

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
MENGUNAKAN *K-MEANS CLUSTERING***

**EKKA PUJO ARIESANTO AKHMAD
BUDI PRIYONO**



MONOGRAF

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

Penulis:

EKKA PUJO ARIESANTO AKHMAD
BUDI PRIYONO

Editor:

Erik Santoso

Layouter :

Tim Kreatif PRCI

Cover:

Rusli

Cetakan Pertama : Januari 2023

Hak Cipta 2023, pada Penulis. Diterbitkan pertama kali oleh:

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

Copyright © 2023 by Perkumpulan Rumah Cemerlang Indonesia
All Right Reserved

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

Hak cipta dilindungi undang-undang

Dilarang memperbanyak buku ini dalam bentuk dan dengan
cara apapun tanpa izin tertulis dari penulis dan penerbit

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

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 kehadiran 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

KATA PENGANTAR	I
DAFTAR ISI	II
DAFTAR TABEL	III
DAFTAR GAMBAR	IV
BAB I KONSEP DASAR METODE <i>CLUSTERING</i>	1
A. Metode Partisi	2
B. Metode Hierarki	3
C. Metode Berbasis Kerapatan	4
BAB II <i>K-MEANS CLUSTERING</i> , APA DAN BAGAIMANA?	6
A. Pengantar <i>K-Means Clustering</i>	6
B. Evaluasi	8
BAB III PENTINGNYA MEMAHAMI PENEREAPAN <i>K-MEANS CLUSTERING</i> DALAM PENGELOMPOKAN DATA	12
A. Pengantar	12
B. Penelitian Pendukung	14
C. Roadmap Kajian	19
BAB IV PROSES PEMECAHAN MASALAH	20
BAB V HASIL PENGELOMPOKAN DATA EKSPOR IKAN SEGAR/DINGIN HASIL TANGKAP MENURUT NEGARA TUJUAN UTAMA MENGGUNAKAN <i>K-MEANS CLUSTERING</i>	25
A. Pengumpulan Data Awal	25
B. Pengolahan Awal (<i>Preprocessing</i>)	27
C. Penerapan algoritma K-Means untuk pengelompokan data Ekspor Ikan Segar atau Dingin Hasil Tangkap	30
D. Evaluasi	42
1. Menghitung <i>SSW</i>	43
2. Menghitung <i>SSB</i>	44
3. Menghitung R	46

4. Menghitung <i>DBI</i>	47
BAB VI PENUTUP	49
DAFTAR PUSTAKA	51
BIOGRAFI PENULIS	55

DAFTAR TABEL

Tabel 5.1 Ekspor Ikan Segar/Dingin Hasil Tangkap menurut Negara Tujuan Utama, 2019-2020	25
Tabel 5.2 Data ekspor ikan segar/dingin hasil tangkap tahun 2019	28
Tabel 5.3 Data ekspor ikan segar/dingin hasil tangkap tahun 2020	29
Tabel 5.4. Nilai Awal <i>Centroid</i> Ekspor Ikan Segar atau Dingin Hasil Tangkap Tahun 2019 (a) dan Tahun 2020 (b)	31
Tabel 5.5 Selisih masing-masing data ke <i>centroid</i> awal buat Ekspor Ikan Segar atau Dingin Hasil Tangkap Tahun 2019 (a) dan Tahun 2020 (b)	36
Tabel 5.6 Daftar Anggota Cluster Ekspor Ikan Segar atau Dingin Hasil Tangkap Tahun 2019 dan Tahun 2020 Setelah Iterasi 1	38
Tabel 5.7 <i>Centroid</i> baru klaster 1, 2, dan 3 Tahun 2019 (a) dan Tahun 2020 (b)	39
Tabel 5.8 Jarak dari setiap data ke centroid baru Tahun 2019 (a) dan Tahun 2020 (b)	40
Tabel 5.9 Anggota Klaster Aktual Ekspor Ikan Segar atau Dingin Tahun 2019 dan Tahun 2020 Setelah Iterasi 2	42

