

Yossy Candra
Andrea Tri Rian Dani

MENGUNGKAP POLA TERSEMBUNYI MENGUNAKAN ANALISIS KLASSTER DIR



Mengungkap Pola TERSEMBUNYI

Menggunakan Analisis Klaster di R

Yossy Candra
Andrea Tri Rian Dani



PENERBIT KBM INDONESIA

Adalah penerbit dengan misi memudahkan proses penerbitan buku buku penulis di tanah air Indonesia. Serta menjadi media sharing proses penerbitan buku.

MENGUNGKAP POLA TERSEMBUNYI MENGGUNAKAN ANALISIS KLAS TER DI R

*Copyright @2025 By Yossy Candra dan Andrea Tri Rian Dani
All right reserved*

Penulis

Yossy Candra
Andrea Tri Rian Dani

Desain Sampul

Aswan Kreatif

Tata Letak

Sofitahm

Editor

Dr. Muhamad Husein Maruapey, Drs., M.Sc
Background isi buku di ambil dari <https://www.freepik.com/>

Official

Depok, Sleman-Jogjakarta (Kantor)
Penerbit Karya Bakti Makmur (KBM) Indonesia
Anggota IKAPI/No. IKAPI 279/JTI/2021
081357517526 (Tlpn/WA)

Website

<https://penerbitkbm.com>
www.penerbitbukumurah.com

Email

naskah@penerbitkbm.com

Distributor

<https://penerbitkbm.com/toko-buku/>

Youtube

Penerbit KBM Sastrabook

Instagram

@penerbit.kbmindonesia
@penerbitbukujogja

ISBN: 978-634-202-848-3

Cetakan ke-1, Oktober 2025
17,6 x 25 cm, viii + 108 halaman

Isi buku diluar tanggungjawab penerbit
Hak cipta merek KBM Indonesia sudah terdaftar di DJKI-Kemenkumham
dan isi buku dilindungi undang-undang.
Dilarang keras menerjemahkan, memfotokopi, atau
memperbanyak sebagian atau seluruh isi buku ini
tanpa seizin penerbit karena beresiko sengketa hukum

Sanksi Pelanggaran Pasal 113
Undang-Undang No. 28 Tahun 2014 Tentang Hak Cipta

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp 100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp 500.000.000,00 (lima ratus juta rupiah).
3. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp 1.000.000.000,00 (satu miliar rupiah).
4. Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp 4.000.000.000,00 (empat miliar rupiah).

KATA PENGANTAR

Puji syukur penyusun panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan kesempatan sehingga buku yang berjudul ***“Mengungkap Pola Tersembunyi Menggunakan Analisis Klaster di R”*** ini dapat diselesaikan dengan baik. Buku ini disusun dengan tujuan untuk mempermudah pemahaman tentang konsep-konsep analisis klaster dan optimasi, serta memberikan panduan praktis untuk mengimplementasi-kannya menggunakan *R*. Analisis klaster merupakan teknik yang digunakan untuk mengelompokkan data berdasarkan kesamaan pola, sementara *Principal Component Analysis* (PCA) digunakan untuk mengoptimalkan klaster dengan mereduksi dimensi data.

Tujuan penyusunan buku ini adalah untuk memberikan kemudahan dalam mempelajari dan mengimplementasikan teknik-teknik tersebut dengan pendekatan yang sistematis dan aplikatif. Buku ini mencakup berbagai teknik analisis klaster dan optimasi PCA yang dapat diterapkan dalam berbagai jenis permasalahan dan dataset.

Pada kesempatan ini, penyusun mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, sehingga buku ini dapat terwujud. Dengan segala kerendahan hati, penyusun persembahkan karya ini, semoga buku ini memberikan manfaat dan dapat membantu dalam memperdalam pengetahuan mengenai analisis klaster dan optimasi data. Penyusun menyadari bahwa masih banyak kekurangan dalam buku ini, dan kritik serta saran yang membangun sangat diharapkan untuk perbaikan lebih lanjut. Kritik dan saran dapat dikirimkan ke email yondra2022@gmail.com. Terimakasih.

