kembali mulai dari langkah 3 hingga nilai pusat klaster tidak mengalami perubahan.

2.3. Orange Data Mining

K-Means diterapkan menggunakan alat penambangan data bernama *Orange Data Mining*. Ialah aplikasi yang dapat digunakan secara terbuka dalam pemanfaatan teknologi pembelajaran mesin. Orange adalah alat visualisasi data, platform pembelajaran mesin, dan alat peringkasan data. Fitur program *visual front-end* ini memungkinkan analisis eksplorasi data kuantitatif yang cepat serta visualisasi data interaktif (Nazuah et al., 2023). Implementasi

pemrograman visual dilakukan melalui antarmuka yang menghubungkan widget yang telah dinyatakan atau dipertimbangkan oleh pengguna untuk membuat alur kerja. *Software* ini menyandang karakteristik khusus yang berguna dalam data mining, seperti pemodelan prakiraan, sistem usulan, serta cakup dimanfaatkan dalam penelitian pada bidang genomik, biomedis, bioinformatika, dan pendidikan (Aghnia et al., 2021).

3. Hasil dan Pembahasan

3.1. Pengimpunan Data

Data sekunder yang diperoleh dari situs publikasi resmi lembaga pemerintahan, yakni Badan Pusat Statistik (BPS), dimanfaatkan dalam studi ini yang kuasa diperoleh melalui www.bps.go.id. Data yang dipergunakan ialah data Indeks Pembangunan Manusia (IPM) berdasarkan provinsi di Indonesia pada tahun 2022. Selanjutnya data akan diproses berdasarkan dari yang sudah didapat sebelumnya menggunakan algoritma k-means melalui penggunaan alat pengolahan data, yaitu *Orange Data Mining*, untuk menentukan jumlah kluster IPM yang sesuai.

3.2. Prapemrosesan Data

Pada tahun 2022, cakupan Indeks Pembangunan Manusia (IPM) berdasarkan keseluruhan provinsi yang ada terdiri dari enam atribut, yaitu provinsi, umur harapan hidup (UHH), harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), pengeluaran per kapita, dan IPM. Totalnya terdapat 34 data atau record yang merepresentasikan 34 provinsi yang ada di Indonesia.

Kemudian, dilakukan prapemrosesan data dengan menghapus atribut IPM karena atribut tersebut tidak digunakan dalam penelitian ini. Sehingga, data yang di pilih untuk penelitian ini (*selection data*) hanya terdiri dari 5 atribut dan 34 record. Gambar 2 merupakan data yang belum melalui prapemrosesan data.

Provinsi	UHH	HLS	RLS	Pengeluaran	IPM
Aceh	70.16	14.37	9.44	9.953	72.5
Sumatera Utara	69.61	12.31	9.71	10,846	72.71
Sumatera Barat	69.3	14.1	9.16	11,130	73.26
Riau	71.95	13.29	9.22	11,158	73.52
Jambi	71.5	13.05	8.68	10,871	72.14
Sumatera Selatan	70.32	12.55	8.37	11,109	70.9
Bengkulu	69.69	13.69	8.91	10,840	72.16
Lampung	70.99	12.74	8.18	10,336	70.45
Kepulauan Bangka Belitung	70.98	12.18	8.11	13,358	72.24
Kepulauan Riau	70.5	12.99	10.37	14,469	76.46
DKI Jakarta	73.32	13.08	11.31	18,927	81.65
Jawa Barat	73.52	12.62	8.78	11,277	73.12
Jawa Tengah	74.57	12.81	7.93	11,377	72.79
DI Yogyakarta	75.08	15.65	9.75	14,482	80.64
Jawa Timur	71.74	13.37	8.03	11,992	72.75
Banten	70.39	13.05	9.13	12,216	73.32
Bali	72.6	13.48	9.39	13,942	76.44
Nusa Tenggara Barat	67.07	13.95	7.61	10,691	69.46
Nusa Tenggara Timur	67.47	13.21	7.7	7,877	65.9
Kalimantan Barat	71.02	12.66	7.59	9,355	68.63
Kalimantan Tengah	70.04	12.75	8.65	11,458	71.63
Kalimantan Selatan	69.13	12.82	8.46	12,469	71.84
Kalimantan Timur	74.62	13.84	9.92	12,641	77.44
Kalimantan Utara	72.67	13.06	9.27	9,350	71.83
Sulawesi Utara	72.08	12.95	9.68	11,179	73.81
Sulawesi Tengah	68.93	13.32	8.89	9,696	70.28
Sulawesi Selatan	70.97	13.53	8.63	11,430	72.82
Sulawesi Tenggara	71.37	13.69	9.25	9,708	72.23
Gorontalo	68.51	13.12	8.02	10,697	69.81
Sulawesi Barat	65.63	12.87	8.08	9,358	66.92
Maluku	66.45	14	10.19	8,876	70.22
Maluku Utara	68.79	13.73	9.24	8,398	69.47
Papua Barat	66.46	13.21	7.84	8,101	65.89
Papua	66.23	11.14	7.02	7,146	61.39

