

# *Klasterisasi Makanan Khas Daerah Indonesia Menggunakan Algoritma Machine Learning Berdasarkan Kandungan Nutrisi*

Jumail Ilham Sahupala

Program Studi : Sistem Informasi, Universitas Darwan Ali

Email : [sahupalamjumaililham@gmail.com](mailto:sahupalamjumaililham@gmail.com)

---

**ABSTRACT**— Indonesia is renowned for its rich traditional culinary diversity, supported by an abundance of spices that serve as a global attraction. Although Indonesian food boasts delicious flavors, excessive consumption can have negative health impacts. In line with this issue, this research aims to analyze and cluster regional specialty foods based on their nutritional content. The dataset used includes information on calories, proteins, fat, and carbohydrate content per 100 grams for various food types. The K-Medoids algorithm was applied to group food types according to their nutritional content, specifically into low and high categories. The results successfully identified two clusters: Cluster 0 contains 1095 items and Cluster 1 contains 251 items. It is hoped that these clustering results can serve as a guide for individuals to independently manage their diet and nutritional intake, and contribute to future nutritional policy planning.

**Keywords**— Data mining, K-Medoids, Culinary, Nutrition

**ABSTRAK**— Indonesia dikenal kaya akan keanekaragaman kuliner tradisionalnya, yang didukung oleh melimpahnya rempah-rempah sebagai daya tarik global. Meskipun cita rasa makanan Indonesia sangat nikmat, konsumsi berlebihan dapat berdampak negatif bagi kesehatan. Sejalan dengan permasalahan ini, penelitian ini bertujuan untuk menganalisis dan mengelompokkan makanan khas daerah berdasarkan kandungan nutrisinya. Dataset yang digunakan mencakup informasi kandungan *calories*, *proteins*, *fat*, dan *carbohydrate* per 100 gram untuk berbagai jenis makanan. Algoritma K-Medoids diterapkan untuk mengelompokkan jenis makanan berdasarkan kandungan nutrisinya, yaitu klaster kategori rendah dan tinggi. Hasil penelitian ini berhasil mengidentifikasi dua klaster: Klaster 0 berisi 1095 item dan Klaster 1 berisi 251 item. Diharapkan hasil klasterisasi ini dapat menjadi panduan bagi individu dalam mengatur pola makan dan asupan nutrisi secara mandiri, serta memberikan kontribusi dalam perencanaan kebijakan gizi di masa mendatang.

**Kata kunci**— Data mining, K-Medoids, Kuliner, Nutrisi

---

## I. PENDAHULUAN

Indonesia, sering disebut sebagai negara kepulauan dan terletak di Asia Tenggara, dikenal luas akan kekayaan adat dan budayanya, serta pemandangan alamnya yang memukau dunia. Selain itu, kuliner tradisional yang lahir dari kreativitas masyarakat di berbagai daerah telah menjadi duta utama dalam memperkenalkan identitas Indonesia ke kancah internasional [1]. Popularitas kuliner Indonesia semakin terbukti dengan berbagai pengakuan global. Contohnya, nasi goreng dan rendang pernah dinobatkan sebagai salah satu makanan tradisional terenak di dunia oleh media internasional terkemuka, menunjukkan daya saing dan citra positif kuliner Indonesia di mata dunia [2]. Ini membuktikan bahwa citra kuliner dari Indonesia juga bisa bersaing dengan produk makanan orang luar negeri.

Karakteristik kuliner Indonesia merupakan refleksi dari perpaduan budaya dan adat istiadat yang

kaya di setiap daerah. Penyebarannya dari Sabang hingga Merauke memperlihatkan keunikan cita rasa dan filosofi lokal. Setiap hidangan tidak hanya menawarkan lezatnya, tetapi juga memiliki makna mendalam serta posisi penting dalam kehidupan sosial dan budaya masyarakat setempat [3].