DAFTAR GAMBAR

Gambar 1.1 <i>Cluster</i> menggunakan partisi	3
Gambar 1.2 <i>Cluster</i> menggunakan hierarki	4
Gambar 1.3 <i>Cluster</i> menurut kerapatan data	5
Gambar 4.1 Alur Penelitian	20
Gambar 4.2 Diagram Alir K-Means Clustering	24

BAB I

KONSEP DASAR METODE *CLUSTERING*

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

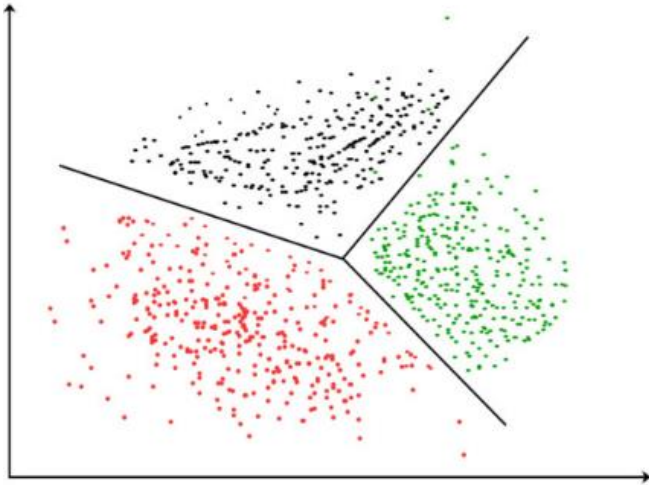
Perbedaan dengan metode klasifikasi, *clustering* mengelompokkan data menurut fitur-fitur yang ada pada data tersebut. Berdasarkan keadaan tersebut, *clustering* tidak membutuhkan data *training* yang telah didapatkan kelasnya. *Clustering* memiliki proses pembelajaran yang bersifat mandiri, karena itu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

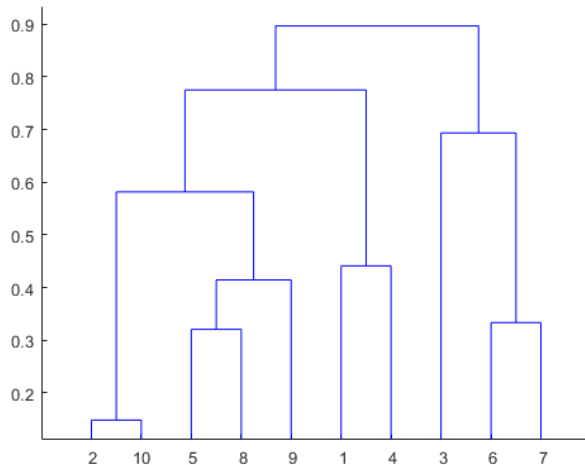


Gambar 1.1 *Cluster* menggunakan partisi

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*.

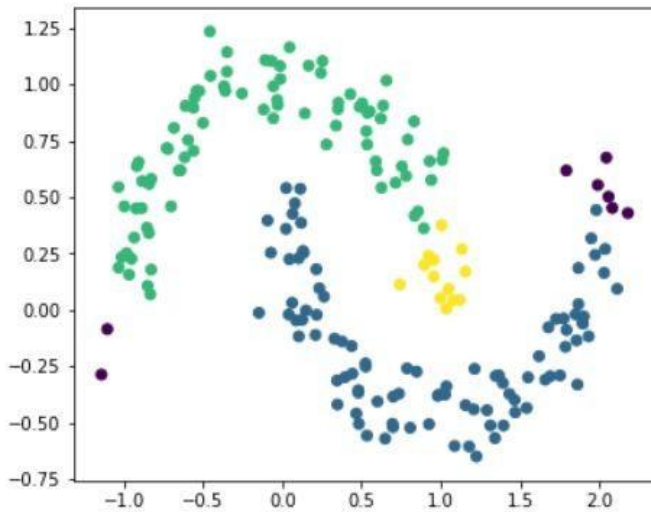


Gambar 1.2 Cluster menggunakan hierarki

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).

