

Manual Book
**Sistem Pemetaan Wilayah Berdasarkan Tingkat Pengangguran
Terbuka dan Tingkat Partisipasi Angkatan Kerja di Indonesia
dengan K-Means dan DBSCAN**



UNTAR
Universitas Tarumanagara

Oleh:
Jessica Ho - 535220187

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TARUMANAGARA
2025**

DAFTAR ISI

DAFTAR ISI.....	2
PENDAHULUAN.....	3
PANDUAN PENGGUNAAN APLIKASI.....	5
Halaman Home (Beranda).....	5
Halaman Dataset.....	8
Halaman Clustering.....	9

PENDAHULUAN

Dokumen ini dirancang untuk memberikan panduan bagi pengguna agar dapat memanfaatkan fitur aplikasi secara jelas. Sistem ini dibangun untuk menganalisis dua indikator ketenagakerjaan, yaitu Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK). Biasanya kondisi kedua indikator ini sangat beragam antar wilayah kabupaten/kota, menyebabkan terjadinya ketimpangan dan perlu adanya analisis untuk mengelompokkan wilayah berdasarkan kesamaan karakteristik kedua indikator tersebut.

Aplikasi ini menggunakan teknik dalam machine learning yang dapat membantu analisis tersebut, sehingga dapat memudahkan pengguna dalam memahami kondisi ketenagakerjaan pada suatu wilayah berdasarkan hasil evaluasi, peta wilayah, dan grafik yang interaktif.

Sistem ini beroperasi menggunakan dua sumber data, yaitu:

1. Data TPT dan TPAK kabupaten/kota di Indonesia tahun 2018-2024 yang disimpan dalam bentuk file Excel yang bersumber dari Badan Pusat Statistik (BPS) dari setiap provinsi.
2. Data geospasial berbentuk GeoJSON berisi data batas wilayah (poligon) untuk setiap kabupaten/kota di Indonesia.

Dataset pada sistem ini melalui alur kerja sebagai berikut:

1. Prapemrosesan Data

- a. Pengisian *missing value*

Pengisian dilakukan dengan metode:

- Forward fill: pengisian nilai kosong dengan nilai terakhir dari baris sebelumnya. Hal ini cocok jika data kosong berada di akhir.
- Interpolasi linear: pengisian nilai kosong dengan cara menghubungkan dua titik. Hal ini cocok jika data kosong berada di tengah.
- Backward fill: pengisian nilai kosong dengan nilai dari baris setelahnya. Hal ini cocok jika data kosong berada di awal.

- b. Normalisasi data

Normalisasi data adalah proses mengatur ulang data agar variabel memiliki bobot yang sama selama proses clustering. Sistem ini menggunakan Z-Score sebagai metode untuk normalisasi data.

2. Algoritma Clustering

Sistem menggunakan dua algoritma clustering, yaitu K-Means dan DBSCAN. K-Means membagi data berdasarkan jumlah cluster (K) yang dipilih oleh pengguna, sesuai dengan jarak terdekat pada masing-masing titik pusat (centroid). Sedangkan, DBSCAN membagi data berdasarkan kepadatan data melalui parameter Epsilon (jarak maksimum antar titik untuk dianggap sebagai tetangga) dan MinPts (jumlah minimum titik dalam radius Epsilon untuk membentuk sebuah cluster).

3. Metrik Evaluasi

Setelah melalui proses clustering, hasil akan dievaluasi untuk mengetahui seberapa baik kualitas suatu cluster dan apakah cluster tersebut lebih kompak. Sistem menggunakan 2 metrik, yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Semakin tinggi nilai Silhouette, data dalam cluster semakin sesuai dengan tempatnya, artinya cluster lumayan benar. Sedangkan, semakin kecil nilai DBI, cluster lebih kompak dan terpisah, menandakan sedikit atau tidak ada tumpang tindih.

PANDUAN PENGGUNAAN APLIKASI

Sistem terdiri dari tiga halaman utama yang dapat diakses, yaitu halaman Home, Dataset, dan Clustering. Halaman tersebut dapat diakses melalui navigasi utama yang berada di atas layar.



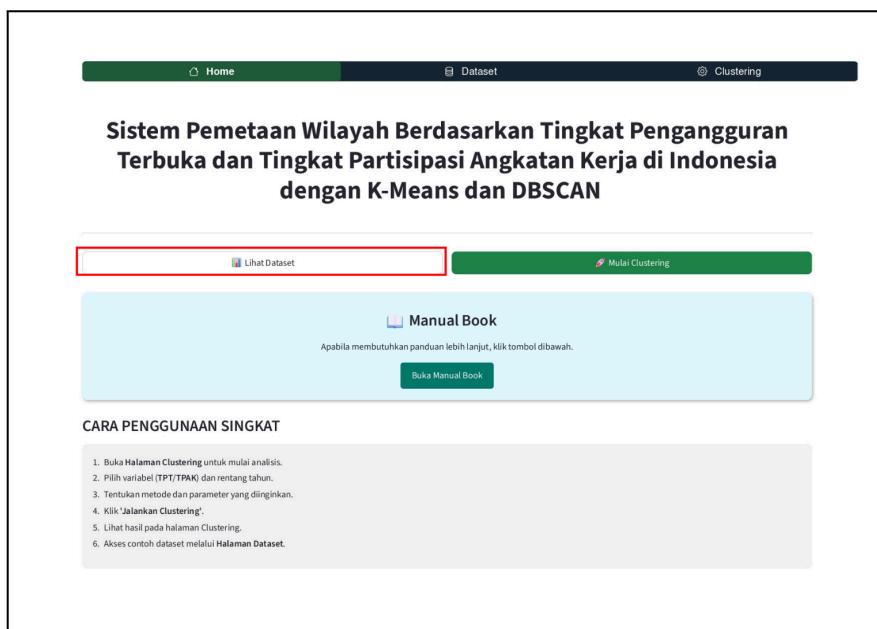
Sistem Pemetaan Wilayah Berdasarkan Tingkat Pengangguran Terbuka dan Tingkat Partisipasi Angkatan Kerja di Indonesia dengan K-Means dan DBSCAN

