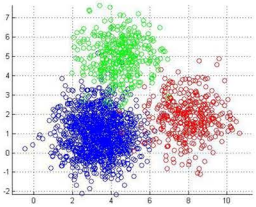


K-MEANS CLUSTERING

MONOGRAF

PENGELOMPOKAN DATA EKSPOR IKAN SEGAR/DINGIN HASIL TANGKAP MENURUT NEGARA TUJUAN UTAMA MENGGUNAKAN K-MEANS CLUSTERING

EKKA PUJO ARIESANTO AKHMAD
BUDI PRIYONO

MONOGRAF

PENGELOMPOKAN DATA EKSPOR IKAN SEGAR/DINGIN HASIL TANGKAP MENURUT NEGARA TUJUAN UTAMA MENGGUNAKAN *K-MEANS CLUSTERING*

MONOGRAF

PENGELOMPOKAN DATA EKSPOR IKAN SEGAR/DINGIN HASIL TANGKAP MENURUT NEGARA TUJUAN UTAMA MENGGUNAKAN *K-MEANS CLUSTERING*

EKKA PUJO ARIESANTO AKHMAD
BUDI PRIYONO

RC
Rumah Cemerlang

MONOGRAF

PENGELOMPOKAN DATA EKSPOR IKAN SEGAR/DINGIN HASIL TANGKAP MENURUT NEGARA TUJUAN UTAMA MENGGUNAKAN *K-MEANS CLUSTERING*

Penulis:
EKKA PUJO ARIESANTO AKHMAD
BUDI PRIYONO

Editor:
Erik Santoso

Layouter :
Tim Kreatif PRCI

Cover:
Rusli

Cetakan Pertama : Januari 2023

Perkumpulan Rumah Cemerlang Indonesia
ANGGOTA IKAPI JAWA BARAT
Pondok Karisma Residence Jalan Raflesia VI D.151
Panglayungan, Cipedes Tasikmalaya – 085223186009

Website : www.rcipress.rcipublisher.org
E-mail : rumahcemerlangindonesia@gmail.com

- Cet. I – : Perkumpulan Rumah Cemerlang Indonesia, 2023
; 14,8 x 21 cm
ISBN : 978-623-448-413-7

Undang-undang No.19 Tahun 2002 Tentang Hak Cipta
Pasal 72

Barang siapa dengan sengaja melanggar dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam pasal ayat (1) atau pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling sedikit 1 (satu) bulan dan/atau denda paling sedikit Rp.1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp.5.000.000.000,00 (lima miliar rupiah).

Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran hak cipta terkait sebagai dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp.500.000.000,00 (lima ratus juta rupiah)

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa yang telah memberikan rahmat serta karunia-Nya kepada kami sehingga kami berhasil menyelesaikan Buku dengan judul MONOGRAF Pengelompokan Data Ekspor Ikan Segar/Dingin Hasil Tangkap Menurut Negara Tujuan Utama Menggunakan *K-Means Clustering* sesuai yang ditargetkan. Buku dengan judul MONOGRAF Pengelompokan Data Ekspor Ikan Segar/Dingin Hasil Tangkap Menurut Negara Tujuan Utama Menggunakan *K-Means Clustering* ini berisikan mengenai aplikasi *K-Means Clustering* yang digunakan dalam pengelompokan data ekspor ikan segar/dingin menurut negara tujuan. Hasil ini memberikan gambaran agar pengelompokan data bisa dilakukan dengan efektif dan dengan hasil yang baik. Kami menyadari bahwa Buku ini masih jauh dari sempurna, oleh karena itu kritik dan saran dari semua pihak yang bersifat membangun selalu kami harapkan demi kesempurnaan buku ini.

Akhir kata, kami sampaikan terima kasih kepada semua pihak yang telah berperan serta dalam penyusunan Buku ini dari awal sampai akhir. Semoga Tuhan Yang Maha Esa senantiasa meridhoi segala usaha kita. Amin.

Januari 2023, Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
KONSEP DASAR METODE *CLUSTERING*

Metode *clustering* digunakan untuk melakukan penngelompokan data menjadi beberapa kategori (Adinugroho & Sari, 2018).

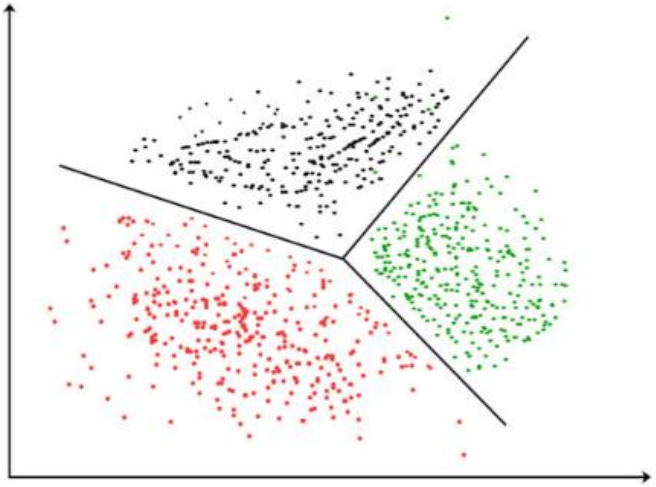
Perbedaan dengan metode klasifikasi, *clustering* mengelompokkan data menurut fitur-fitur yang ada pada data tersebut. Berdasarkan keadaan tersebut, *clustering* tidak membutuhkan data *training* yang telah didapatkan kelasnya. *Clustering* memiliki proses pembelajaran yang bersifat mandiri, karena itu dikenal dengan istilah *unsupervised learning.*