Sumber: Hasil Penelitian (2023)

Gambar 2. Dataset IPM sebelum prapemrosesan data

Pada gambar 2 menunjukkan Dataset IPM sebelum prapemrosesan data.

Provinsi	UHH	HLS	RLS	Pengeluaran
Aceh	70.18	14.37	9.44	9,963
Sumatera Utara	69.61	13.31	9.71	10,848
Sumatera Barat	69.9	14.1	9.18	11,130
Riau	71.95	13.29	9.22	11,158
Jambi	71.5	13.05	8.68	10,871
Sumatera Selatan	70.32	12.55	8.37	11,109
Bengkulu	69.69	13.68	8.91	10,840
Lampung	70.99	12.74	8.18	10,336
Kepulauan Bangka Belitung	70.98	12.18	8.11	13,358
Kepulauan Riau	70.5	12.99	10.37	14,469
DKI Jakarta	73.32	13.08	11.31	18,927
Jawa Barat	73.52	12.62	8.78	11,277
Jawa Tengah	74.57	12.81	7.93	11,377
DI Yogyakarta	75.08	15.65	9.75	14,482
Jawa Timur	71.74	13.37	8.03	11,992
Banten	70.39	13.05	9.13	12,216
Bali	72.6	13.48	9.39	13,942
Nusa Tenggara Barat	67.07	13.96	7.61	10,681
Nusa Tenggara Timur	67.47	13.21	7.7	7,877
Kalimantan Barat	71.02	12.66	7.59	9,355
Kalimantan Tengah	70.04	12.75	8.65	11,458
Kalimantan Selatan	69.13	12.82	8.46	12,469
Kalimantan Timur	74.62	13.84	9.92	12,641
Kalimantan Utara	72.67	13.06	9.27	9,350
Sulawesi Utara	72.08	12.95	9.68	11,179
Sulawesi Tengah	68.93	13.32	8.89	9,696
Sulawesi Selatan	70.97	13.53	8.63	11,430
Sulawesi Tenggara	71.37	13.69	9.25	9,708
Gorontalo	68.51	13.12	8.02	10,687
Sulawesi Barat	65.63	12.87	8.08	9,358
Maluku	66.45	14	10.19	8,876
Maluku Utara	68.79	13.73	9.24	8,398
Papua Barat	66.46	13.21	7.84	8,101
Papua	66.23	11.14	7.02	7,146

Sumber: Hasil Penelitian (2023)

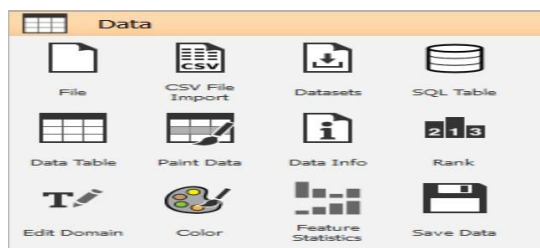
Gambar 3. Dataset IPM sesudah prapemrosesan data

Pada gambar 3 menunjukkan data yang sudah melalui prapemrosesan data.