Salah satu ciri khas utama kuliner Indonesia adalah penggunaan rempah dan bumbu-bumbu khas daerah yang melimpah, menghasilkan cita rasa yang kompleks dan menarik. Kekayaan rempah ini bahkan memiliki jejak historis penting, di mana Indonesia pernah menjadi perebutan bangsa-bangsa besar di masa lalu. Rempah-rempah kala itu dianggap sebagai komoditas berharga karena kelangkaan pasokan yang membuatnya memiliki nilai ekonomis tinggi di pasar global [4].

Namun komposisi yang tinggi pada penggunaan bahan berupa rempah dan bumbu lainnya, patut menjadi perhatian. Meskipun pada dasarnya bertujuan untuk memperkaya cita rasa, jika

di konsumsi berlebihan dapat menimbulkan risiko kesehatan. Sebagai contoh, banyak jajanan pasar manis menggunakan gula dalam jumlah tinggi sebagai bahan utama, yang jika dikonsumsi berlebihan secara terus-menerus dapat memicu masalah kesehatan serius seperti obesitas dan diabetes. Demikian pula, beberapa hidangan utama yang kaya santan atau minyak dapat berkontribusi pada asupan lemak jenuh yang berlebih hingga melahirkan kolesterol [5].

Oleh karena itu, pengendalian asupan nutrisi menjadi krusial bagi setiap individu. Organisasi Kesehatan Dunia (WHO) menekankan bahwa setiap individu memiliki jumlah kebutuhan kalori harian yang bervariasi sesuai dengan usia individu, jenis kelamin, tingkat aktivitas rutin, dan kondisi kesehatan tubuhnya. Penting bagi kita untuk memahami dan memantau kebutuhan kalori serta nutrisi lainnya untuk menjaga kesehatan optimal [6]. Seiring dengan kemajuan teknologi, perhitungan dan pemantauan kebutuhan kalori serta asupan nutrisi harian kini dapat dibantu oleh berbagai aplikasi dan perangkat digital. Inovasi ini mempermudah individu dalam upaya mengelola konsumsi makanan agar tidak berlebihan maupun kekurangan nutrisi esensial [7].

Berdasarkan permasalahan yang di dapat, penelitian ini mempunyai tujuan untuk menganalisis dan mengelompokkan makanan khas daerah Indonesia berdasarkan kandungan nutrisinya. Penulis telah mengumpulkan dataset yang mencakup informasi kalori, lemak, karbohidrat, dan protein per 100 gram untuk berbagai hidangan. Data ini akan dimanfaatkan untuk mengidentifikasi klaster makanan dengan kandungan nutrisi tinggi dan rendah.

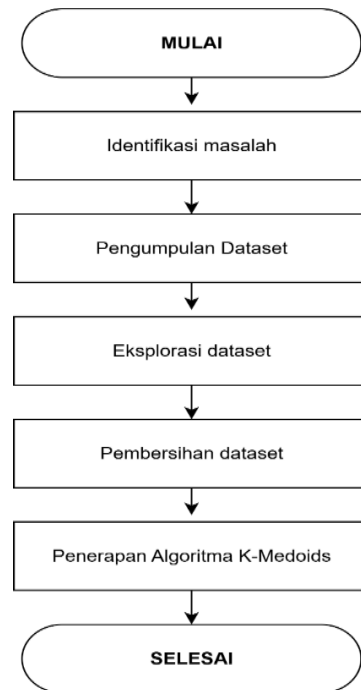
Untuk mendukung otomatisasi dan efisiensi dalam proses analisis data, penelitian ini mencoba menerapkan pendekatan keilmuan data mining sebagai salah satu cabang dari pembelajaran mesin (ML) [8]. Pendekatan ini mencakup tahapan pengumpulan data, analisis data, hingga identifikasi pola-pola unik dari informasi yang tersedia [9]. Beberapa metode yang dapat diterapkan dalam proses ini antara lain prediksi, estimasi, asosiasi, dan klustering. Hasil dari proses data mining ini diharapkan dapat memberikan manfaat yang sangat membantu dalam perencanaan dan pengambilan kebijakan di masa mendatang [10].

Dalam konteks klusterisasi dataset nutrisi kuliner ini, algoritma yang digunakan adalah K-

Medoids. Salah satu model pembelajaran mesin yang digunakan untuk mengelompokkan data berdasarkan nilai terdekat ke medoid [8][9][10], dengan harapan bisa mengelompokkan daftar makanan khas daerah berdasarkan kandungan nutrisinya.

