

FORTEX 6.0 Data Science

Analisis Klaster Danau Berdasarkan Kualitas Air dan Karakteristik DAS

PENDEKATAN DENSITY-BASED CLUSTERING
MENGGUNAKAN ALGORITMA HDBSCAN DENGAN
EVALUASI SILHOUETTE DAN DBI

Data Entrying

Dhifya Fetryan, Teddy Agustinus, Verlene Angelica
Universitas Tarumanagara





Latar Belakang

Danau adalah ekosistem vital. Kesehatannya ditentukan oleh kualitas air dan interaksi dengan karakteristik DAS (tutupan lahan, aktivitas manusia). Data kualitas air dan DAS bersifat multidimensi dan kompleks. Sifat non-linier data ini menyulitkan analisis dan pemetaan pola dengan metode konvensional.



METODE CLUSTERING

Clustering adalah solusi untuk mengungkap pola tersembunyi. Pengelompokan ini penting untuk identifikasi variabilitas, prioritas konservasi, dan perumusan strategi pengelolaan danau yang spesifik.



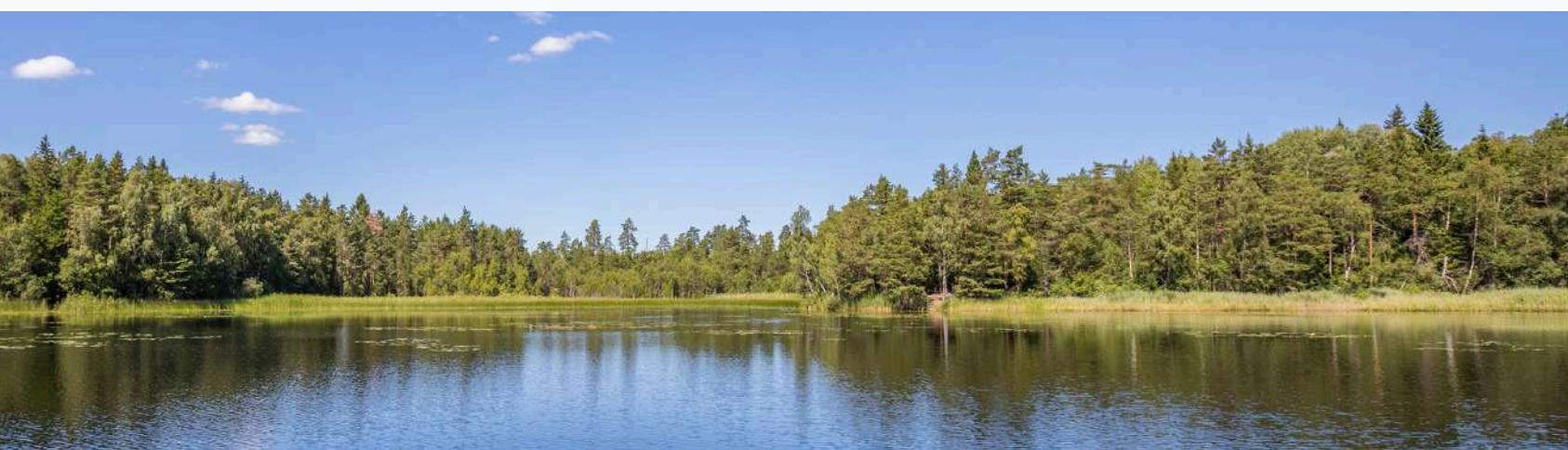


Rumusan Masalah

- Bagaimana cara mengelompokkan danau berdasarkan kualitas air dan karakteristik DAS secara optimal?
- Bagaimana performa algoritma HDBSCAN dalam mengidentifikasi pola klaster pada data danau?
- Bagaimana hasil evaluasi kualitas klaster menggunakan metrik Silhouette Coefficient dan Davies-Bouldin Index (DBI)?

Tujuan

- Mengelompokkan danau berdasarkan kemiripan kualitas air dan karakteristik DAS menggunakan algoritma HDBSCAN.
- Mengevaluasi hasil klustering dengan metrik Silhouette dan DBI untuk menilai kualitas hasil pengelompokan.



Manfaat



Memberikan informasi berbasis data yang dapat mendukung perencanaan strategi pengelolaan danau.



Menjadi dasar rekomendasi kebijakan terkait konservasi dan rehabilitasi danau sesuai dengan karakteristik kelompoknya.



Landasan Teori



Clustering

teknik unsupervised learning yang bertujuan mengelompokkan data berdasarkan kemiripan fitur tanpa label.