Halaman Home (Beranda)

Halaman Home merupakan halaman awal sebagai pengantar sistem. Halaman ini berisi judul sistem, cara penggunaan singkat, dan informasi perihal latar belakang. Adapun beberapa fitur pada halaman ini:

a. Tombol " Lihat Dataset"

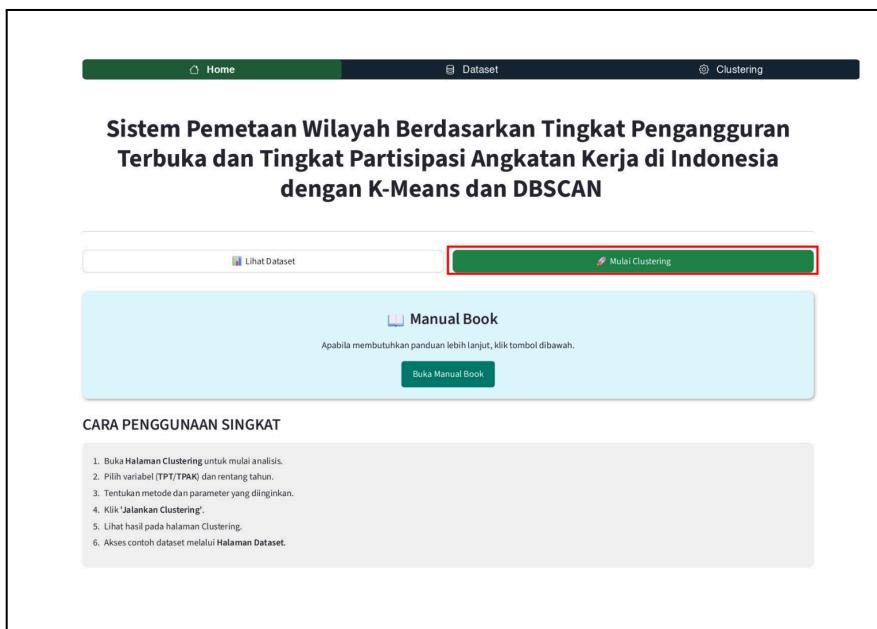
Tombol ini mengarahkan pengguna untuk ke halaman Dataset.



1. Buka Halaman Clustering untuk mulai analisis.
2. Pilih variabel (TP/T/TPAK) dan rentang tahun.
3. Tentukan metode dan parameter yang diinginkan.
4. Klik 'Jalankan Clustering'.
5. Lihat hasil pada halaman Clustering.
6. Akses contoh dataset melalui Halaman Dataset.

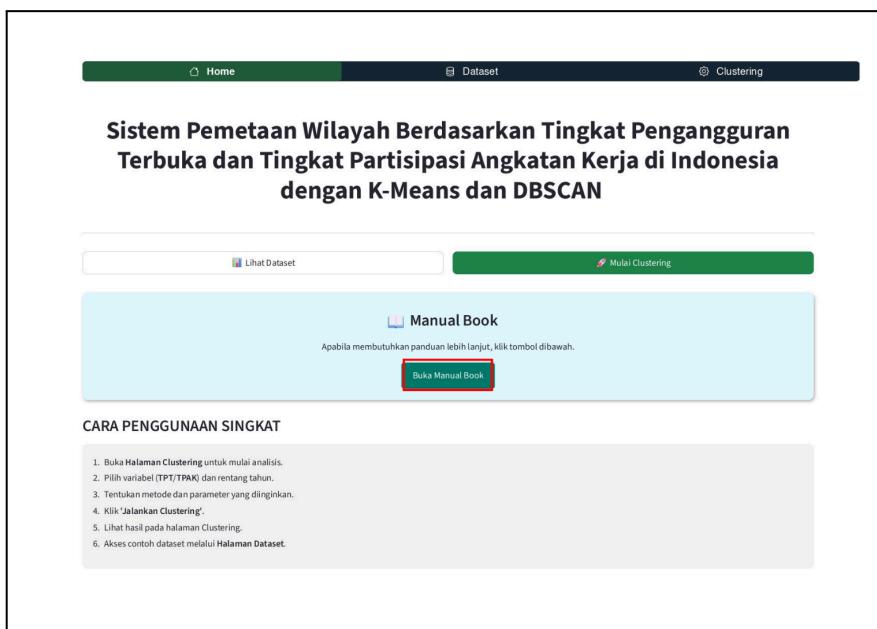
b. Tombol "🚀 Mulai Clustering"

Tombol ini mengarahkan pengguna untuk ke halaman Clustering.



c. Tombol “Buka Manual Book”

Tombol ini mengarahkan pengguna untuk mengakses manual book.



d. FAQ (Frequently Asked Questions)

Berisi jawaban atas pertanyaan umum, seperti definisi TPT/TPAK dan penjelasan singkat tentang K-Means, DBSCAN, dan metrik evaluasi yang digunakan. Ketika pertanyaan tersebut dibuka, maka akan tampil jawaban dari pertanyaan tersebut.

A screenshot of a web-based FAQ section. At the top, there is a note: "• TPT & TPAK rendah -> potensi tenaga kerja belum dimanfaatkan optimal." Below this is a descriptive text about the research methodology. Two questions are listed under a red-bordered "FAQ" heading:

- > Apa itu Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK)?
- > Metode clustering dan metrik evaluasi apa yang digunakan dalam sistem ini?

A screenshot of the same FAQ section after expanding the two questions. The first question, "Apa itu Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK)?", reveals a detailed answer about the indicators being used to measure labor force participation. The second question, "Metode clustering dan metrik evaluasi apa yang digunakan dalam sistem ini?", also reveals a detailed answer, including sections on K-Means and DBSCAN clustering methods and evaluation metrics like Silhouette Score and Davies-Bouldin Index.