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

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).

$$a_{il} = \begin{cases} 1 & d = \min\{D(x_i, c_i)\} \\ 0 & \text{lainnya} \end{cases} \quad (3)$$

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

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

n sama dengan banyak data, k sama dengan banyak golongan, a_{i1} sama dengan angka keanggotaan titik

data x_i ke golongan c_1 yang diikuti, a memiliki nilai 0 atau 1. Bila data merupakan anggota suatu kelompok, nilai a_{i1} sama dengan 1, sebaliknya bila tidak, maka nilai a_{i1} 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 kluster dari kluster yang diiringi. Sebaliknya pemisahan (separasi) didasarkan pada jarak antara titik pusat kluster kepada klasternya. Pengukuran *DBI* mengoptimalkan jarak *inter-cluster* antara kluster C_i serta C_j serta pada durasi yang serupa berupaya buat meminimalkan jarak antar titik dalam suatu kluster (Sitompul, 2018; Wani & Riyaz, 2017).

Langkah-langkah perhitungan *DBI* adalah sebagai berikut.

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

Keterikatan sebuah kluster 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 kluster dicari dengan menghitung *Sum of Square Between cluster*. Rumus

yang dipakai buat mengkalkulasi angka *sum of square between cluster*, yaitu (Kusnawi, 2022).

$$SSB_{i,j} = d(c_i, c_j) \quad (6)$$

3. Rasio (*Ratio*)

Angka perbandingan antar kluster ke-i serta kluster ke-j diketahui dengan memakai rasio. Angka rasio dipunyai tiap kluster diukur dengan rumus berikut (Nawrin, dkk., 2017; Sitompul, 2018).

$$R_{i,j} = \frac{SSW_I + SSW_J}{SSB_{i,j}} \quad (7)$$

4. Indeks Davies Bouldin

Angka rasio yang didapatkan dari formula (7) dipakai guna mengetahui angka Davies Bouldin Index memakai rumus sebagai berikut.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (8)$$

Dari rumus (8), k merupakan jumlah kluster. Kian kecil angka DBI yang didapat (bukan negatif, lebih besar atau sama dengan nol), maka tambah

bagus klaster yang didapat dari penggolongan memakai algoritma clustering (Sitompul, 2018; Bates & Kalita, 2016).

BAB III

PENTINGNYA MEMAHAMI PENEREAPAN *K-MEANS CLUSTERING* DALAM PENGELOMPOKAN DATA

A. Pengantar

Kementerian Kelautan dan Perikanan melaporkan total ekspor perikanan RI mencapai 5,2 miliar dollar AS atau setara dengan Rp 72,8 triliun dan ekspor produk perikanan Indonesia justru mengalami peningkatan dan Indonesia naik 2 peringkat menjadi berada di posisi 8 sebagai eksportir utama produk perikanan dunia tahun 2020 (KKP, 2021).

Perikanan tangkap adalah salah satu zona harapan dalam membagikan partisipasi kepada perekonomian nasional lewat capaian angka eksportnya. Perihal ini karena perikanan tangkap dalam negara sanggup berdaya saing dunia serta mempunyai angka tambah besar.

Perikanan tangkap, berbeda dengan perikanan budi daya, adalah usaha penangkapan ikan dan organisme air lainnya di alam liar (laut, sungai, danau,

dan badan air lainnya). Kehidupan organisme air di alam liar dan faktor-faktornya (biotik dan abiotik) tidak dikendalikan secara sengaja oleh manusia. Perikanan tangkap sebagian besar dilakukan di laut, terutama di sekitar pantai dan landasan kontinen. Perikanan tangkap juga ada di danau dan sungai.

Jenis perikanan tangkap meliputi ikan segar/dingin hasil tangkap, udang hasil tangkap, ikan hidup hasil tangkap, kepiting, moluska, kulit kerang dan lain-lain, dan invertebrata air lainnya (BPS, 2021).