Density-Based Clustering

membentuk klaster berdasarkan kepadatan titik data. Titik dengan kepadatan tinggi dikelompokkan bersama, sedangkan titik yang terisolasi dianggap sebagai noise.

Davies-Bouldin Index

mengukur jarak rata-rata antara klaster yang berbeda relatif terhadap ukuran klaster (nilai ideal mendekati 0).

Silhouette Coefficient

mengukur seberapa mirip suatu objek dengan klasternya sendiri dibandingkan dengan klaster lain (nilai ideal mendekati 1).

HDBSCAN

pengembangan dari DBSCAN yang mampu mendeteksi jumlah klaster secara otomatis, bekerja lebih baik pada data kompleks, dan memiliki ketahanan tinggi terhadap noise.



Metode Penelitian



Pengumpulan Data

Mengambil data kualitas air dan karakteristik DAS dari beberapa danau.



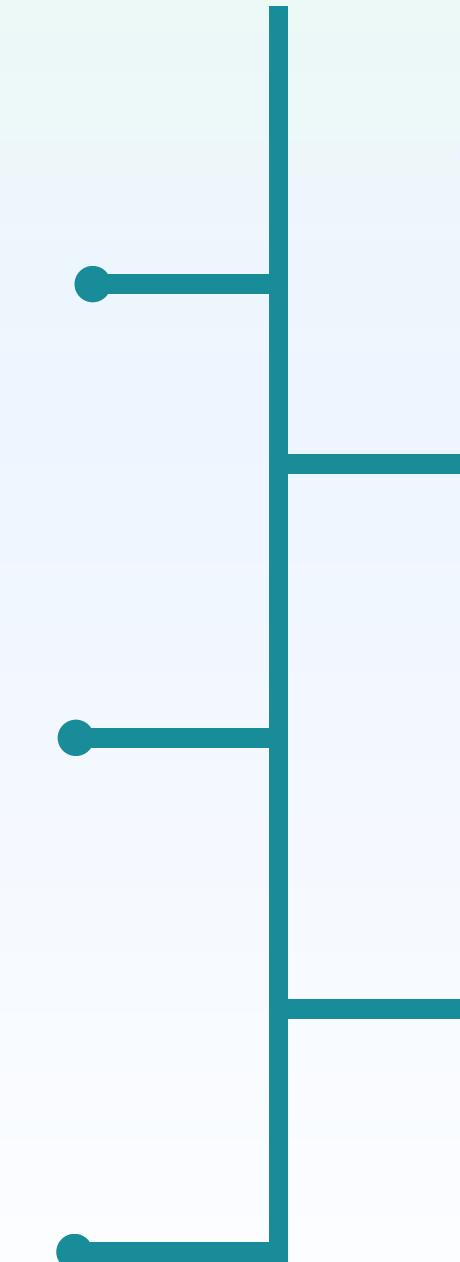
Reduksi Dimensi

Menggunakan PCA untuk memvisualisasikan hasil klastering dalam 2 dimensi.



Evaluasi

Menggunakan Silhouette Coefficient dan Davies-Bouldin Index untuk menilai kualitas hasil klastering.



Pra-pemrosesan

- Pembersihan data dari nilai kosong dan anomali.
- Normalisasi data agar semua variabel memiliki skala yang sama.