## II. METODOLOGI PENELITIAN

Berdasarkan permasalahan yang sudah penulis jelaskan, penelitian ini akan dijelaskan menjadi beberapa tahapan agar lebih mudah bagi para pembaca nya untuk memahaminya.



Gambar 1. Tahapan penelitian

### A. Identifikasi Masalah

Proses ini perlu dilakukan di awal, bertujuan untuk mengetahui langkah-langkah yang ingin di lakukan berikutnya. Sesuai dengan tujuan penelitian ini yaitu untuk mengetahui klaster makanan berdasarkan kandungan nutrisi yang ada didalamnya, dan maka dari itu penulis mencoba mendapatkan data berupa ragam nutrisi yang di perlukan dalam tubuh. Di karena kan penelitian ini mengambil contoh kasus yang ada di Indonesia, maka dari itu data yang di kumpulkan lebih mengutamakan jenis makanan khas di tiap daerah Indonesia. Kemudian menerapkan algoritma K- Medoids untuk menemukan pola yang tersembunyi dalam data, pada akhirnya hasil perhitungan bisa digunakan untuk mengelompokkan makanan khas daerah Indonesia yang tinggi nutrisi dan yang rendah nutrisi.

B. Pengumpulan dataset

Dataset yang di gunakan merupakan dataset Sekunder, artinya dataset yang di kumpulkan oleh pihak ketiga atau peneliti lain kemudian membagi nya ke publik agar bisa digunakan kembali oleh peneliti yang lain [11]. Disini penulis menggunakan dataset yang ditemukan dari website <https://www.kaggle.com/datasets/anasfikrihanif/indonesian-food-and-drink-nutrition-dataset>.

C. Eksplorasi dataset

Dataset yang akan digunakan dalam perhitungan ini terdiri dari 1.346 record dengan beberapa atribut utama, yaitu: *id*, *calories*, *proteins*, *fat*, *carbohydrate*, *name*, dan *image*. Data ini merepresentasikan informasi nutrisi dari berbagai jenis kuliner di Indonesia yang dapat dianalisis untuk menemukan pola-pola tersembunyi. Dengan struktur data yang kaya dan terukur, dataset ini memungkinkan penerapan teknik data mining untuk menggali informasi yang relevan.

D. Pembersihan dataset

Pada tahap ini, penulis menentukan atribut-atribut yang relevan untuk dianalisis menggunakan algoritma K-Medoids. Tidak semua kolom dalam dataset dapat digunakan, karena algoritma ini hanya dapat memproses data numerik. Oleh karena itu, atribut *name* akan dijadikan sebagai identitas (*id*) dari masing-masing record, sedangkan atribut *calories*, *proteins*, *fat*, dan *carbohydrate* digunakan sebagai variabel numerik yang akan digunakan dalam proses klustering.

TABEL I  
 DATASET MAKANAN KHAS DAERAH

name (id)	calori es	protein s	fat	carbohydrate ate
Abon	280	9.2	28.4	0
Agar-agar	0	0	0.2	0
Andewi	25	1.6	0.2	5.3
Babat	113	17.6	4.2	0
Bacang	102.9	3.3	2.7	16.3

E. Penerapan algoritma K-Medoids

Algoritma K-Medoids merupakan salah satu bagian dari kluster algoritma *partitioning*, serupa dengan algoritma K-means. K-Medoids berfungsi membagi data ke dalam sejumlah kluster tanpa membentuk struktur hierarki antar kluster. K-Medoids memilih objek aktual dari kumpulan data sebagai representasi pusat kluster [12]. Memiliki keunggulan dalam ketahanan terhadap outlier dan

performa yang lebih stabil, terutama saat digunakan pada dataset berukuran kecil [13]. Pada dasarnya, algoritma K-Medoids beroperasi dengan secara iteratif memilih medoid terbaik dan menetapkan titik data ke medoid terdekat hingga kriteria yang telah ditentukan terpenuhi. Kriteria ini biasanya melibatkan minimisasi jumlah ketidakserupaan (jarak) antara titik data dan medoid yang dipilih [14]. Agar lebih mudah memahaminya, cara kerja algoritma K-Medoids di bagi menjadi beberapa tahapan yaitu:

- 1) Inisialisasi: Tahap awal ini melibatkan penentuan nilai *k*, yaitu jumlah kluster atau kluster yang ingin ditemukan. Setelah nilai *k* ditetapkan, sejumlah titik data dipilih secara acak dari dataset untuk menjadi medoid awal, yang akan berfungsi sebagai pusat sementara bagi kluster yang akan terbentuk.
- 2) Penugasan: setiap titik data non-medoid dihitung jaraknya terhadap setiap medoid yang telah dipilih. Titik data kemudian ditugaskan ke kluster yang medoidnya memiliki jarak terdekat. Dalam penelitian ini, metrik Jarak Euclidean (Euclidean Distance) digunakan untuk mengukur kedekatan jarak antar kedua titik data. Rumus Jarak Euclidean *y* didefinisikan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$d(x, y)$  melambangkan nilai jarak Euclidean antara 2 titik data *x* dan *y*, semakin kecil nilai  $d(x, y)$  maka semakin dekat kedua titik tersebut. *n* menunjukkan jumlah atribut yang tersedia untuk setiap titik data, Contoh nya: *calories*, *proteins*, *fat*, dan *carbohydrate* [8].

- 3) Pembaruan: Setelah semua titik data ditugaskan ke kluster masing-masing, algoritma mencari medoid baru. Medoid baru dipilih dari titik data yang ada dalam kluster tersebut yang memiliki total jarak terkecil ke semua titik data lain di kluster yang sama. Proses ini memastikan bahwa medoid yang dipilih benar-benar merupakan representasi paling sentral dari klasternya.
- 4) Penilaian: Pada setiap iterasi, algoritma menghitung total biaya (*cost*) klustering, yaitu jumlah total jarak antara setiap titik data dan medoid klasternya. Jika medoid baru yang terpilih menghasilkan total biaya yang lebih

rendah dibandingkan iterasi sebelumnya, maka medoid baru tersebut akan digunakan. Jika tidak ada perbaikan, atau jika biaya meningkat, medoid lama akan dipertahankan.

- Iterasi: Tahap 2, 3, dan 4 diulang secara berkesinambungan hingga tidak ada perubahan signifikan pada posisi medoid, atau hingga total biaya kluster yang stabil telah terbentuk.

Dalam penelitian ini, penulis menetapkan dua kluster utama agar proses analisis hasilnya bisa dilakukan lebih sederhana, yaitu: Kluster 1 yang merepresentasikan makanan dengan kandungan nutrisi rendah, dan Kluster 2 yang merepresentasikan makanan dengan kandungan nutrisi tinggi.

Untuk mempermudah proses perhitungan serta analisis data, penulis menggunakan Rapid Miner yang menyediakan berbagai alat untuk menganalisis data dan mendukung penerapan beberapa model AI khususnya model klustering K-Medoids, yang akan digunakan dalam penelitian ini [15].

#### F. Mengevaluasi model K-Medoids

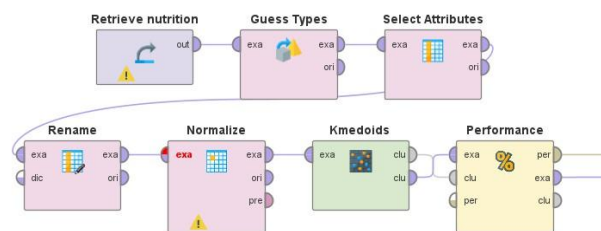
Kemudian mengukur performa model yang di bangun dengan metode *Average within centeroid distance* dan *Davies bouldin index*. Yang sangat umum di gunakan untuk mengukur ke akuratan hasil klasterisasi. *Average within centeroid distance* berkerja dengan menghitung rata-rata dari jarak tiap data ke titik centeroid nya masing masing. Kemudian *Davies bouldin index*, berkerja dengan mengukur rasio penyebaran data di dalam kluster (intra-kluster distance) dan jarak antar kluster (inter-kluster distance), intinya DBI mengukur seberapa mirip data dalam kluster yang sama dan seberapa jauh jarak kluster 1 dengan kluster lainnya [16].