Halaman Dataset

Halaman ini bertujuan untuk menunjukkan dataset awal yang digunakan untuk proses clustering, berisi:

a. Tampilan dataset

Menampilkan tabel data lengkap berupa data TPT dan TPAK tahun 2018-2024 pada setiap kabupaten/kota di Indonesia. Data yang ditampilkan sudah melalui proses pembersihan data non-numerik.

The screenshot shows a web application interface with a dark header bar containing 'Home', 'Dataset' (which is highlighted in green), and 'Clustering'. Below the header is a section titled 'TAMPILAN DATASET ASLI' with a subtitle 'Dataset bersumber dari BPS (Badan Pusat Statistik) per provinsi.' A large table follows, showing data for 10 provinces (ACEH, Simeulue, Aceh Singkil, Aceh Selatan, Aceh Tengara, Aceh Timur, Aceh Tengah, Aceh Barat, Aceh Besar, Pidie, Bireuen) across 16 years (2018-2024). The table has columns for province, kab_kota, and various TPT and TPAK values. At the bottom left of the table area, it says 'Total 514 baris dan 16 kolom.' Below the table is a button labeled 'Penjelasan Kolom Dataset'.

prov	kab_kota	TPT_2018	TPT_2019	TPT_2020	TPT_2021	TPT_2022	TPT_2023	TPT_2024	TPAK_2018	TPAK_2019	TPAK_2020	TPAK_2021	TPAK_2022	TPAK_2023	TPAK_2024
0	ACEH Simeulue	4.95	5.82	5.47	5.71	6	5.85	5.5	64.37	62.5	70.37	71.15	64.44	70.57	71.08
1	ACEH Aceh Singkil	7.96	8.58	8.24	8.36	6.88	6.84	6.44	63.11	59.09	61.97	62.85	57.33	57.81	57.82
2	ACEH Aceh Selatan	6.03	6.54	6.54	6.46	4.82	4.73	4.56	65.42	59	61.41	60.85	60.69	58.87	59.24
3	ACEH Aceh Tengara	3.75	3.45	5.72	6.43	5.09	5	4.79	72.81	67.41	71.33	69.62	67.82	70.43	70.51
4	ACEH Aceh Timur	6.92	7.61	7.26	7.13	8.07	8.03	7.75	65.75	59.37	61.92	59.48	58.45	60.48	60.84
5	ACEH Aceh Tengah	2.11	2.65	3.05	2.61	4.44	4.42	4.21	72.49	72.5	79.96	76.3	76.37	76.85	77.41
6	ACEH Aceh Barat	8.58	7.41	7.3	7.09	6.09	6.07	5.58	54.27	60.9	59.41	60.05	57.8	59.62	60.23
7	ACEH Aceh Besar	7.29	7.67	7.62	7.7	8.28	8.17	7.93	58.51	56.18	60.02	61.67	65.31	66.86	67.05
8	ACEH Pidie	7.24	6.83	6.49	7.28	5.94	5.92	5.74	61.73	63.05	65.55	60.29	65.47	61.57	62.17
9	ACEH Bireuen	3.52	3.83	4.12	4.32	4.2	4.14	3.93	64.79	66.88	67.06	65.91	64.39	66.96	67.24

b. Penjelasan kolom dataset

Merupakan penjelasan masing-masing kolom dalam tabel supaya mudah dipahami.

The screenshot shows a modal window titled 'Penjelasan Kolom Dataset' with a list of 24 items, each starting with a bullet point. The items describe various columns related to provinces, kabupaten/kota, and specific TPT and TPAK values for the years 2018 through 2024.

- Provinsi: Nama provinsi.
- Kabupaten/Kota: Nama kabupaten/kota.
- TPT_2018: Tingkat pengangguran terbuka tahun 2018.
- TPT_2019: Tingkat pengangguran terbuka tahun 2019.
- TPT_2020: Tingkat pengangguran terbuka tahun 2020.
- TPT_2021: Tingkat pengangguran terbuka tahun 2021.
- TPT_2022: Tingkat pengangguran terbuka tahun 2022.
- TPT_2023: Tingkat pengangguran terbuka tahun 2023.
- TPT_2024: Tingkat pengangguran terbuka tahun 2024.
- TPAK_2018: Tingkat partisipasi angkatan kerja tahun 2018.
- TPAK_2019: Tingkat partisipasi angkatan kerja tahun 2019.
- TPAK_2020: Tingkat partisipasi angkatan kerja tahun 2020.
- TPAK_2021: Tingkat partisipasi angkatan kerja tahun 2021.
- TPAK_2022: Tingkat partisipasi angkatan kerja tahun 2022.
- TPAK_2023: Tingkat partisipasi angkatan kerja tahun 2023.
- TPAK_2024: Tingkat partisipasi angkatan kerja tahun 2024.

Pengguna dapat mengunduh dataset dalam bentuk .csv melalui tombol download yang ada pada tabel dataset, melakukan search, dan menampilkan tabel secara fullscreen. Selain itu, pengguna juga dapat melakukan sort secara ascending atau descending, melakukan pin column, dan hide column.

Halaman Clustering

Halaman ini merupakan halaman untuk melakukan proses analisis dan visualisasi. Halaman ini memungkinkan pengguna untuk memilih variabel, tahun, metode, dan parameter yang diinginkan. Setelah proses pemilihan selesai, pengguna dapat menjalankan analisis dan sistem akan menampilkan hasil berupa: hasil metrik evaluasi dan silhouette plot, tabel jumlah wilayah per cluster, box plot, peta wilayah interaktif, scatter pair plot, dan tabel hasil.

The screenshot shows the 'Clustering' tab selected in the top navigation bar. The main title is 'Sistem Pemetaan Wilayah Berdasarkan Tingkat Pengangguran Terbuka dan Tingkat Partisipasi Angkatan Kerja di Indonesia dengan K-Means dan DBSCAN'. On the left, there's a section titled 'PILIH VARIABEL & TAHUN' with two checkboxes: 'Tingkat Pengangguran Terbuka (TPT)' and 'Tingkat Partisipasi Angkatan Kerja (TPAK)'. Below these is a button 'Pilih variabel (TPT/TPAK) terlebih dahulu'. To the right, there's a section titled 'HASIL METRIK & GRAFIK' which is currently empty. At the bottom, there's a slider for 'Pilih Range Tahun' with values 2021 and 2024, and a message 'Data tidak valid.'.