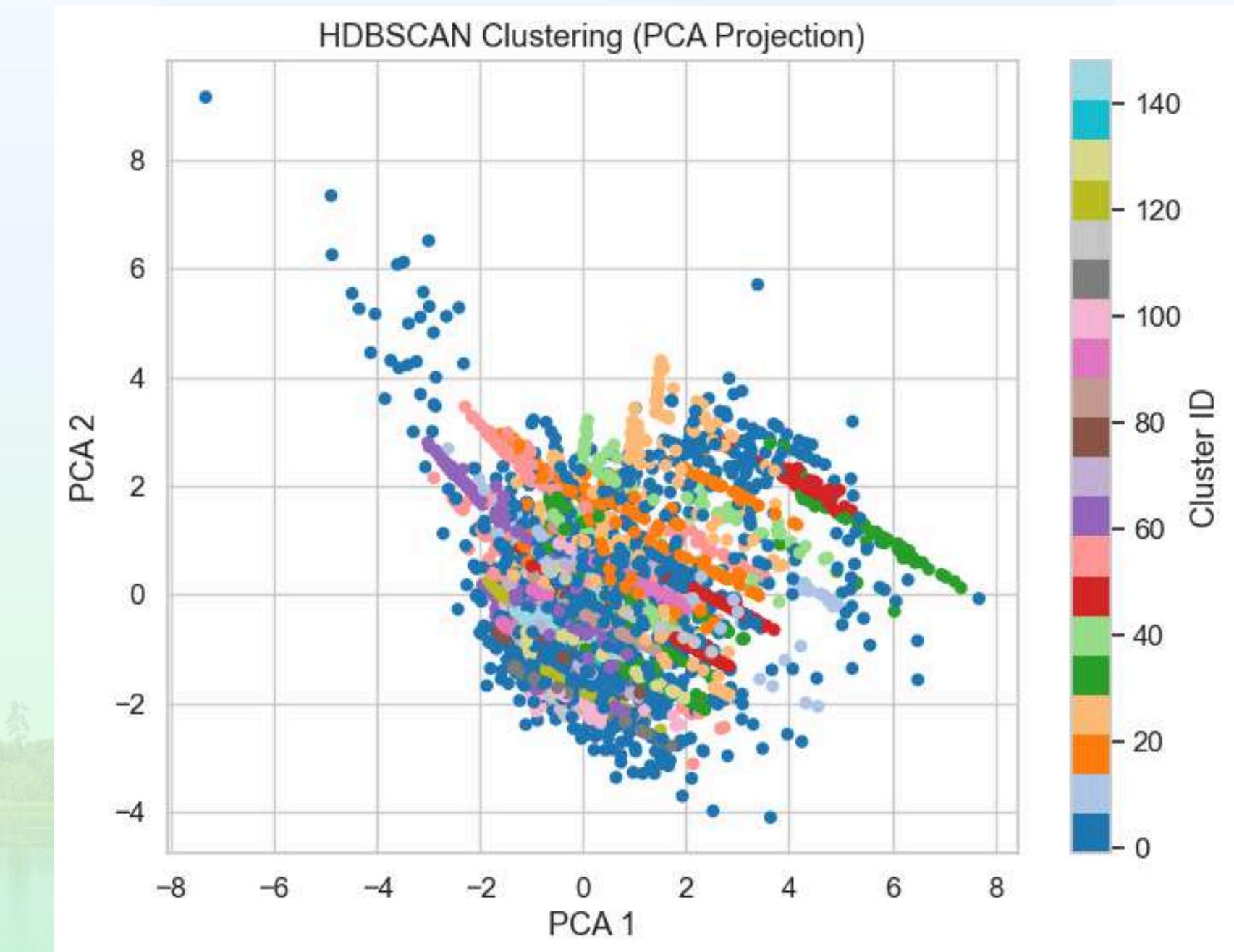