3.3. Analisis K-Means

Setelah dilakukan prapemrosesan data, selanjutnya akan dilakukan analisis *K-Means Clustering* menggunakan aplikasi *Orange Data Mining* untuk mengelompokkan data berdasarkan indikator penyusun Indeks Pembangunan Manusia untuk menentukan mana yang tergolong sedang hingga ke tinggi serta tinggi hingga ke sangat tinggi. Langkah-langkahnya adalah sebagai berikut.

- 1) Pilih *Widget File* pada tabulasi Data di Orange.

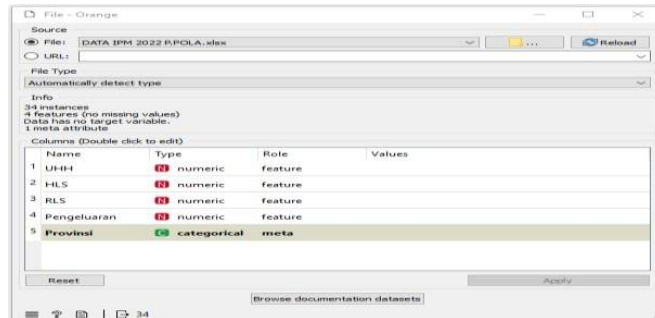


Sumber: Hasil Penelitian (2023)

Gambar 4. Tampilan *Widget File* pada Orange

Pada gambar 4 menunjukkan Tampilan *Widget File* pada Orange.

- 2) Klik pada *Widget File* untuk memilih file data yang akan diproses.



Sumber: Hasil Penelitian (2023)

Gambar 5. Penginputan Data pada *Orange*

Pada gambar 5 menunjukkan Penginputan Data pada *Orange*.

- 3) Untuk proses *clustering*, lalu hubungkanlah *Widget File* dengan *Widget K-Means*.

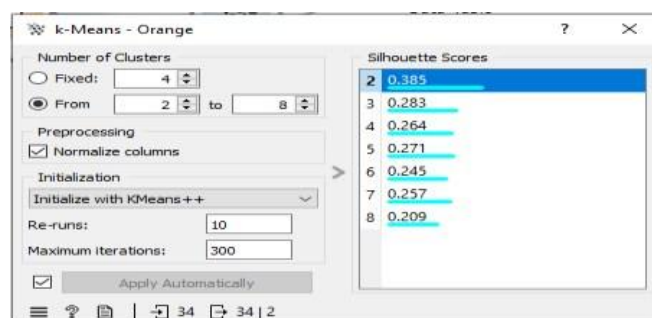


Sumber: Hasil Penelitian (2023)

Gambar 6. Tampilan *Widget K-Means* pada *Orange*

Pada gambar 6 menunjukkan Tampilan *Widget K-Means* pada *Orange*

- 4) Lalu tentukan jumlah *cluster* yang sudah dibuat otomatis oleh aplikasi. Setelah fitur *k-means* diatur maka diperoleh jumlah *cluster* yang ideal yaitu 2 *cluster* dengan *silhouette score* tertinggi yaitu 0,385.

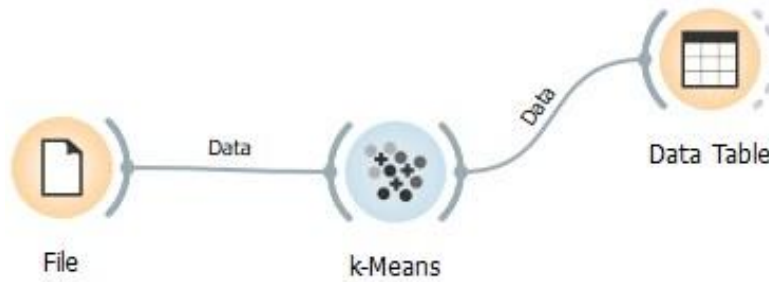


Sumber: Hasil Penelitian (2023)

Gambar 7. Proses Menentukan *Cluster*

Pada gambar 7 menunjukkan Proses Menentukan *Cluster*.