Data ekspor ikan Indonesia disusun dan disajikan dalam bentuk tabel menurut kelompok negara tujuan ekspor. Namun, tabel ekspor ikan belum dibagi dalam bentuk tingkat ekspor.

Masalah pengelompokan data ekspor ikan akan diselesaikan dengan metode *clustering*. Penelitian ini menggunakan metode *clustering* berbasis partisi, yaitu *K-Means* (Arora, et al., 2016). Algoritma *K-Means* amat simpel serta gampang buat diimplementasikan. Tidak hanya itu, algoritma ini relatif kilat dalam membagi informasi (Adinugroho & Sari, 2018).

Tujuan penelitian berikut ini merupakan uraian hasil yang akan dicapai melalui penelitian, yaitu menerapkan *K-Means clustering* untuk membagi data ekspor ikan berdasarkan tingkat ekspor.

Keutamaan penelitian adalah membagi sekumpulan data ekspor ikan menjadi sekumpulan grup (klaster) tingkat ekspor (tinggi, menengah, atau rendah), alhasil data-data dalam satu klaster mempunyai banyak kesamaan, tetapi berlainan dengan data-data yang terletak pada klaster yang lain. Kesamaan antara informasi bisa dihitung memakai bermacam tata cara pengukuran jarak.

B. Penelitian Pendukung

Penelitian terdahulu yang mendasari penelitian ini adalah sebagai berikut.

Ekspor kopi merupakan kegiatan yang difokuskan pemerintah agar terus mendorong ekspor Indonesia secara keseluruhan. Salah satu strategi, yaitu mengelompokkan negara-negara tujuan ekspor kopi Indonesia. Oleh karena itu, riset ini mengerjakan pengelompokan data ekspor kopi sesuai dengan

negara destinasi dengan memakai *framework CRISP-DM* yang mempunyai 6 jenjang. Riset ini menimbang Algoritma *K-means* serta *K-medoids* dengan memakai perangkat lunak *RapidMiner*. Luaran algoritma yang paling optimal dinilai berdasarkan Indeks *Davies Bouldin* yang terkecil. Hasil yang didapatkan akan divisualisasikan agar dapat menjadi informasi yang dapat dianalisa. Hasil Analisa didapatkan Ekspor kopi Indonesia dapat ditingkatkan ke negara Uni Eropa dikarenakan memiliki tingkat konsumsi yang sangat tinggi terlebih negara Jerman yang memiliki nilai impor yang cukup tinggi. Ekspor kopi Indonesia terhadap negara Amerika Serikat harus dipertahankan (Gaghana, 2022).

Iskandar, dkk. (2020) mengelompokkan Ekspor dan Impor di Indonesia berdasarkan Bulan (Januari-Desember tahun 2018), agar diketahui perbandingan tingkat ekspor dan impor, tinggi ataupun rendah. Penggunaan algoritma *K-Means* dilakukan untuk pengelompokkan nilai Ekspor dan Impor di Indonesia berdasarkan bulan. Pengelompokan terbagi menjadi 2 kluster: Ekspor dan Impor kluster tinggi dan Ekspor

dan Impor klasteri rendah. Hasil untuk Ekspor klaster tinggi terdiri dari 5 bulan, dan kategori rendah terdiri dari 7 bulan. Sedangkan hasil untuk Impor klaster tinggi terdiri dari 7 bulan, dan kategori rendah terdiri dari 5 bulan.

Clustering *K-Means* Pada Data Ekspor (Studi Kasus: PT. Gaikindo) memberikan gambaran dan solusi kepada PT. Gaikindo dalam menentukan pasar yang terbaik untuk mengekspor produk mobilnya dengan memaksimalkan penggunaan metode clustering *k-means*. Permasalahan pengelompokan negara untuk tempat ekspor mobil dapat diselesaikan dengan menggunakan metode clustering *k-means*. Walaupun mampu menyelesaikan data dengan jumlah banyak, namun masih belum efisien dalam mengelompokkan dokumen secara tepat. Hasil akhir juga sangat bergantung pada centroid yang ditentukan, karena jika berbeda centroid yang ditentukan, maka hasilnya pun akan berbeda juga (Afifi, dkk., 2020).