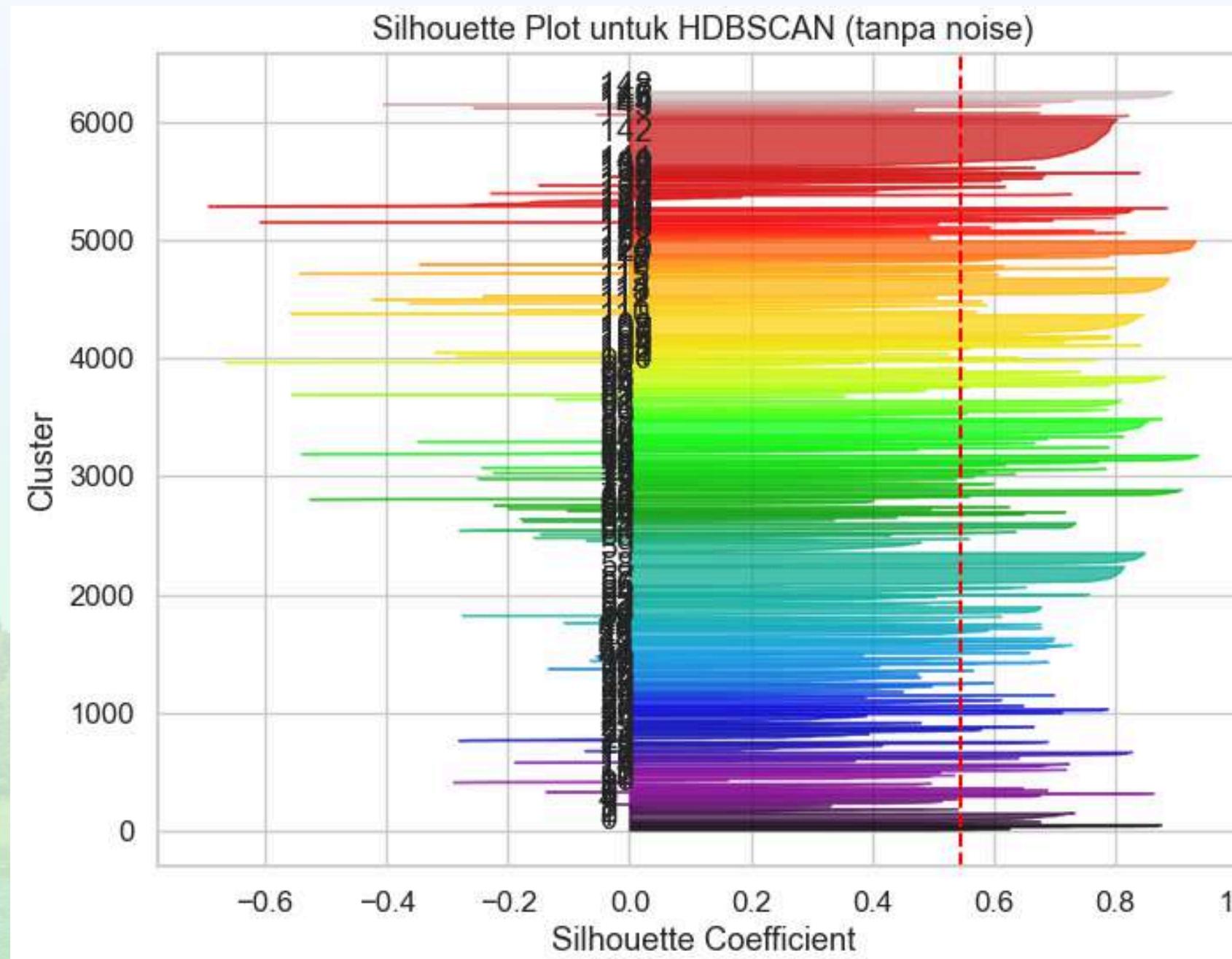