Berikut merupakan fitur yang terdapat pada halaman Clustering.

a. Memilih variabel dan tahun

Pengguna dapat memilih salah satu variabel (TPT atau TPAK) atau keduanya. Selain itu, pengguna juga dapat memilih range tahun pada slider dalam range 2018-2024.

This screenshot is similar to the previous one, but the 'PILIH VARIABEL & TAHUN' section is highlighted with a red box. The rest of the interface remains the same, including the empty 'HASIL METRIK & GRAFIK' area and the 'Data tidak valid.' message at the bottom.

Apabila pengguna belum memilih variabel sama sekali, maka akan tampil informasi "Pilih variabel (TPT/TPAK) terlebih dahulu" dan tidak bisa melakukan pemilihan range tahun. Sehingga diperlukan pemilihan variabel terlebih dahulu, baru dilanjutkan dengan pemilihan range tahun.

b. Memilih Metode Clustering dan Parameter

Setelah pemilihan variabel dan tahun, pengguna dapat memilih metode clustering yang diinginkan: K-Means ataupun DBSCAN. Adapun pengguna juga dapat memilih parameter optimal. Berikut merupakan contoh ketika melakukan pemilihan metode K-Means. Jika pengguna memilih metode K-Means, pengguna dapat memilih nilai parameter K

dalam range 2 sampai dengan 6, bahkan memilih Nilai K Optimal berdasarkan nilai Silhouette.

METODE & PARAMETER

K-Means
 DBSCAN

Nilai K Optimal ⓘ

K (Jumlah cluster yang ingin dibentuk)

2

K pada K-Means berpengaruh pada seberapa banyak kelompok/cluster yang terbentuk. Apabila pengguna memilih K=2, maka jumlah kelompok/cluster yang akan terbentuk adalah 2. Begitu juga, jika pengguna memilih K=5, maka jumlah kelompok/cluster yang terbentuk adalah 5.

Sedangkan, berikut merupakan contoh ketika melakukan pemilihan metode DBSCAN. Jika pengguna memilih metode DBSCAN, pengguna dapat memilih parameter epsilon (radius jarak maksimum) dalam range 0.1 sampai 25.00 dan MinPts dalam range 2 sampai 20 (jumlah minimum titik yang diperlukan untuk membentuk sebuah klaster inti). Adapun fitur apabila pengguna ingin menerapkan reduksi dimensi dengan PCA jika $D \geq 3$, bahkan memilih parameter Epsilon dan MinPts yang optimal.

METODE & PARAMETER

K-Means

DBSCAN

Nilai Epsilon & MinPts Optimal ?

Epsilon (Jarak maksimum antar titik untuk dianggap sebagai tetangga)

0.50

MinPts (Jumlah minimum titik dalam radius Epsilon untuk membentuk sebuah cluster)

3

Terapkan PCA (jika $D \geq 3$) ?

Parameter DBSCAN tidak mampu langsung menentukan jumlah cluster yang terbentuk. Nilai Epsilon dan MinPts berpengaruh terhadap pembentukan cluster dengan mempertimbangkan kepadatan titik-titik data yang memenuhi kriteria tersebut. Oleh karena itu, diberi kemudahan dengan fitur “Nilai Epsilon & MinPts Optimal”.

c. Tombol “Jalankan & Clustering Peta”

Tombol ini digunakan untuk menjalankan analisis clustering sehingga dapat menghasilkan output berupa hasil metrik evaluasi, silhouette plot, peta wilayah, dan grafik lainnya. Sebelum menjalankan analisis, diharapkan pengguna dapat memilih variabel, range tahun, metode, dan nilai parameter yang diinginkan.

METODE & PARAMETER

K-Means

DBSCAN

Epsilon (Jarak maksimum antar titik untuk dianggap sebagai tetangga)

0.50

MinPts (Jumlah minimum titik dalam radius Epsilon untuk membentuk sebuah cluster)

2

Jalankan Clustering & Peta

Jika pengguna tidak memilih variabel apapun, maka sistem akan menampilkan alert yang menunjukkan bahwa untuk menjalankan proses analisis perlu memilih minimal satu variabel terlebih dahulu. Berikut merupakan contoh alert yang muncul.

Silakan pilih minimal satu variabel (TPT atau TPAK)
terlebih dahulu sebelum menjalankan.

Tombol akan melalui proses sebagai berikut.

1. Mengambil data sesuai pilihan variabel dan tahun yang dipilih
2. Melakukan preprocessing (pengisian missing value & normalisasi).
3. Menjalankan algoritma dengan parameter yang ditentukan.
4. Menghitung metrik evaluasi (Silhouette & DBI), serta waktu proses clustering.
5. Menghasilkan visualisasi: Box Plot, Peta wilayah interaktif, Pair Plot, dan tabel hasil.
6. Menyimpan semua hasil ke dalam *session* aplikasi untuk ditampilkan.

d. Menampilkan Output

Setelah proses clustering selesai, hasil akan muncul berupa:

1) Metrik & Grafik

HASIL METRIK & GRAFIK

Data tidak valid.

a) Silhouette score

Semakin tinggi nilai Silhouette, data dalam cluster semakin sesuai dengan tempatnya, artinya cluster lumayan benar.