Semakin kecil nilai *Average within centeroid distance* dan DBI maka model yang di bangun memiliki performa yang lebih baik.

### III. DESAIN, HASIL DAN PEMBAHASAN

#### A. Desain

Di bagian ini penulis akan memaparkan desain proses yang diimplementasikan menggunakan Rapid miner. Desain proses ini mencakup serangkaian operator yang digunakan untuk memproses dan menganalisis dataset.



Gambar 2. Desain proses pada Rapidminer

Berikut adalah penjelasan mengenai operator yang digunakan dalam desain proses:

- Operator Retrive Nutrition: berfungsi untuk memuat objek ke dalam proses, termasuk data tabular, model, atau objek RapidMiner lainnya. Pada penelitian ini, operator ini digunakan untuk memuat dataset nutrisi makanan yang telah disiapkan.
- Operator Guess Types: secara otomatis menetapkan tipe data dari tiap atribut (misalnya, numerik, nominal, atau tanggal), yang sangat membantu penulis menghemat waktu selama proses persiapan data serta meminimalkan kesalahan yang mungkin timbul akibat ketidaksesuaian format data.
- Operator Rename: memungkinkan perubahan nama kolom secara langsung dalam proses Rapid Miner, tanpa perlu memodifikasi sumber dataset. Ini dilakukan agar lebih mudah bagi penulis untuk memahami konteks dari tiap kolom/atribut.

old name	new name
name	nama
calories	kalori
carbohydrate	karbohidrat
fat	lemak
proteins	protein

Gambar 3. Perubahan pada tiap attrbut

- Operator Normalize: ini digunakan untuk menyeragamkan skala nilai dari setiap atribut agar berada dalam rentang yang sama. Normalisasi ini penting untuk meningkatkan performa algoritma klustering dan mengurangi dampak dari data *outlier*. Parameter yang diatur adalah
- Operator K-Medoids: merupakan bagian krusial dari penelitian ini. Operator ini bertanggung jawab untuk memanggil dan menjalankan

algoritma K-Medoids dalam lingkungan Rapid Miner, sehingga memungkinkan proses klusterisasi data berdasarkan kemiripan antar item.

k	2
max runs	20
max optimization steps	100

Gambar 4. Parameter yang K-Medoids gunakan

Parameter yang diatur adalah jumlah kluster yang ingin dibentuk yaitu  $k = 2$  sesuai dengan tujuan penelitian di awal. Kemudian max run 20 dan max optimization steps 100. di atur lebih tinggi agar hasil yang lebih presisi, namun bisa di lakukan percobaan lebih mendalam untuk menentukan nilai yang paling optimal.

- 6) Operator Performance: bertipe *Klaster Distance Performance* artinya operator ini memang digunakan khusus untuk mengevaluasi performa dari algoritma klustering.

**Performance (Cluster Distance Performance)**

main criterion: Avg. within centroid distance

main criterion only

normalize

maximize

Gambar 5. Parameter operator Performance

Parameter yang digunakan adalah opsi main criterion yang di matikan agar bisa menggunakan 2 metode penilaian sekaligus, yaitu: Avg. Within centeroid distance dan Davies Bouldin. Opsi normalize dinyalakan agar hasil penilaian di bagi dengan jumlah atribut sehingga hasil akhir bisa lebih spesifik dengan setiap atribut yang digunakan. Opsi maximize juga dinyalakan karena Rapid Miner secara default akan mengalikan hasil penilaian dengan

-1, otomatis hasil penilaian nya akan bernilai negatif.

**B. Hasil dan Pembahasan**

Penelitian tentang mengelompokkan data makanan tradisional Indonesia menggunakan algoritma K-Medoids ini menghasilkan dua kluster, yaitu: kluster makanan dengan nutrisi rendah dan

nutrisi tinggi. Artinya algoritma K-Medoids terbukti mampu membagi dataset besar menjadi sub-kluster yang homogen berdasarkan kemiripan data nya. Hasil dari perhitungan K-Medoids adalah sebagai berikut.