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI.....	iii
DAFTAR TABEL.....	v
DAFTAR GAMBAR.....	vii
BAB 1 KNOWLEDGE DISCOVERY IN DATA MINING.....	1
1.1 Definsi dan Langkah-Langkah Knowledge Discovery in Data Mining.....	1
1.2 Posisi <i>Data Mining</i> dalam Berbagai Disiplin Ilmu.....	4
1.3 Pekerjaan dalam <i>Data Mining</i>	5
BAB 2 KONSEP ANALISIS KLASTER.....	9
2.1 Pengantar Analisis Klaster	9
2.2 Analisis Klaster Menurut Para Ahli.....	13
2.3 Karakteristik Klaster yang Baik.....	15
2.4 Pengukuran Jarak Kemiripan.....	16
BAB 3 PENGELOMPOKAN HIERARKI.....	21
3.1 Konsep Pengelompokan Hierarki	21
3.2 Pendekatan dalam Pengelompokan Hierarki.....	22
3.3 Definisi, Persamaan, dan Tahapan Algoritma <i>Single Linkage</i>	24
3.4 Definisi, Persamaan, dan Tahapan Algoritma <i>Complete Linkage</i>	25
3.5 Definisi, Persamaan, dan Tahapan Algoritma <i>Average Linkage</i>	26
3.6 Definisi, Persamaan, dan Tahapan Algoritma <i>Centroid</i>	26
3.7 Definisi, Persamaan, dan Tahapan Algoritma <i>Ward</i>	27
3.8 Implementasi Pengelompokan Hierarki Menggunakan <i>Software R</i>	28
BAB 4 PENGELOMPOKAN NON HIERARKI	35
4.1 Konsep Pengelompokan Non-Hierarki.....	35
4.2 Definisi, Persamaan, dan Tahapan Algoritma K-Means	36

4.3	Definisi, Persamaan, dan Tahapan Algoritma <i>K-Medoids</i>	38
4.4	Definisi, Persamaan, dan Tahapan Algoritma <i>K-Modes</i>	40
4.5	Definisi, Persamaan, dan Tahapan Algoritma Fuzzy C-Means.....	41
4.6	Implementasi Pengelompokan Non-Hierarki Menggunakan <i>Software R</i>	43
BAB 5	UJI VALIDITAS KLASTER	57
5.1	Konsep dan Definisi Uji Validitas Kluster.....	57
5.2	Definisi, Persamaan, dan Implementasi <i>R</i> Koefisien Korelasi <i>Cophenetic</i>	58
5.3	Definisi, Persamaan, dan Implementasi <i>R</i> <i>Davies Bouldin Index</i> (DBI)	60
5.4	Definisi, Persamaan, dan Implementasi <i>R</i> <i>Silhouette</i> <i>Coefficient</i>	64
5.5	Definisi, Persamaan, dan Implementasi <i>R</i> <i>Dunn Index</i>	68
BAB 6	OPTIMASI KLASTER DENGAN <i>PRINCIPAL COMPONENT</i> <i>ANALYSIS</i> (PCA)	73
6.1	Konsep Optimasi Kluster	73
6.2	Penerapan Optimasi PCA dalam Pengelompokan Menggunakan Algoritma <i>Centroid</i> (Hierarki).....	75
DAFTAR PUSTAKA		103
PROFIL PENULIS		107

DAFTAR TABEL

Tabel 3.1 Data Kementerian Agama Delapan Variabel Penelitian.....	28
Tabel 5.1 Kategori Nilai SC	66
Tabel 6.1 Data Kementerian Agama Lima Belas Variabel Penelitian	82
Tabel 6.2 Nilai VIF pada Setiap Variabel.....	87

DAFTAR PUSTAKA

- Anderberg, M. (1973). *Cluster analysis for application*. Academic Press.
- Ansari, M. Saleh. 2011. *Analisis Statistik dengan Microsoft Excel dan SPSS*. Yogyakarta: Graha Ilmu.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis (5th ed.)*. Wiley.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From data mining to knowledge discovery in databases*. AI Magazine, 17(3), 37–54.
<https://doi.org/10.1609/aimag.v17i3.1230>
- García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-10247-4>
- Gujarati, D. (2009). *Dasar-dasar ekonometrika*. Erlangga.
- Hair, J. F., Black, W. C., & Babin, B. J. (2009). *Multivariate data analysis*. Prentice Hall.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Morgan Kaufmann.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction (2nd ed.)*. Springer.
- Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.). (2015). *Handbook of cluster analysis*. CRC Press.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice Hall.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. ACM Computing Surveys (CSUR), 31(3), 264–323.
<https://doi.org/10.1145/331499.331504>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R (2nd ed.)*. Springer.

- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. Pearson Prentice Hall.
- Jolliffe, I. T., & Cadima, J. (2016). *Principal component analysis: A review and recent developments*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Wiley.
- Mardi, Y., (2017), Data Mining: Klasifikasi Menggunakan Algoritma C4.5. Jurnal Edik Informatika, Vol. 2, No. 2, Hal. 213-219.
- Nielsen, F. (2016). *Introduction to HPC with MPI for data science*. Springer.
- Nugroho, H. (2008). *Statistika multivariat terapan*. UNIB Press.
- Pramesti, D. F., Furqon, M. T., & Dewi, C., (2017), Impelementasi Metode K-Medoids Clustering untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, 964X.Y. Vol. 1, No. 9, Hal. 723-732.
- Pramana, S., Yuniarto, B., Mariyah, S., Santoso, L., & Nooraeni, R., (2018), *Data Mining dengan R: Konsep serta Impelementasi*. Bogor: IN MEDIA.
- Prasetyo, E. (2012). *Data mining: Konsep dan aplikasi menggunakan MATLAB*. Andi Offset.
- Santosa, B., & Umam, A. (2018). *Data Mining dan Big Data Analytics*. Yogyakarta: Penebar Media Pustaka.
- Santoso, B. (2007). *Data mining: Teknik pemanfaatan data untuk keperluan bisnis*. Graha Ilmu.
- Suntoro, J. (2019). *Data mining: Algoritma dan implementasi dengan pemrograman PHP*. Elex Media Komputindo.
- Suyanto, Dr. (2019). *Data Mining untuk Klasifikasi dan Klasterisasi Data (Edisi Revisi)*. Informatika.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Harper Collins.
- Uyun, S., & Subanar. 2007. *Pengantar Statistika Matematika dan Analisis Data*. Yogyakarta: Graha Ilmu.

- Widarjono, A. (2015). *Analisis multivariat terapan dengan program SPSS*. UPP STIM YKPN.
- Yuan, C., & Yuang, H., (2019), Research on K-Value Selection Method of K-Means Clustering Algorithm. *Journal Multidisciplinary Scientific*, Vol. 2, No. 2, Hal. 226-235.

PROFIL PENULIS



Yossy Candra, merupakan Statistisi di Kementerian Agama Republik Indonesia. Menyelesaikan pendidikan sarjana pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA), Universitas Mulawarman. Saat ini bertugas dalam mendukung kegiatan pengumpulan, pengolahan, analisis, dan penyajian data statistik di lingkungan instansi. Peran tersebut berkontribusi pada penyediaan informasi yang akurat dan relevan guna menunjang pengambilan keputusan serta perumusan kebijakan berbasis data. Di awal karier, fokus utama diarahkan pada penelitian-penelitian di bidang *Data Mining*, khususnya pada Analisis Kluster, sebagai bagian dari pengembangan pemanfaatan data dalam mendukung kebijakan dan layanan publik yang berbasis bukti.



Andrea Tri Rian Dani, merupakan seorang dosen tetap di Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman. Lulus S1 dari Program Studi Statistika, FMIPA Universitas Mulawarman dengan gelar S.Stat., dan S2 dari Departemen Statistika, Fakultas Sains dan Analitika Data (FSAD), Institut Teknologi Sepuluh Nopember dengan gelar M.Stat. Penelitian-penelitian yang ditekuni adalah bidang regresi nonparametrik dan semiparametrik, khususnya pengembangan model estimator campuran. Beberapa penelitian telah berhasil dipublikasi, baik pada jurnal internasional bereputasi, prosiding internasional, jurnal nasional terindeks dan prosiding nasional. Saat ini mengampu mata kuliah di S1 Statistika Unmul, diantaranya Pengantar Statistika Matematika 1, Statistika Matematika 2, Pengantar Regresi

Nonparametrik, Analisis Data Kategorik, Matematika Keuangan, Pengantar Metode Bayesian, Modern Prediksi dan Machine Learning, Analisis Risiko, dan Teori Antrian