Clustering

Menerapkan algoritma HDBSCAN untuk membentuk kelompok danau berdasarkan kemiripan karakteristik.

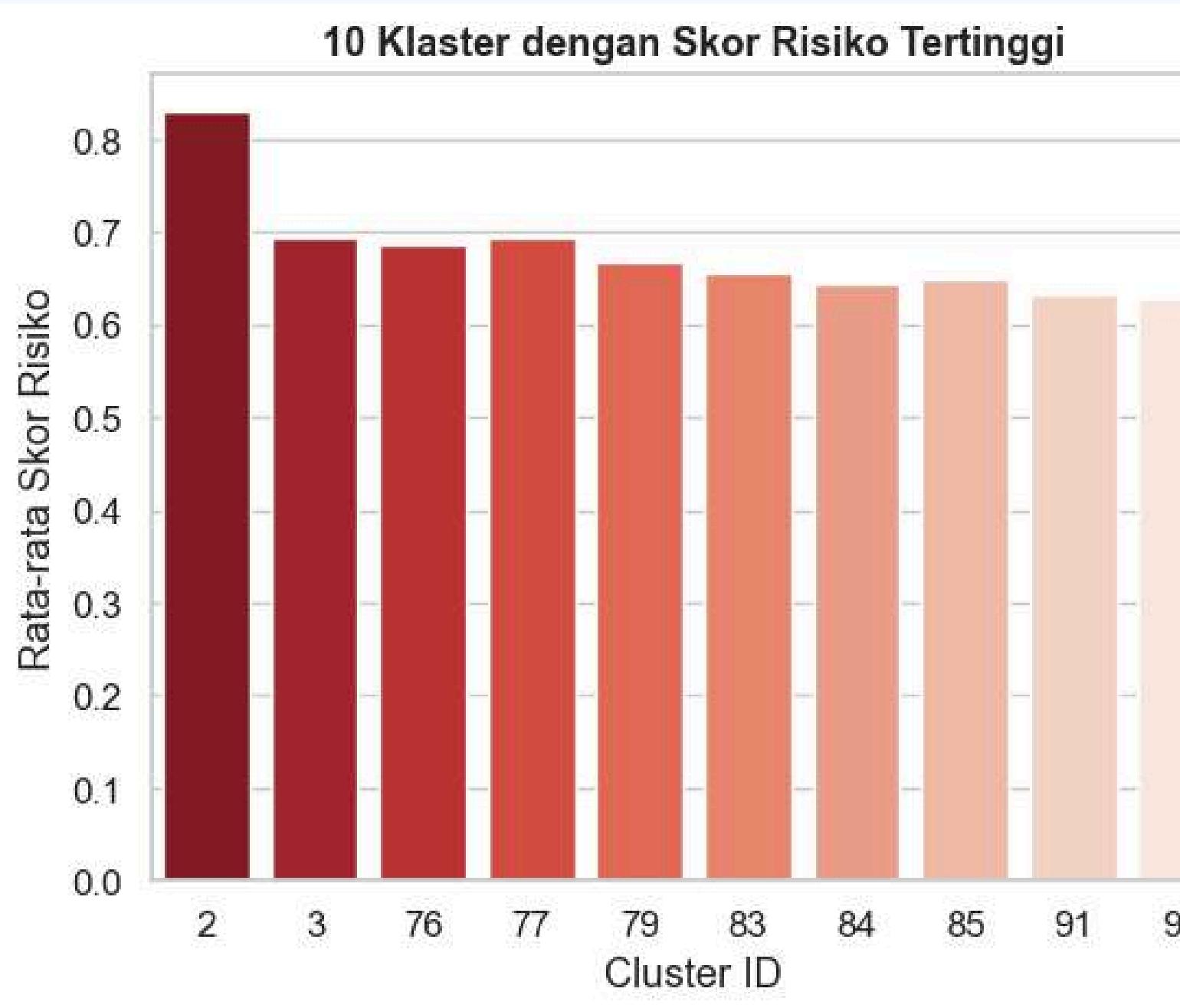


PEMBAHASAN





