



OPTIMALISASI PENGGEROMBOLAN PROVINSI BERDASARKAN TUJUAN PEMBANGUNAN BERKELANJUTAN DENGAN PENGGEROMBOLAN BERHIERARKI, *K-MEDOIDS*, DAN DBSCAN

©Hak cipta milik IPB University

FERISTA WAHYU SAPUTRI



**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Optimalisasi Penggerombolan Provinsi Berdasarkan Tujuan Pembangunan Berkelanjutan dengan Penggerombolan Berhierarki, *K-Medoids*, dan DBSCAN” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini. Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2024

Ferista Wahyu Saputri
G1401201008

ABSTRAK

FERISTA WAHYU SAPUTRI. Optimalisasi Penggerombolan Provinsi Berdasarkan Tujuan Pembangunan Berkelanjutan dengan Penggerombolan Berhierarki, *K-Medoids*, dan DBSCAN. Dibimbing oleh AGUS MOHAMAD SOLEH dan AAM ALAMUDI.

Analisis gerombol merupakan teknik statistika untuk mengidentifikasi gerombol bermakna dari objek data multivariat berdasarkan ukuran kemiripan. Selain memilih ukuran kemiripan, ketepatan dalam menerapkan metode pembentukan gerombol juga menjadi hal fundamental. Karakteristik data dapat memengaruhi performa analisis gerombol. Data multidimensional dengan jumlah dimensi yang besar memiliki tantangan adanya *curse of dimensionality*. Salah satu cara mengatasi masalah tersebut adalah dengan melakukan reduksi dimensi. Penelitian ini mengevaluasi performa metode reduksi dimensi analisis komponen utama dan t-SNE, serta membandingkan tiga metode penggerombolan yaitu metode hierarki, metode partisi dengan algoritma *K-Medoids*, dan metode berbasis kepekatan dengan algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) pada data multidimensi, studi kasus penggerombolan provinsi berdasarkan tujuan pembangunan berkelanjutan (TPB). Evaluasi kinerja hasil penggerombolan didasarkan pada tiga indeks validitas. Hasil penelitian menunjukkan bahwa kombinasi t-SNE dengan algoritma hierarki pautan tunggal atau dengan DBSCAN menghasilkan gerombol paling optimal dengan nilai *Silhouette* tertinggi (0,56), indeks *Davies-Bouldin* terendah (0,47), dan indeks *Calinski-Harabasz* tertinggi (75,04) dibandingkan hasil penggerombolan algoritma lainnya. Penggerombolan optimal tersebut menghasilkan tujuh gerombol provinsi dengan karakteristik pencapaian TPB yang berbeda. Terdapat dua gerombol yang menunjukkan pencapaian TPB rendah, sehingga memerlukan perhatian ekstra sesuai dengan karakteristiknya masing-masing. Gerombol-gerombol tersebut terdiri atas provinsi di wilayah timur Indonesia.

Kata kunci: analisis gerombol, analisis komponen utama, DBSCAN, t-SNE, tujuan pembangunan berkelanjutan.



ABSTRACT

FERISTA WAHYU SAPUTRI. Optimization of Provincial Clustering Based on Sustainable Development Goals with Hierarchical Clustering, K-Medoids, and DBSCAN. Supervised by AGUS MOHAMAD SOLEH and AAM ALAMUDI.

Cluster analysis is a statistical technique used to identify meaningful clusters from multivariate data objects based on similarity measures. In addition to choosing the right similarity measures, the accuracy in applying clustering methods is also fundamental. The characteristics of the data can influence the performance of cluster analysis. Multidimensional data with many dimensions face the challenge known as the curse of dimensionality. Dimensionality reduction helps address this issue. This study evaluates the performance of two dimensionality reduction methods, principal component analysis and t-SNE, and compares three clustering methods: hierarchical clustering, K-Medoids, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The study focuses on clustering provinces based on sustainable development goals (SDGs). Clustering performance is evaluated using three validity indices. Results indicate that combining t-SNE with the single-linkage hierarchical algorithm or DBSCAN produces the most optimal clusters. These methods achieve the highest Silhouette score (0.56), the lowest Davies-Bouldin index (0.47), and the highest Calinski-Harabasz index (75.04) compared to other algorithms. The optimal clustering results in seven distinct clusters of provinces with different SDGs achievement characteristics. Two clusters, comprising provinces in eastern Indonesia, show low SDGs achievement and require extra attention based on their specific characteristics.

Keywords: cluster analysis, DBSCAN, principal component analysis, sustainable development goals, t-SNE.