**Cluster Model**

Cluster 0: 1095 items  
 Cluster 1: 251 items  
 Total number of items: 1346

Gambar 6. Hasil perhitungan K-Medoids

Kluster 0 merepresentasikan makanan dengan kandungan nutrisi rendah, dan Kluster 1 merepresentasikan makanan dengan kandungan nutrisi tinggi. Dengan nilai medoid untuk masing- masing Kluster dijelaskan pada tabel berikut.

TABEL 2  
 Centeroid table

Attribute	Klaster_0	Klaster_1
karbohidrat	0.006	0.004
protein	0.040	0.175
lemak	0.025	0.423
kalori	0.055	0.481

Kemudian hasil dari evaluasi model nya adalah

```
PerformanceVector:
Avg. within centroid distance: 0.015
Avg. within centroid distance_cluster_0: 0.010
Avg. within centroid distance_cluster_1: 0.036
Davies Bouldin: 0.204
```

Gambar 7. Performance vektor

Klusterisasi ini memberikan gambaran yang jelas mengenai distribusi makanan tradisional berdasarkan profil nutrisinya, sesuai dengan penetapan di metodologi. Hasil ini memiliki implikasi penting, yaitu dapat menjadi panduan awal bagi individu untuk mengatur pola makan dan asupan nutrisi secara mandiri, dengan kesadaran akan kategori nutrisi makanan yang dikonsumsi. Lebih lanjut, temuan ini diharapkan dapat memberikan kontribusi signifikan dalam perencanaan dan pengambilan kebijakan terkait konsumsi makanan dan gizi di masa mendatang, mendukung upaya peningkatan kesehatan masyarakat melalui informasi nutrisi yang lebih terstruktur.

#### IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis dan mengelompokkan makanan tradisional Indonesia berdasarkan kandungan nutrisinya menggunakan algoritma K-Medoids. Dengan memanfaatkan dataset sekunder yang mencakup informasi kalori, lemak, karbohidrat, dan protein dari 1.346 jenis makanan, penelitian ini berhasil mengidentifikasi pola nutrisi tersembunyi dalam kuliner tradisional.

Penerapan algoritma K-Medoids dalam RapidMiner menghasilkan dua klaster utama. Klaster 0 berisi 1095 item dan Klaster 1 berisi 251 item. Klasterisasi ini memberikan gambaran awal mengenai distribusi makanan berdasarkan profil nutrisinya, di mana Klaster 0 diasumsikan merepresentasikan makanan dengan kandungan nutrisi rendah dan Klaster 1 dengan kandungan nutrisi tinggi, sesuai dengan penetapan di metodologi. Hasil evaluasi juga memperlihatkan nilai mendekati 0 yang artinya tiap data di masing-masing kluster memiliki kemiripan satu sama lain,

Hasil ini diharapkan dapat berkontribusi dalam perencanaan dan pengambilan kebijakan terkait konsumsi makanan dan nutrisi di masa mendatang.