- 5) Lalu hubungkan *Widget K-Means* dengan *Widget Data Table* untuk melihat outputnya.



Sumber: Hasil Penelitian (2023)

Gambar 8. Tampilan *Widget Data Table* Pada *Orange*

Pada gambar 8 menunjukkan Tampilan *Widget Data Table* Pada *Orange*.

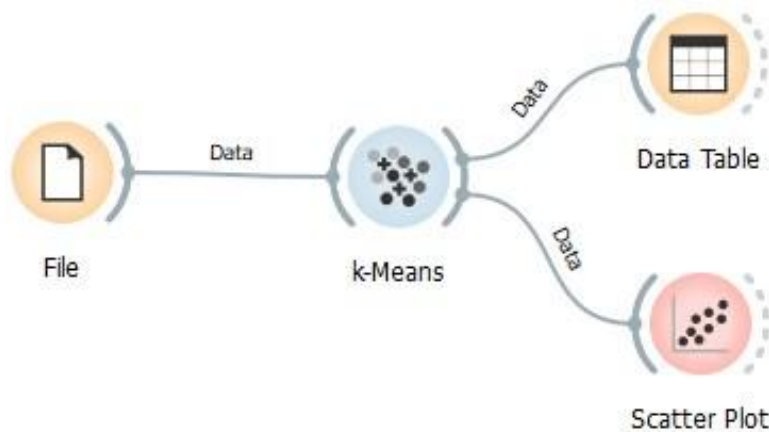
Province	Cluster	Silhouette	UH+	HLS	RLS	Pengeukuran
1 Aceh	C2	0.581858	70.18	14.37	9.44	9963
2 Sumatera Utara	C2	0.606669	69.61	13.31	9.71	10640
3 Sumatera Barat	C2	0.594501	69.90	14.10	9.18	11130
4 Riau	C2	0.589735	71.95	13.29	9.22	11158
5 Jambi	C2	0.644388	71.50	13.05	8.68	10871
6 Sumatera Selatan	C2	0.652941	70.32	12.55	8.37	11109
7 Bengkulu	C2	0.633363	69.69	13.60	8.91	10540
8 Lampung	C2	0.658209	70.99	12.74	8.18	10336
9 Kepulauan Ban...	C2	0.601068	70.98	12.18	8.11	13358
10 Kepulauan Riau	C1	0.538918	70.50	12.99	10.37	14469
11 DKI Jakarta	C1	0.607063	73.32	13.08	11.31	18927
12 Jawa Barat	C2	0.598206	73.52	12.62	8.78	11277
13 Jawa Tengah	C2	0.588893	74.57	12.81	7.93	11377
14 D I Yogyakarta	C1	0.576783	75.08	15.65	9.75	14482
15 Jawa Timur	C2	0.620879	71.74	13.37	8.03	11992
16 Banten	C2	0.607576	70.39	13.05	9.13	12216
17 Bali	C1	0.52921	72.60	13.48	9.39	13942
18 Nusa Tenggara ...	C2	0.633583	67.07	13.96	7.61	10681
19 Nusa Tenggara ...	C2	0.655179	67.47	13.21	7.70	7877
20 Kalimantan Barat	C2	0.652838	71.02	12.66	7.59	9355
21 Kalimantan Ten...	C2	0.649562	70.04	12.75	8.65	11458
22 Kalimantan Sel...	C2	0.635786	69.13	12.82	8.46	12468
23 Kalimantan Tim...	C1	0.566146	74.62	13.84	9.92	12641
24 Kalimantan Utara	C2	0.613749	72.07	13.06	9.27	9350
25 Sulawesi Utara	C2	0.566622	72.08	12.95	9.68	11179
26 Sulawesi Tengah	C2	0.658785	68.93	13.32	8.89	9696
27 Sulawesi Selatan	C2	0.630288	70.97	13.53	8.63	11430
28 Sulawesi Tengah	C2	0.617979	71.37	13.69	8.25	9708
29 Gorontalo	C2	0.661451	68.51	13.12	8.02	10687
30 Sulawesi Barat	C2	0.648151	65.63	12.87	8.08	9358
31 Maluku	C2	0.589843	66.45	14.00	10.19	8876

Sumber: Hasil Penelitian (2023)

Gambar 9. Hasil dari *Clustering*

Pada gambar 9 menunjukkan Hasil dari *Clustering*.