@Hak Cipta milik IPB University



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



OPTIMALISASI PENGGEROMBOLAN PROVINSI BERDASARKAN TUJUAN PEMBANGUNAN BERKELANJUTAN DENGAN PENGGEROMBOLAN BERHIERARKI, *K-MEDOIDS*, DAN DBSCAN

© Hak cipta milik IPB University

FERISTA WAHYU SAPUTRI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

IPB University

**Tim Penguji pada Ujian Skripsi:
1 Akbar Rizki, S.Stat., M.Si.**



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Optimalisasi Penggerombolan Provinsi Berdasarkan Tujuan
Pembangunan Berkelanjutan dengan Penggerombolan Berhierarki,
K-Medoids, dan DBSCAN

Nama : Ferista Wahyu Saputri
NIM : G1401201008

Disetujui oleh

Pembimbing 1:

Dr. Agus Mohamad Soleh S.Si., M.T.

Pembimbing 2:

Ir. Aam Alamudi, M.Si.

Diketahui oleh

Ketua Departemen Statistika:

Dr. Bagus Sartono, M.Si.
197804112005011002

Tanggal Ujian:
11 Juni 2024

Tanggal Lulus:

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanahu wa taala atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan karya ilmiah yang berjudul “Optimalisasi Penggerombolan Provinsi Berdasarkan Tujuan Pembangunan Berkelanjutan dengan Penggerombolan Berhierarki, *K-Medoids*, dan DBSCAN”. Terima kasih penulis ucapkan kepada Bapak Dr. Agus Mohamad Soleh S.Si., M.T. dan Bapak Ir. Aam Alamudi, M.Si. yang telah membimbing penulis selama ini. Ucapan terima kasih juga disampaikan kepada dosen penguji Bu Akbar Rizki, S.Stat., M.Si. yang telah memberikan masukan dan saran demi perbaikan penelitian ini. Kemudian terima kasih juga disampaikan kepada ibu, bapak, serta seluruh keluarga yang telah memberikan dukungan, doa, dan kasih sayangnya. Semoga karya ilmiah ini dapat memberikan wawasan dan kontribusi yang positif bagi pembaca dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2024

Ferista Wahyu Saputri



DAFTAR ISI

DAFTAR ISI	ix
DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	2
2.1 Tujuan Pembangunan Berkelanjutan	2
2.2 Analisis Komponen Utama (<i>Principal Component Analysis/PCA</i>)	3
2.3 <i>t-Distributed Stochastic Neighbor Embedding</i> (t-SNE)	4
2.4 Analisis Gerombol	6
2.5 Ukuran Kemiripan	6
2.6 Metode Penggerombolan	7
2.7 Kinerja Penggerombolan	9
III METODE	11
3.1 Data	11
3.2 Prosedur Analisis Data	13
IV HASIL DAN PEMBAHASAN	15
4.1 Eksplorasi Data	15
4.2 Reduksi Dimensi	16
4.3 Analisis Gerombol	19
4.4 Perbandingan Hasil Gerombol Optimal Algoritma Hierarki, <i>K-Medoids</i> , dan DBSCAN	22
4.5 Karakteristik Hasil Penggerombolan Terbaik	22
V SIMPULAN	30
DAFTAR PUSTAKA	31
LAMPIRAN	33
RIWAYAT HIDUP	36



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Tujuan pembangunan berkelanjutan	3
2	Peubah yang digunakan dalam penelitian	11
3	Hasil perhitungan nilai <i>eigen</i>	17
4	Nilai validitas hasil penggerombolan Hierarki pautan tunggal	19
5	Nilai validitas hasil penggerombolan Hierarki pautan lengkap	19
6	Nilai validitas hasil penggerombolan Hierarki pautan rata-rata	19
7	Hasil penggerombolan optimal metode Hierarki <i>agglomerative</i>	20
8	Nilai validitas hasil penggerombolan <i>K-Medoids</i>	21
9	Hasil penggerombolan optimal algoritma <i>K-Medoids</i>	21
10	Evaluasi hasil penggerombolan DBSCAN untuk <i>epsilon</i> optimal	21
11	Evaluasi hasil penggerombolan optimal setiap algoritma	22
12	Karakteristik peubah hasil penggerombolan terbaik	25

DAFTAR GAMBAR

1	Diagram alir penelitian	14
2	<i>Boxplot</i> semua peubah (X1 – X48) yang sudah dibakukan	15
3	Matriks korelasi <i>Pearson</i> antar pasangan peubah	16
4	<i>Scree plot</i> hasil PCA	17
5	<i>Scatter plot</i> hasil reduksi dimensi t-SNE	18
6	Dendogram hasil Hierarki <i>agglomerative</i> pautan tunggal	20
7	Provinsi Indonesia berdasarkan penggerombolan indikator SDGs	23
8	<i>Scatter plot</i> hasil penggerombolan optimal data reduksi t-SNE	23
9	Pencapaian target indikator SDGs setiap gerombol	23
10	<i>Radar Chart</i> peubah numerik hasil penggerombolan terbaik pada (a) Gerombol 1, (b) Gerombol 2, (c) Gerombol 3, (d) Gerombol 4, (e) Gerombol 5, (f) Gerombol 6, (g) Gerombol 7, (h) gabungan	24

DAFTAR LAMPIRAN

1	Statistik deskriptif peubah-peubah numerik pada penelitian (X1-X48)	34
---	---	----