Anjelita, dkk. (2019) membahas tentang ekspor perhiasan dan barang berharga dengan menggunakan

data mining *K-Means Clustering*. Data dikelompokkan menjadi 2, *cluster* ekspor perhiasan tingkat tinggi (C1) dan *cluster* ekspor perhiasan tingkat rendah (C2). Dengan menggunakan nilai *centroid* awal untuk ekspor tingkat tinggi (C1) 6.657,4 dan ekspor tingkat rendah (C2) 41,5. Metode *K-Means Clustering* pada penelitian ini menghasilkan 4 iterasi, sehingga diperoleh hasil akhir yaitu : *cluster* ekspor perhiasan tingkat tinggi terdapat pada Negara Uni Emirat Arab, Afrika Selatan, Taiwan, Amerika Serikat, India, Australia, Italia, dan Negara lainnya, sedangkan *cluster* ekspor perhiasan tingkat rendah yaitu Swiss, Singapura, dan Hongkong.

Penelitian mengkaji penggunaan metode *K-means* dalam studi kasus ekspor produk Indonesia ke Korea Selatan dilakukan untuk pengendalian persediaan produk atau sebagai referensi *Indonesia Trade Promotion Center (ITPC)* di Busan untuk melihat produk ekspor apa saja yang harus dipertahankan dan produk ekspor apa yang harus ditingkatkan untuk dipromosikan di Korea Selatan, sehingga dapat

meningkatkan jumlah produk ekspor Indonesia ke Korea Selatan (Pandin, C R M & Fahrudin, 2019).

Windarto, Agus Perdana (2017) mempraktikkan *K-Means* untuk mengkluster ekspor buah bersumber pada negeri destinasi. Evaluasi bersumber pada indikator ekspor buah, yakni dua negeri kluster tingkatan ekspor besar ialah India serta Pakistan, 3 negeri kluster tingkatan ekspor menengah ialah Singapore, Bangladesh serta negeri yang lain serta 6 negeri kluster tingkatan ekspor kecil, ialah Cina, Hongkong, Iran, Nepal, Malaysia, serta Vietnam. Hasil dari riset bisa dipakai buat mengenali ekspor buah bagi negeri tujuan.

Riset yang dilaksanakan penulis sekarang memiliki kesamaan dengan penelitian sebelumnya, yaitu menggunakan metode *K-Means Clustering*. Kebaruan penelitian (novelty), adalah hasil cluster lebih baik, karena memiliki nilai variansi dalam cluster yang tinggi dan nilai variansi antar kluster yang rendah.

C. Roadmap Kajian

Tahun 2021-2022: Pengelompokan data ekspor ikan segar/dingin hasil tangkap menggunakan K-Means Clustering;

Tahun 2022-2023: Pengelompokan data ekspor ikan segar/dingin hasil tangkap menggunakan K-Means Clustering dan K-Medoids Clustering;

Tahun 2023-2024 Pengelompokan data ekspor ikan segar/dingin hasil tangkap menggunakan K-Means Clustering dan Fuzzy C-Means Clustering;

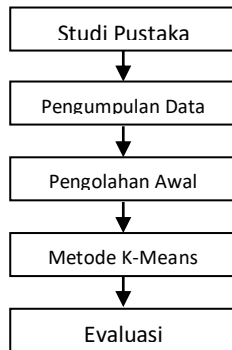
Tahun 2024-2025: Pengelompokan data ekspor ikan segar/dingin hasil tangkap menggunakan Hierarchy Agglomerative Clustering;

Tahun 2025-2026: Pengelompokan data ekspor ikan segar/dingin hasil tangkap menggunakan DBSCAN (Density Based Spatial Clustering of Applications with Noise).