PEMBAHASAN



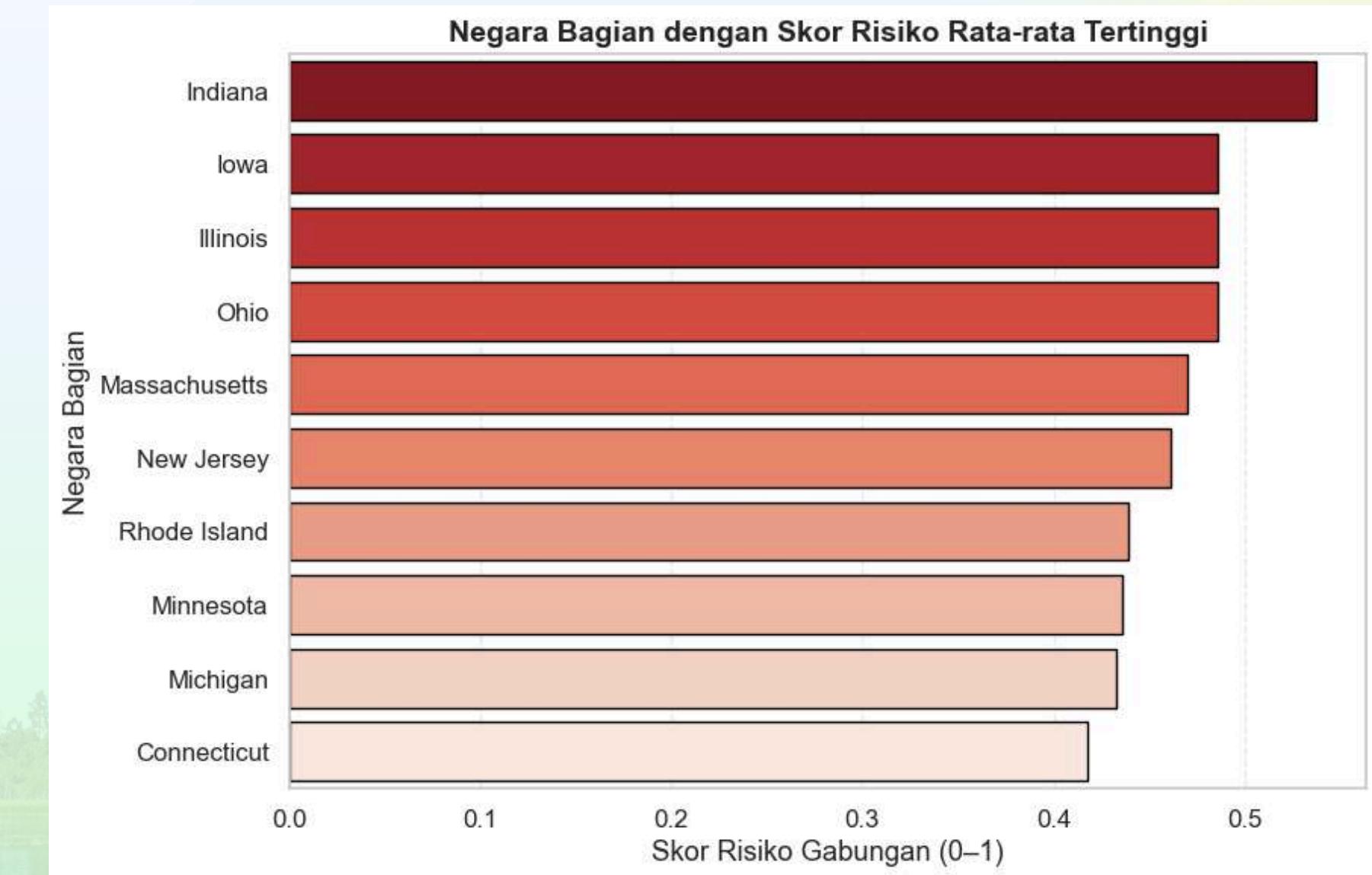
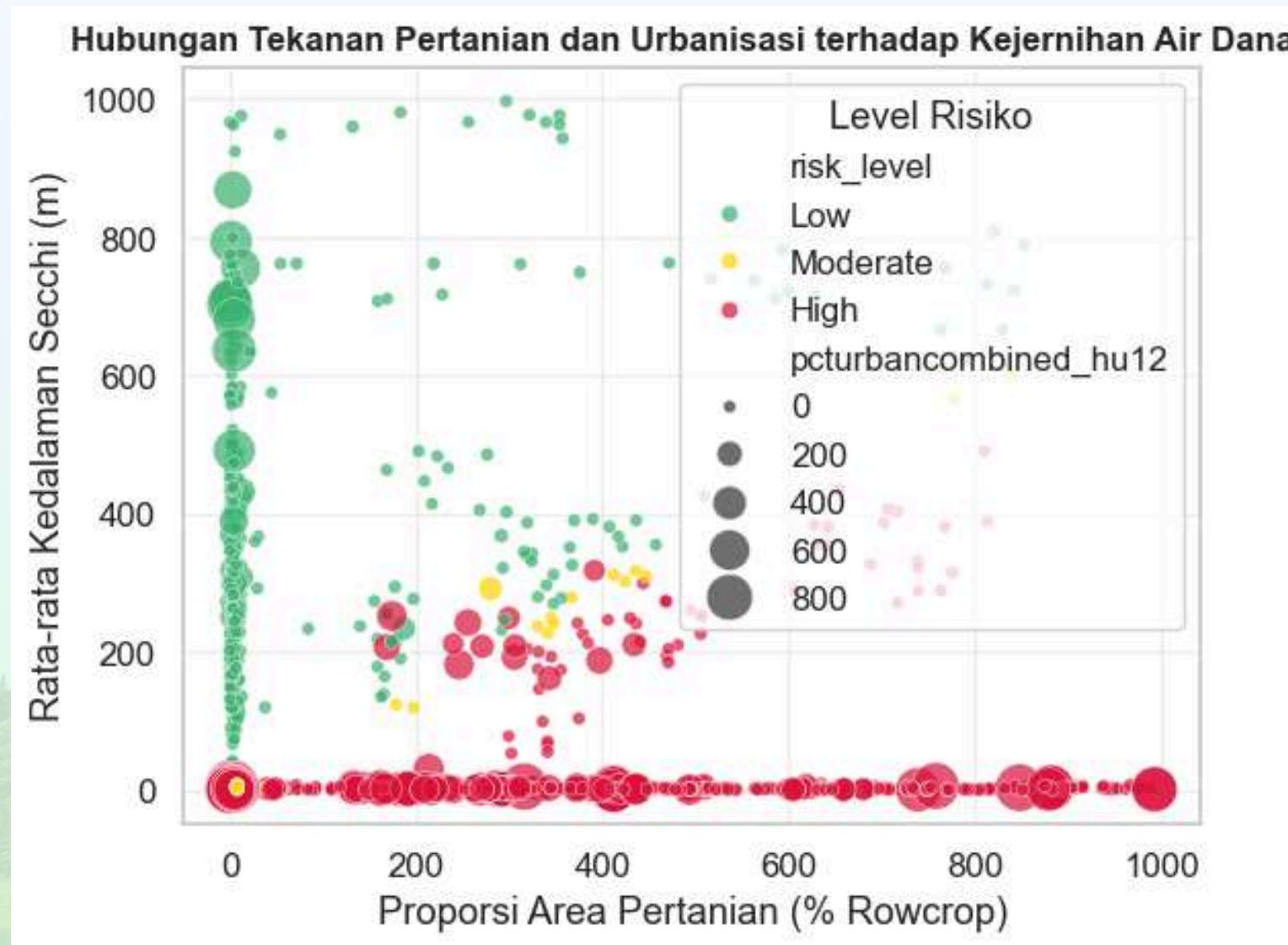
**Profil 15 Klaster Danau Paling Berisiko
Berdasarkan Faktor Tekanan Antropogenik**