Proses *clustering* mempunyai tujuan pokok membagi sekumpulan data menjadi sekumpulan grup *(cluster)*, sehingga data-data dalam satu *cluster* memiliki banyak kemiripan, namun berbeda dengan data-data yang berada pada *cluster* lainnya. Kemiripan antar data dapat dihitung menggunakan berbagai metode pengukuran jarak, misal Euclidean, Manhattan, dan Chebyshev.

Pembentukan *cluster* dari sekumpulan data dapat dilakukan dengan berbagai cara. Secara umum, algoritma *clustering* dapat dikategorikan menjadi tiga metode (Adinugroho & Sari, 2018)

A. Metode Partisi

Metode partisi bertujuan membagi data menjadi beberapa region/cluster, sehingga setiap data hanya dapat menjadi anggota dari suatu region saja. Dengan kata lain, region yang terbentuk tidak saling tumpang tindih. Secara umum, jumlah region yang akan dibentuk harus didefinisikan terlebih dahulu. Selanjutnya, penentuan keanggotaan dilakukan berdasarkan kemiripan atau jarak setiap data dengan titik pusat (centroid) dari masing-masing region. Algoritma yang menggunakan metode partisi antara lain *K-means* dan *K-medoids*.

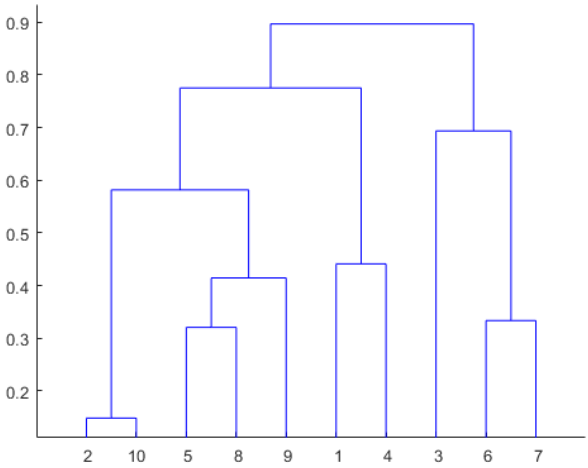


Gambar 1.1 *Cluster* menggunakan partisi

B. Metode Hierarki

Konsep *clustering* menggunakan metode hierarki memiliki kesamaan dengan metode partisi, yaitu sekumpulan data dibagi menjadi beberapa region. Bedanya, pada metode hierarki setiap region terdiri dari dua sub-region yang lebih kecil. Oleh karena itu, sub-region terkecil hanya memiliki satu data saja. Secara umum, hasil dari *clustering* menggunakan metode hierarki disajikan dalam bentuk *tree* atau pohon. Salah satu algoritme yang

memanfaatkan konsep hierarki adalah *agglomerative clustering.*

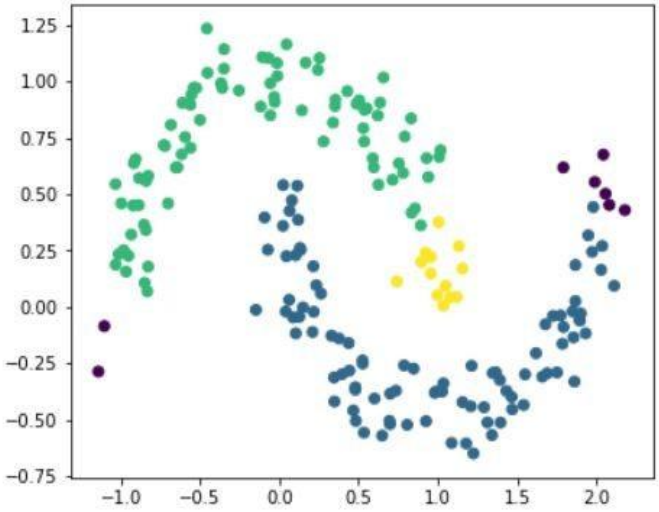


Gambar 1.2 *Cluster* menggunakan hierarki

C. Metode Berbasis Kerapatan

Metode ini membentuk *cluster* dari data-data yang saling berdekatan, sedangkan data yang saling berjauhan tidak akan menjadi anggota *cluster* manapun dan berfungsi sebagai pemisah antar *cluster.* Biasanya, tingkat kerapatan data dihitung berdasarkan jumlah data dalam suatu radius tertentu. Metode yang

menggunakan prinsip kerapatan adalah DBSCAN *(Density Based Spatial Clustering of Applications with Noise)* dan OPTICS *(Ordering Points to Identify the Clustering Structure).*



Gambar 1.3 *Cluster* menurut kerapatan data

BAB II
K-MEANS CLUSTERING, APA DAN BAGAIMANA?