- 6) Lalu kaitkan juga *Widget K-Means* dengan *Widget Scatter Plot* untuk menampilkan visualisasi hasil *K-Means Clustering*.

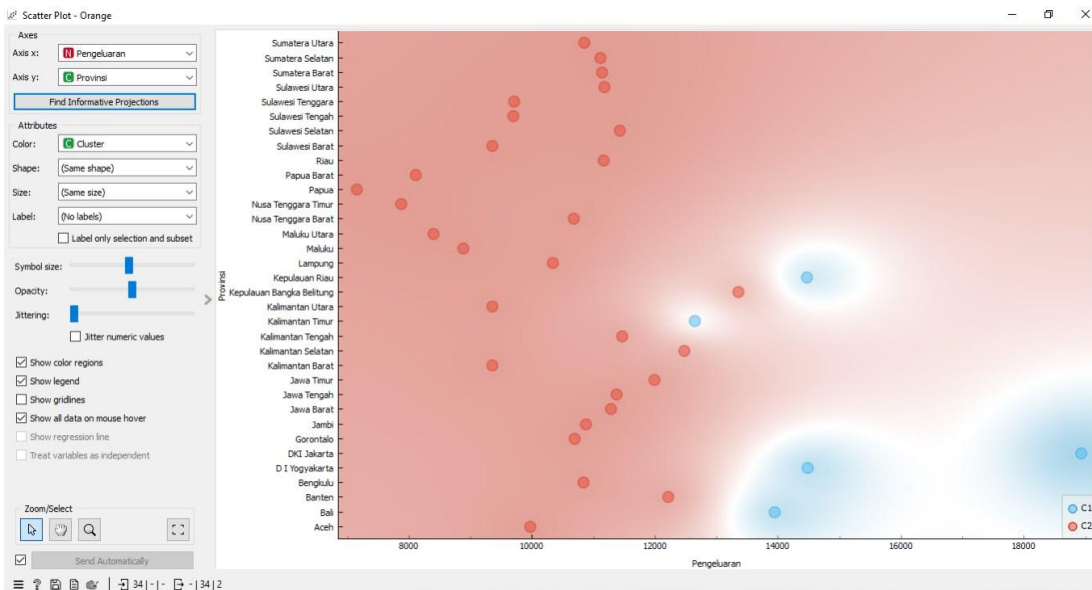


Sumber: Hasil Penelitian (2023)

Gambar 10. Pemodelan *K-Means Clustering*

Pada gambar 10 menunjukkan Pemodelan *K-Means Clustering*.