#### V. REFERENSI

- [1] S. Utami, "Kuliner Sebagai Identitas Budaya: Perspektif Komunikasi Lintas Budaya," *Cover. J. Strateg. Commun.*, vol. 8, no. 2, pp. 36–44, 2018, doi: 10.35814/coverage.v8i2.588.
- [2] C. Indonesia, "Rendang & Nasi Goreng Dipilih Jadi Makanan Terenak di Dunia," *cnnindonesia.com*. Accessed: Jul. 07, 2025. [Online]. Available: <https://www.cnnindonesia.com/gaya-hidup/20170715172743-307-228130/rendang-nasi-goreng-dipilih-jadi-makanan-terenak-di-dunia>
- [3] G. Weichart, "Identitas Minahasa: Sebuah Praktik Kuliner," *Antropol. Indones.*, vol. 0, no. 74, pp. 59–80, 2014, doi: 10.7454/ai.v0i74.3510.
- [4] M. Aprilia, "Dunia Melayu Dibawah Kekuasaan Kolonial," *Hadharah J. Keislam. dan Perad.*, vol. 18, no. 1, pp. 15–25, 2024, doi: 10.15548/h.v18i1.9499.
- [5] I. M. Ayu *et al.*, "Edukasi tentang Pengaruh Kebiasaan Mengonsumsi Makanan dan Minuman Manis secara Berlebih bagi Kesehatan," *J. Abdimas Kesehat.*, vol. 5, no. 3, p. 454, 2023, doi: 10.36565/jak.v5i3.421.
- [6] dr. R. Fadli, "Berapa Banyak Kalori yang Dibutuhkan Tubuh? Cek Faktanya!," *.halodoc.com*. Accessed: Jul. 07, 2025. [Online]. Available: <https://www.halodoc.com/artikel/berapa-banyak-kalori-yang-dibutuhkan-tubuh-cek-faktanya>
- [7] M. S. Ramdhan, N. H. P. Anwar, Tsabitah. Adinda Wira Zahra, and A. S. Fitri, "Desain Aplikasi Mobile Pesan Antar Makanan Dengan Fitur Inovatif Perhitungan Nutrisi Dan Rekomendasi Menu Berbasis UML," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 3, pp. 11731–11743, 2024.
- [8] S. R. Ningsih, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, and A. Wanto, "Analisis K-Medoids Dalam Pengelompokan Penduduk Buta Huruf Menurut Provinsi," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 721, 2019, doi: 10.30645/senaris.v1i0.78.
- [9] D. S. M. Simanjuntak, I. Gunawan, S. Sumarno, P. Poningsih, and I. P. Sari, "Penerapan Algoritma K-Medoids Untuk Pengelompokan Pengangguran Umur 25 tahun Keatas Di Sumatera Utara," *J. Krisnadana*, vol. 2, no. 2, 2023, doi: 10.58982/krisnadana.v2i2.264.
- [10] E. T. Ena Tasia and M. Afdal, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Klastering Daerah Rawan Banjir Di Kabupaten Rokan Hilir," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 65–73, 2023, doi: 10.57152/ijirse.v3i1.523.
- [11] H. Galih, "Memahami Peran Dataset dalam Penelitian Kecerdasan Buatan," no. October, 2024, doi: 10.13140/RG.2.2.34468.49288.
- [12] R. K. Dinata, S. Retno, and N. Hasdyna, "Minimization of the Number of Iterations in K-Medoids Klastering with Purity Algorithm," *Rev. d'Intelligence Artif.*, vol. 35, no. 3, pp. 193–199, 2021, doi: 10.18280/ria.350302.
- [13] A. A. Panji Bintoro, Ratnasari, Edy Wihardjo, Indah Pratiwi Putri, *Pengantar Machine Learning*, vol. 3, no. 1. 2024. [Online]. Available: <http://dx.doi.org/10.1016/j.bpj.2015.06.056%0Ahttps://academic.oup.com/bioinformatics/article-abstract/34/13/2201/4852827%0Ainternal-pdf://semisupervised-3254828305/semisupervised.ppt%0Ahttp://dx.doi.org/10.1016/j.str.2013.02.005%0Ahttp://dx.doi.org/10.10>
- [14] P. Kumar and D. Sirohi, "Comparative analysis of FCM and HCM algorithm on Iris data set," *Int. J. Comput. Appl.*, vol. 5, no. 2, pp. 33–37, 2010, doi: 10.5120/888-1261.

- [15] Z. Widarto Rachbini, *DATA SCIENCE untuk Pemula dengan RapidMiner*. CV. AA. Rizky, 2025, 2025.
- [16] S. Petrovic, "A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Klusters," *11th Nord. Work. Secur. IT-systems*, pp. 53–64, 2006, [Online]. Available: [https://xp-dev.com/svn/b\\_frydrych.../silhouetteIndexRegulaStopu.pdf](https://xp-dev.com/svn/b_frydrych.../silhouetteIndexRegulaStopu.pdf)