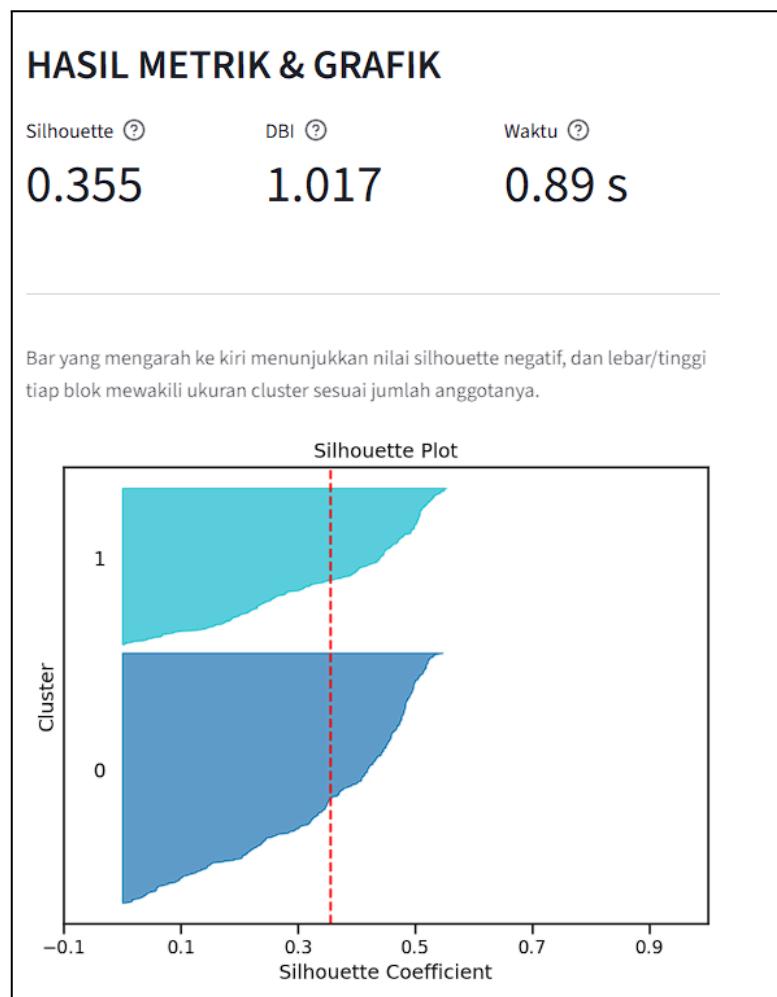
b) Davies-Bouldin Index (DBI)

Semakin kecil nilai DBI, cluster lebih kompak dan terpisah, menandakan sedikit atau tidak ada tumpang tindih.

c) Waktu: Waktu komputasi yang diperlukan (dalam detik).

- d) Silhouette Plot: Visualisasi untuk menilai kualitas klaster berdasarkan silhouette score. Cara membaca plot: Setiap bar mewakili satu cluster. Lebar/tinggi tiap blok mewakili jumlah anggotanya. Bar yang mengarah ke kiri menunjukkan nilai silhouette negatif. Garis putus-putus merah adalah nilai silhouette rata-rata.

Contoh tampilan hasil metrik dan grafik adalah sebagai berikut.



Masing-masing metrik memiliki tooltip yang jika dihover maka akan tampil penjelasan mengenai tiap metrik. Adapun penjelasan cara membaca silhouette plot pada layar.

Jika cluster yang terbentuk hanya 1, maka akan muncul tampilan seperti berikut.

Silhouette plot memerlukan minimal 2 cluster valid.

➤ Mengapa Silhouette Plot tidak muncul?

Penjelasan mengapa Silhouette Plot tidak muncul, dapat dilihat pada dropdown berikut.

▼ Mengapa Silhouette Plot tidak muncul?

Silhouette plot berguna untuk membandingkan seberapa mirip sebuah titik dengan **cluster-nya sendiri** dibandingkan dengan **cluster tetangga terdekat**. Agar perbandingan ini bisa dilakukan, sistem membutuhkan minimal **dua cluster valid** (selain Noise).

1. Semua Wilayah Menjadi "Noise" (Label -1)

- Arti: Tidak ada wilayah yang cukup padat untuk membentuk cluster.
- Penyebab: Nilai **Epsilon** terlalu kecil atau **MinPts** terlalu tinggi.

2. Semua Wilayah Menjadi 1 Cluster (Label 0)

- Arti: Semua wilayah termasuk dalam satu cluster besar tanpa ada yang dianggap Noise.
- Penyebab: Nilai **Epsilon** terlalu besar. Semua titik dianggap 'tetangga' satu sama lain.

3. Hanya Ada 1 Cluster + Noise (Label 0 dan -1)

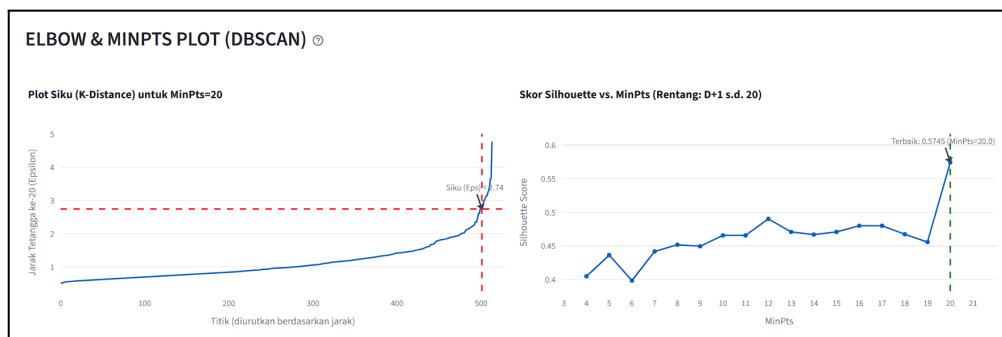
- Arti: Hanya ada satu cluster yang terbentuk, sementara wilayah lainnya dianggap Noise karena terlalu tersebar.
- Penyebab: Nilai **Epsilon** dan **MinPts** sudah tepat untuk membentuk satu cluster, tetapi tidak cukup untuk membentuk cluster tambahan.