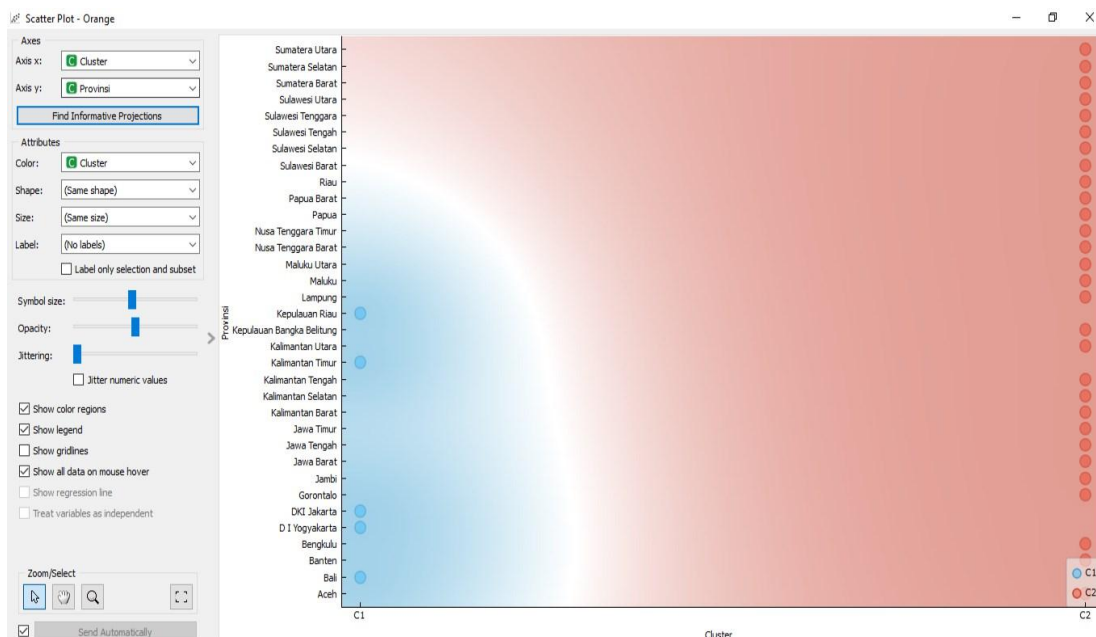
- 7) Klik *Widget Scatter Plot*, Lalu klik *Find Informative Projections* untuk melihat hasil visualisasi paling bagus.



Sumber: Hasil Penelitian (2023)

Gambar 11. Hasil *Clustering* hubungan Pengeluaran dengan Provinsi

Pada gambar 11 menunjukkan Hasil *Clustering* hubungan Pengeluaran dengan Provinsi.



Sumber: Hasil Penelitian (2023)

Gambar 12. Hasil *Clustering* hubungan *Cluster* dengan Provinsi

Pada gambar 12 menunjukkan penggunaan *Orange Data Mining* mengimplementasikan *K-Means Clustering* dengan melibatkan beberapa widget seperti *File*, *K-Means*, *Data Table*, dan *Scatter Plot*. Dari 34 data yang dianalisis, terbentuk 2 klaster, di mana 5 data termasuk dalam klaster 1, sedangkan 29 data lainnya termasuk ke dalam klaster 2. Tabel 1 menampilkan jumlah klaster untuk setiap nilai K yang didapatkan.

Tabel 1. Anggota Setiap Cluster

Cluster	Anggota Cluster
<i>Cluster 1</i>	Provinsi Kepulauan Riau, Provinsi DKI Jakarta, Provinsi D.I Yogyakarta, Provinsi Bali, Provinsi Kalimantan Timur
<i>Cluster 2</i>	Provinsi Aceh, Provinsi Sumatera Utara, Provinsi Sumatera Barat, Provinsi Riau, Provinsi Jambi, Provinsi Sumatera Selatan, Provinsi Bengkulu, Provinsi Lampung, Provinsi Kepulauan Bangka Belitung, Provinsi Jawa Barat, Provinsi Jawa Tengah, Provinsi Jawa Timur, Provinsi Banten, Provinsi Nusa Tenggara Barat, Provinsi Nusa Tenggara Timur, Provinsi Kalimantan Barat, Provinsi Kalimantan Tengah, Provinsi Kalimantan Selatan, Provinsi Kalimantan Utara, Provinsi Sulawesi Utara, Provinsi Sulawesi Tengah, Provinsi Sulawesi Selatan, Provinsi Sulawesi Tenggara, Provinsi Gorontalo, Provinsi Sulawesi Barat, Provinsi Maluku, Provinsi Maluku Utara, Provinsi Papua Barat, Provinsi Papua

Sumber: Hasil Penelitian (2023)

Pada Tabel 1 menunjukkan Anggota Setiap Cluster.

4. Kesimpulan

Berdasarkan temuan dari penelitian ini, dapat disimpulkan bahwa algoritma *K-Means Clustering* sukses diaplikasikan atas data parameter Indeks Pembangunan Manusia (IPM) Provinsi di Indonesia pada tahun 2022 menggunakan *Orange Data Mining*. Algoritma ini berhasil mengelompokkan nilai IPM menjadi dua klaster, yang mencakup kategori IPM tergolong sedang hingga tinggi, serta kategori IPM tinggi hingga sangat tinggi. Hal ini

didasarkan pada indikator penyusun IPM, seperti Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran per Kapita.