		Skor Risiko (Gabungan)	Kekeruhan Air (Invers Secchi)	Tekanan Pertanian (%)	Tekanan Urban (%)
ID Klaster	1	1.00	0.17	0.82	0.96
	2	0.38	0.22	0.00	1.00
	3	0.38	0.01	1.00	0.00
	4	0.34	0.11	0.94	0.01
	5	0.26	0.00	0.85	0.00
	6	0.20	0.10	0.78	0.00
	7	0.17	0.52	0.73	0.00
	8	0.14	0.80	0.70	0.01
	9	0.09	1.00	0.64	0.00
	10	0.06	0.10	0.61	0.00
	11	0.06	0.75	0.60	0.01
	12	0.06	0.28	0.41	0.19
	13	0.03	0.23	0.57	0.00
	14	0.01	0.62	0.53	0.00
	15	0.00	0.21	0.52	0.01

Variabel Faktor Risiko



PEMBAHASAN





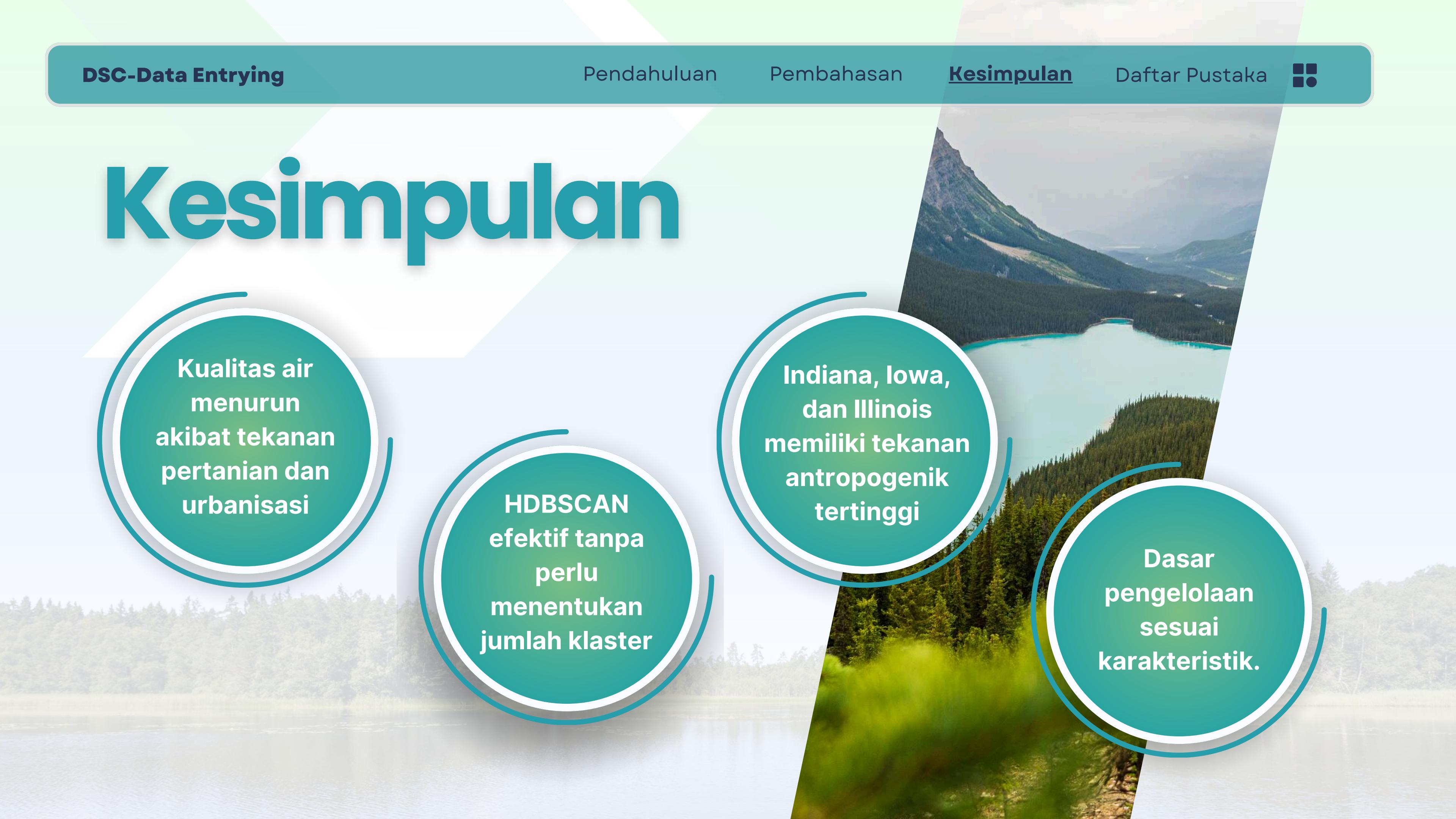
Kesimpulan

Kualitas air menurun akibat tekanan pertanian dan urbanisasi

HDBSCAN efektif tanpa perlu menentukan jumlah klaster

Indiana, Iowa, dan Illinois memiliki tekanan antropogenik tertinggi

Dasar pengelolaan sesuai karakteristik.





Rekomendasi

Integrasi Data Lingkungan

Gabungkan hasil klaster dengan data tata guna lahan, curah hujan, dan aktivitas manusia.

Strategi Tiap Klaster

Susun prioritas konservasi, rehabilitasi, atau monitoring sesuai kelompok danau.

Analisis Noise

Teliti danau noise untuk memahami karakteristik unik dan strategi pengelolaan khusus.

Pembaruan Berkala

Lakukan analisis rutin untuk memantau perubahan karakteristik danau.



Validasi Algoritma

Bandingkan dengan algoritma lain untuk meningkatkan akurasi hasil.



Daftar Pustaka

Liu P, Yuan H, Ning Y, Chakraborty B, Liu N, Peres MA. A modified and weighted Gower distance-based clustering analysis for mixed type data: a simulation and empirical analyses. *BMC Medical Research Methodology*. 2024 Dec 18;24(1):305.

Costa E, Papatsouma I, Markos A. Benchmarking distance-based partitioning methods for mixed-type data. *Advances in Data Analysis and Classification*. 2023 Sep;17(3):701-24.

Szepannek G, Aschenbruck R, Wilhelm A. Clustering large mixed-type data with ordinal variables. *Advances in Data Analysis and Classification*. 2024 May 27:1-9.