2) Pencarian K Optimal (K-Means) atau Elbow & MinPts Plot (DBSCAN)

Apabila pengguna memilih fitur K optimal pada K-Means, maka akan tampil tabel hasil pencarian K optimal dengan isi berupa jumlah K dan nilai Silhouettenya. K yang optimal akan ditandai dengan kolom berwarna biru.

PENCARIAN K-OPTIMAL (K-MEANS) Ⓜ	
Jumlah K	Silhouette
0	2 0.5147
1	3 0.4059
2	4 0.3344
3	5 0.2960
4	6 0.2652

Sedangkan, apabila pengguna memilih metode DBSCAN dengan fitur parameter optimal, maka akan tampil pencarian MinPts berdasarkan nilai Silhouette tertinggi dan pencarian epsilon menggunakan elbow plot. Hasil keduanya akan menghasilkan kombinasi parameter optimal.



3) Jumlah wilayah per cluster

JUMLAH WILAYAH PER CLUSTER Ⓜ	
Tabel ini menunjukkan jumlah wilayah yang termasuk dalam setiap cluster hasil clustering.	Belum ada hasil clustering untuk menampilkan jumlah wilayah per cluster.

- Menampilkan tabel yang merangkum jumlah kabupaten/kota yang masuk ke dalam setiap klaster. (Untuk DBSCAN, Klaster -1 berarti *Noise* atau *Outlier*).
- Adapun tooltip yang dapat menunjukkan penjelasan tentang tabel tersebut.

JUMLAH WILAYAH PER CLUSTER Ⓜ	
Tabel ini menunjukkan jumlah wilayah yang termasuk dalam setiap cluster hasil clustering.	Belum ada hasil clustering untuk menampilkan jumlah wilayah per cluster.

Contoh tampilan tabel tersebut setelah dilakukan proses clustering.

JUMLAH WILAYAH PER CLUSTER ⓘ

	Cluster	Jumlah_wilayah
0	0	316
1	1	198

4) Distribusi Cluster (Box Plot)

Box plot bertujuan sebagai visualisasi yang dapat menginterpretasi makna setiap cluster. Grafik ini juga membantu interpretasi peta wilayah.

DISTRIBUSI CLUSTER (BOX PLOT) ⓘ

Data tidak valid.

Cara Membaca: Grafik ini menunjukkan distribusi nilai (median, kuartil, min/max) untuk setiap variabel yang dipilih, dipecah per klaster. Penjelasan box plot dapat dilihat dengan cara melakukan hover/menekan tombol tooltip.

Box plot digunakan untuk melihat distribusi nilai variabel pada tiap cluster.

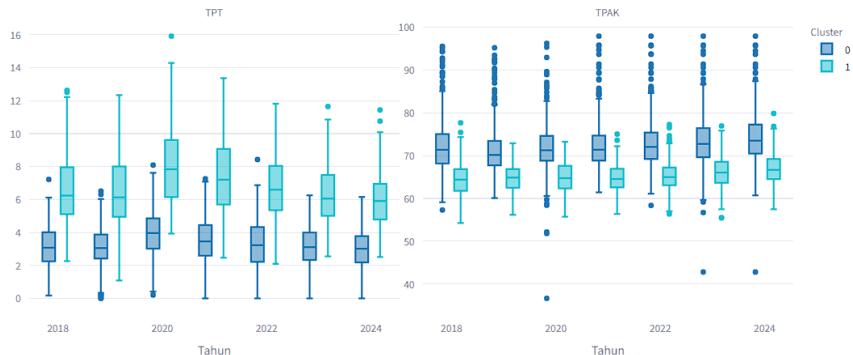
DISTRIBUSI CLUSTER (BOX PLOT) ⓘ ↗

Data tidak valid.

Contoh hasil box plot setelah dilakukan proses clustering adalah sebagai berikut.

DISTRIBUSI CLUSTER (BOX PLOT) ⓘ

Distribusi Nilai Variabel per Cluster dan Tahun



Contoh Interpretasi: Cluster 0 (biru tua) adalah cluster dengan nilai TPT yang rendah dan nilai TPAK yang tinggi. Sedangkan, Cluster 1 (biru muda) adalah sebaliknya."

5) Peta Wilayah Interaktif

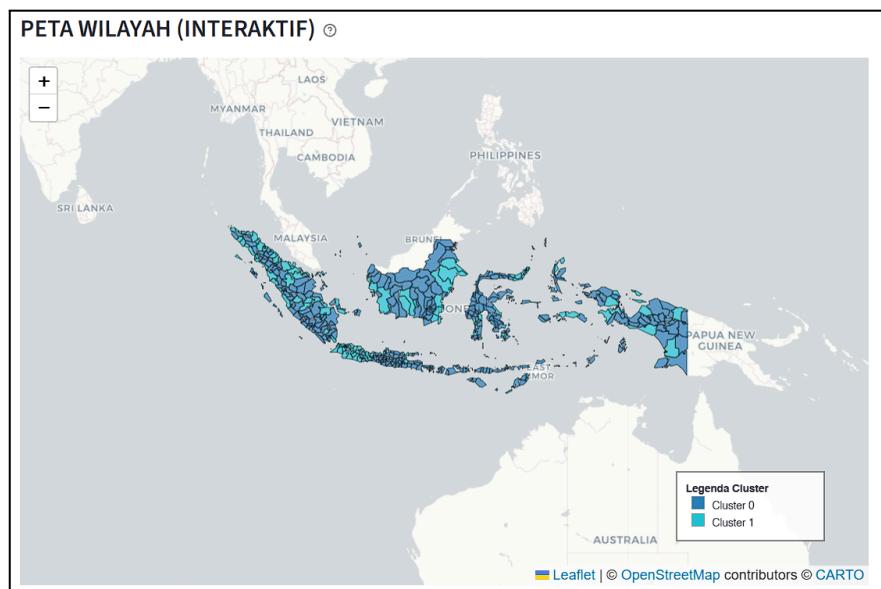
Peta wilayah (kabupaten/kota) di Indonesia ditampilkan diwakili oleh sebuah poligon kabupaten/kota yang memiliki warna sesuai dengan cluster hasil analisis. Cara membacanya adalah dengan menggunakan Box Plot, sesuai dengan penjelasan yang ada pada tooltip tersebut.