- *Cluster 1* yaitu Provinsi dengan indikator IPM tinggi hingga sangat tinggi terdapat 5 data, yaitu Provinsi Kepulauan Riau, Provinsi DKI Jakarta, Provinsi D.I Yogyakarta, Provinsi Bali, Provinsi Kalimantan Timur.
- *Cluster 2* yaitu Provinsi dengan indikator IPM sedang hingga tinggi terdapat 29 data, yaitu Provinsi Aceh, Provinsi Sumatera Utara, Provinsi Sumatera Barat, Provinsi Riau, Provinsi Jambi, Provinsi Sumatera Selatan, Provinsi Bengkulu, Provinsi Lampung, Provinsi Kepulauan Bangka Belitung, Provinsi Jawa Barat, Provinsi Jawa Tengah, Provinsi Jawa Timur, Provinsi Banten, Provinsi Nusa Tenggara Barat, Provinsi Nusa Tenggara Timur, Provinsi Kalimantan Barat, Provinsi Kalimantan Tengah, Provinsi Kalimantan Selatan, Provinsi Kalimantan Utara, Provinsi Sulawesi Utara, Provinsi Sulawesi Tengah, Provinsi Sulawesi Selatan, Provinsi Sulawesi Tenggara, Provinsi Gorontalo, Provinsi Sulawesi Barat, Provinsi Maluku, Provinsi Maluku Utara, Provinsi Papua Barat, Provinsi Papua.

Daftar Pustaka

- Aghnia, R., Wiguna, R., & Rifai, A. I. (2021). Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining. *Journal of Information Systems and Informatics*, 3(1). <http://journal-isi.org/index.php/isi>
- Akhmad Budi, dan. (2019). *KLASTERISASI INDEKS PEMBANGUNAN MANUSIA (IPM) PER KABUPATEN DI INDONESIA DENGAN MENGGUNAKAN ALGORITMA K-MEANS*. <http://www.kwikkiangie.ac.id>
- INDEKS PEMBANGUNAN MANUSIA 2022*. (n.d.). Retrieved June 27, 2023, from <https://www.bps.go.id>
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Nazuah, S., Shofia Hilabi, S., Hananto, A., Huda, B., Sistem Informasi, P., Ilmu Komputer, F., & Buana Perjuangan Karawang, U. (2023). *Jurnal Sistem dan Teknologi Informasi Indonesia Seleksi Penerimaan Beasiswa Dengan Metode K-Means Clustering Menggunakan Orange Selection of Scholarship Recipients Using the K-Means Clustering Method Using Orange*. 8(1), 18–27. <https://doi.org/10.32528/justindo.v8i1.212>
- Sibarani, H., Saputra, W., Gunawan, I., Masruro Nasution Program Studi Teknik Informatika STIKOM Tunas Bangsa Pematangsiantar, Z., & Jendral Sudirman Blok, J. A. (2022). PENERAPAN METODE K-MEANS UNTUK PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI SUMATERA UTARA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 1).

- Syukron Nawawi, M., Sembiring, F., & Erfina, A. (2021). " *PROGRAM STUDI TEKNIK INFORMATIKA-UNIVERSITAS PGRI MADIUN*" | 789 *Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris.*
- United Nations Development Programme. (n.d.). *Human development report 2021/2022 : uncertain times, unsettled lives : shaping our future in a transforming world.*
- Utami, N. W., Istri, A. A., Paramitha, I., Program,), Sistem, S., Akuntansi, I., Informasi, S. S., & Primakara, S. (2021). *PENERAPAN DATA MINING UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI DI STMIK PRIMAKARA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING.*
- Widiastari, A. (2021). (media cetak) Analisa Datamining dengan Metode Klasifikasi C4.5 Sebagai Faktor Penyebab Tanah Longsor. *Journal of Computer System and Informatics (JoSYC)*, 2(3).