Kachigunda B, Mengersen K, Perera DI, Coupland GT, van der Merwe J, McKirdy S. Use of mixed-type data clustering algorithm for characterizing temporal and spatial distribution of biosecurity border detections of terrestrial non-indigenous species. *Plos one*. 2022 Aug 9;17(8):e0272413.

Ros F, Riad R, Guillaume S. PDBI: A partitioning Davies-Bouldin index for clustering evaluation. *Neurocomputing*. 2023 Apr 1;528:178-99.

Batool F, Hennig C. Clustering with the average silhouette width. *Computational Statistics & Data Analysis*. 2021 Jun 1;158:107190.

Wani AA. Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*. 2024 Aug 29;10:e2286.



Daftar Pustaka

Shutaywi M, Kachouie NN. Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*. 2021 Jun 16;23(6):759.

Lee S, Kim J, Hwang J, Lee E, Lee KJ, Oh J, Park J, Heo TY. Clustering of time series water quality data using dynamic time warping: A case study from the Bukhan River water quality monitoring network. *Water*. 2020 Aug 27;12(9):2411.

Qian J, Liu H, Qian L, Bauer J, Xue X, Yu G, He Q, Zhou Q, Bi Y, Norra S. Water quality monitoring and assessment based on cruise monitoring, remote sensing, and deep learning: A case study of Qingcaosha Reservoir. *Frontiers in Environmental Science*. 2022 Oct 11;10:979133.

Stewart, Geoffrey, and Mahmood Al-Khassaweneh. "An implementation of the HDBSCAN* clustering algorithm." *Applied Sciences* 12.5 (2022): 2405.

Melvin, Ryan L., et al. "Visualizing correlated motion with HDBSCAN clustering." *Protein Science* 27.1 (2018): 62-75.

Strobl, Michael, et al. "Model-based clustering with `hdbscan`." *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Cham: Springer International Publishing, 2020.



FORTEX 6.0 Data Science

TERIMA KASIH

Data Entrying