Fitur yang dimiliki:

- Zoom & Pan: Peta dapat diperbesar/diperkecil dan digeser.
- Hover/Tooltip: Kursor dapat diarahkan ke sebuah wilayah untuk melihat nama Provinsi, nama Kabupaten/Kota, dan nomor clusternya.
- Legenda: Terdapat legenda di pojok kanan bawah yang menjelaskan warna untuk setiap klaster.

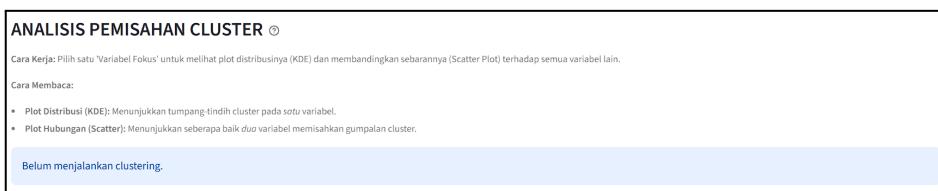
Sebagai gambaran peta wilayah hasil clustering adalah sebagai berikut.



6) Analisis Pemisahan Cluster

Analisis ini terdiri dari plot distribusi (KDE) dan plot hubungan (scatter plot). Keduanya berfungsi untuk menunjukkan tumpang-tindihnya cluster pada variabel tertentu dan menunjukkan seberapa baik kedua variabel memisahkan gumpalan cluster.

Pair plot membantu memvisualisasikan bagaimana cluster terpisah dalam ruang fitur (jika menggunakan lebih dari satu variabel).



Pada bagian ini, pengguna dapat memilih variabel fokus (untuk sumbu Y) agar dapat dilihat hubungannya dengan variabel lainnya yang dipilih pengguna di awal clustering.

ANALISIS PEMISAHAN CLUSTER

Cara Kerja: Pilih satu 'Variabel Fokus' untuk melihat plot distribusinya (KDE) dan membandingkan sebarannya (Scatter Plot) terhadap semua variabel lain.

Cara Membaca:

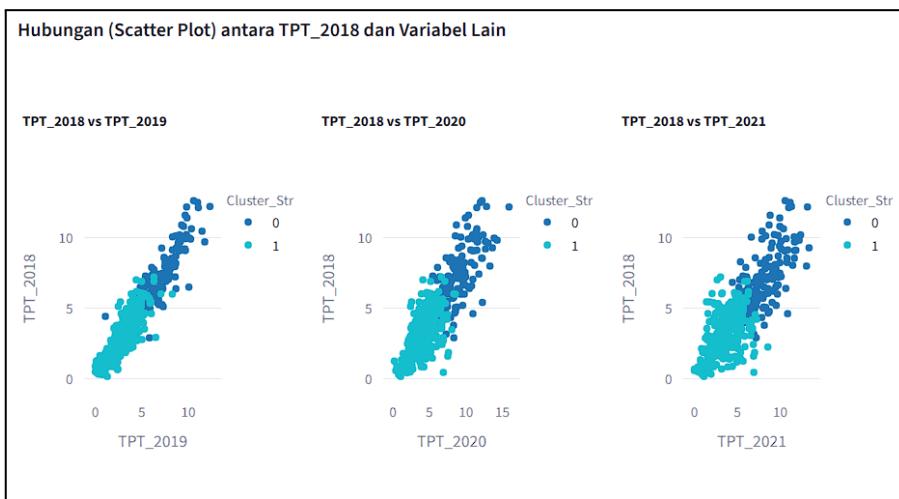
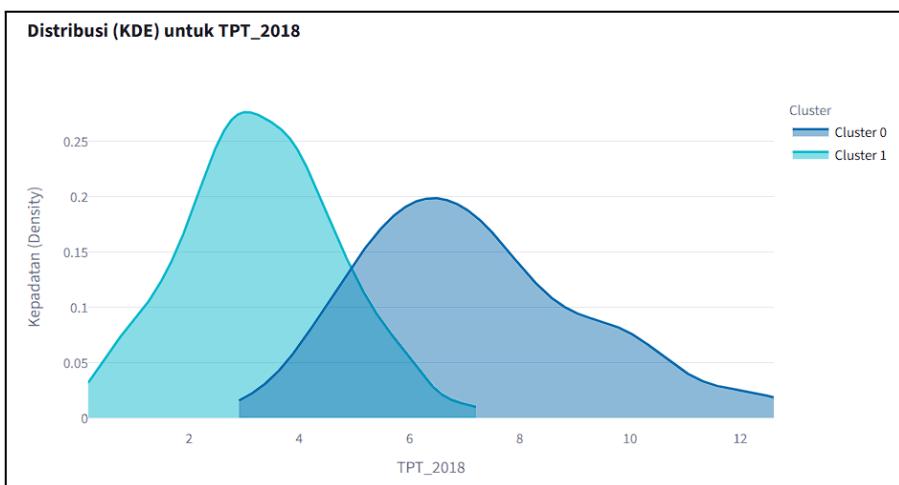
- Plot Distribusi (KDE): Menunjukkan tumpang-tindih cluster pada *satu* variabel.
- Plot Hubungan (Scatter): Menunjukkan seberapa baik *dua* variabel memisahkan gumpalan cluster.

Pilih Variabel Fokus (untuk Sumbu Y)

TPT_2018



Setelah pemilihan variabel fokus, akan didapatkan plot distribusi KDE dan Scatter plot seperti berikut.



7) Tabel dan Laporan Hasil

Bagian ini menampilkan tabel akhir yang berisi provinsi, kab_kota, nilai variabel yang dianalisis, dan kolom cluster hasil prediksi. Selain itu jika menggunakan DBSCAN, juga akan ada kolom point

type yang menggambarkan wilayah tersebut termasuk Core/Border/Noise.

TABEL & LAPORAN HASIL ↗

Belum ada hasil clustering...