A. Pengantar *K-Means CLustering*

Klasifikasi data tata cara K- Means dicoba lewat algoritma berikut (Dhuhita, 2015; Anjelita, dkk., 2019)

1. Pastikan jumlah golongan yang hendak dipecah.
2. Distribusikan data ke dalam golongan dengan cara random pusat golongan.
3. Hitung centroid atau pusat data dari data yang terdapat di tiap-tiap golongan. Posisi centroid tiap golongan didapat dari rata-rata seluruh angka data tiap fitur. Bila M menunjukkan angka informasi dalam suatu golongan, i menjelaskan fitur ke- i dalam suatu golongan, serta p menerangkan dimensi buat data, hingga rumus mengkalkulasi pusat data fitur ke- i dipakai rumus (1). Rumus dicoba sebesar p dimensi atas i=1, hingga i=p, memakai formula (1).

$$Ci = \frac{1}{M} \sum_{j=1}^{M} xj \quad (1)$$

4. Distribusikan tiap-tiap data ke pusat data ataupun

rata-rata paling dekat. Terdapat sebagian metode yang bisa dicoba buat mengukur jarak data ke pusat data, antara lain memakai *Euclidean*. Pengukuran jarak *Euclidean* bisa dicari memakai rumus (2).

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (2)$$

Pendistribusian ulang data masuk tiap- tiap golongan pada tata cara *K- Means* mengikuti perbandingan jarak antara data dengan *centroid* tiap golongan yang ada. Data distribusikan balik sesuai golongan yang memiliki *centroid* jarak terdekat lewat cara yang pasti. Pendistribusian data ini memakai rumus (3).

$$ail = \begin{cases} 1 & d = \min\{D(xi, ci)\} \\ 0 & \text{lainnya} \end{cases} \quad (3)$$

K-Means clustering memakai fungsi rasional berdasarkan jarak serta nilai keanggotaan pada golongan, menggunakan rumus (4).

$$J = \sum_{i=1}^{n} \quad \sum_{i=1}^{k} a_{ic} D\left(x_i, c_1\right)^2 \quad (4)$$

n sama dengan banyak data, k sama dengan banyak golongan, ai_1 sama dengan angka keanggotaan titik

data x_i ke golongan c_l yang diikuti, a memiliki nilai 0 atau 1. Bila data merupakan anggota suatu kelompok, nilai ai_1 sama dengan 1, sebaliknya bila tidak, maka nilai ai_1 sama dengan 0.

5. Jika ada data yang berganti kelompok, atau ada peralihan nilai *centroid* yang telah ditentukan, atau ada fluktuasi angka fungsi rasional yang dipakai di atas nilai ambang yang ditentukan, maka ulangi langkah 3.

Algoritma *K-Means* amat simpel serta mudah buat diimplementasikan. Tidak hanya itu, algoritma ini relatif cepat dalam menggolongkan data (Barakbah & Kiyoki, 2009).

B. Evaluasi

Davies Bouldin Index (*DBI*) digunakan untuk menilai luaran klaster terbaik metode *K-Means*. Klaster akan dikira mempunyai konsep *clustering* maksimum ialah yang memiliki DBI minimal (Nawrin, dkk., 2017; Sitompul, 2018).

DBI ialah salah satu tata cara yang dipakai buat mengukur keabsahan klaster pada suatu tata cara

pengelompokan, keterikatan (kohesi) dideskripsikan selaku jumlah dari kedekatan data kepada titik pusat klaster dari klaster yang diiringi. Sebaliknya pemisahan (separasi) didasarkan pada jarak antara titik pusat klaster kepada klasternya. Pengukuran *DBI* mengoptimalkan jarak *inter-cluster* antara klaster C_i serta C_j serta pada durasi yang serupa berupaya buat meminimalkan jarak antar titik dalam suatu klaster (Sitompul, 2018; Wani & Riyaz, 2017).

Langkah-langkah perhitungan DBI adalah sebagai berikut.

1. *Sum of Square Within-cluster* (*SSW*)

Keterikatan sebuah klaster ke-i diukur dengan angka *Sum of Square Within-cluster*. Formula yang dipakai untuk mendapatkan angka SSW adalah (Kusnawi, 2022).

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (5)$$

2. *Sum of Square Between-cluster* (*SSB*)

Pemisahan antar klaster dicari dengan menghitung *Sum of Square Between cluster*. Rumus