Berikut merupakan contoh hasil tabel setelah analisis clustering dijalankan. Dalam hal ini menggunakan metode K-Means.

TABEL & LAPORAN HASIL

	prov	kab_kota	Cluster	TPT_2018	TPT_2019	TPT_2020	TPT_2021	TPT_2022	TPT_2023	TPT_2024	TPAK_2018	TPAK_2019	1
0	ACEH	Simeulue	1	4.95	5.82	5.47	5.71	6	5.85	5.5	64.37	62.5	
1	ACEH	Aceh Singkil	1	7.96	8.58	8.24	8.36	6.88	6.84	6.44	63.11	59.09	
2	ACEH	Aceh Selatan	1	6.03	6.54	6.54	6.46	4.82	4.73	4.56	65.42	59	
3	ACEH	Aceh Tenggara	0	3.75	3.45	5.72	6.43	5.09	5	4.79	72.81	67.41	
4	ACEH	Aceh Timur	1	6.92	7.61	7.26	7.13	8.07	8.03	7.75	65.75	59.37	
5	ACEH	Aceh Tengah	0	2.11	2.65	3.05	2.61	4.44	4.42	4.21	72.48	72.5	
6	ACEH	Aceh Barat	1	8.58	7.41	7.3	7.09	6.09	6.07	5.58	54.27	60.9	
7	ACEH	Aceh Besar	1	7.29	7.67	7.62	7.7	8.28	8.17	7.93	58.51	56.18	
8	ACEH	Pidie	1	7.24	6.83	6.45	7.28	5.94	5.92	5.74	61.73	63.05	
9	ACEH	Bireuen	0	3.52	3.83	4.12	4.32	4.2	4.14	3.93	64.79	66.88	

Adapula beberapa fitur berikut:

- Mengunduh tabel, mencari data, dan menampilkan data secara fullscreen.

Download Search Fullscreen				
TPT_2022	TPT_2023	TPT_2024	TPAK_2018	TPAK_2019
6	5.85	5.5	64.37	62.5
6.88	6.84	6.44	63.11	59.09
4.82	4.73	4.56	65.42	59
5.09	5	4.79	72.81	67.41
8.07	8.03	7.75	65.75	59.37

- Mensortir data, membuat ukuran tabel menjadi auto, pin column, dan menyembunyikan column tertentu.

	prov	kab_kota	:	Cluster	TPT_2018	TPT_2019
0	A	A	≡ kab_kota ✖	1	4.95	5.82
1	A		↑ Sort ascending	1	7.96	8.58
2	A		↓ Sort descending	1	6.03	6.54
3	A		↔ Autosize	0	3.75	3.45
4	A		⤿ Pin column	1	6.92	7.61
5	A		⤿ Hide column	0	2.11	2.65
6	ACEH	Aceh Barat		1	8.58	7.41
7	ACEH	Aceh Besar		1	7.29	7.67

Selain itu, terdapat fitur tambahan seperti:

- a. Tombol "Simpan Laporan PDF"

Tombol untuk menyimpan hasil analisis ke dalam laporan PDF yang dapat diunduh. Tombol ini berada paling bawah halaman ketika selesai analisis clustering.



Berikut merupakan isi laporan PDF:

1. Halaman Judul & Parameter Analisis (Metode, Variabel, Tahun, Parameter).
2. Hasil Metrik Evaluasi (Skor Silhouette, DBI, Waktu).
3. Visualisasi Silhouette Plot.
4. Visualisasi Peta Persebaran Cluster (versi statis).
5. Visualisasi Box Plot Distribusi Nilai (versi statis).
6. Statistik Deskriptif (mean, std, min, max) per Klaster.
7. Tabel Data Hasil Clustering (lengkap).

Setelah tombol diklik, akan muncul tombol "Unduh Laporan PDF" yang dapat digunakan untuk mengunduh file.

Preview isi PDF

Laporan Analisis Clustering TPT & TPAK

1. Parameter Analisis

Variabel yang Digunakan: tpt_tpak
Rentang Tahun Analisis: 2018 - 2024
Metode Clustering: DBSCAN
Parameter: eps=5.3, minpts=20

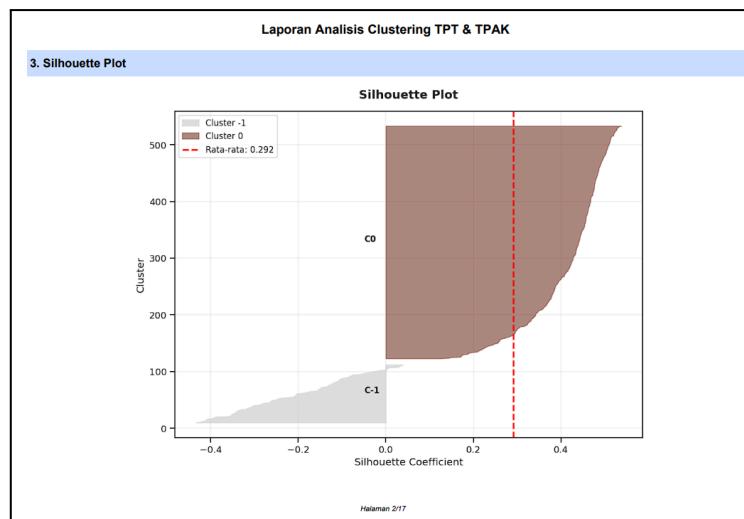
2. Hasil Metrik Evaluasi

Silhouette Score: 0.2915
- Interpretasi: Cukup

Davies-Bouldin Index: 6.5370
- Interpretasi: Kurang Baik

Waktu Komputasi: 0.26 detik
Jumlah Total Data: 514 wilayah
Jumlah Cluster Terbentuk: 2

Halaman 1/17



b. Tombol " Hapus Hasil Sekarang"

Tombol yang digunakan untuk membersihkan semua hasil analisis dari layar.

Simpan Laporan PDF **Hapus Hasil Sekarang**