BAB IV

PROSES PEMECAHAN MASALAH

Penelitian ini mempunyai bagan alir penelitian dapat dilihat pada Gambar 3.1.



Gambar 4.1 Alur Penelitian

Studi Pustaka

Peneliti melakukan studi pustaka pada riset terdahulu yang berhubungan dengan poin riset. Asal muasal studi pustaka didapat dari buku serta tulisan riset *data mining* mengenai *clustering*.

Pengumpulan Data

Pengumpulan data dicoba dengan memakai kumpulan data ekspor ikan segar atau dingin hasil tangkap yang diperoleh dari Badan Pusat Statistik tahun 2019-2020 dengan menggunakan empat buah indikator (atribut), yaitu negara, tahun, berat bersih (ton), dan nilai *FOB (US Dollar)*.

Pengolahan Awal

Saat melaksanakan data mining, ada beberapa tahap melaksanakan *preprocessing*. Awal, melaksanakan pembersihan (*cleaning*) data melingkupi antara lain; membuang data ganda, mengecek informasi yang inkonsisten, serta mengkoreksi kekeliruan pada data.

Metode K-Means

Integrasi data butuh dicoba buat menyatukan data yang terpisah ke dalam basis data terkini supaya bebas dari pengulangan data. Berikutnya melaksanakan pemilahan data dengan cara meminimalkan jumlah data yang hendak dipakai buat berikutnya dicoba cara mining dengan senantiasa menjaga data aslinya. Sehabis melaksanakan

pemilahan data, berikutnya mengubah data ataupun mencampurkan data ke dalam wujud normalisasi. Jenjang berikutnya ialah melaksanakan cara mining ataupun penambangan data memakai tata cara *K-Means Clustering*.

Cara data mining dicoba dengan memakai tata cara clustering. Algoritma yang dipakai merupakan *K-Means*. Cara awal yang dicoba dengan memasukkan data. Berikutnya dicoba clustering memakai algoritma *K-Means* buat memperoleh jumlah klaster. Langkah-langkah prosedur *K-Means* meliputi hal sebagai berikut.

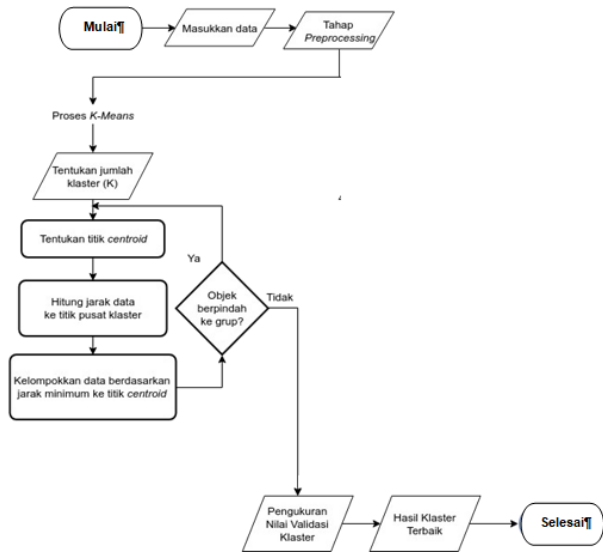
- a. Memasukkan data.
- b. Pengolahan awal (*preprocessing*).
- c. Metode K-Means,
 - 1). Menetapkan jumlah klaster (k);
 - 2). Menetapkan titik pusat data;
 - 3). Mengukur jarak data ke titik pusat klaster;
 - 4). Membagi data bersumber pada jarak minimal ke titik centroid;
 - 5) Mengukur nilai validasi klaster;
 - 6) Mendapatkan hasil klaster terbaik.

d. Melakukan evaluasi hasil kluster.

Evaluasi

Riset ini melaksanakan penilaian kepada hasil kluster terbaik *K-Means* dengan *DBI*. Langkah-langkah perhitungan *DBI* adalah sebagai berikut.

1. Menghitung Sum of Square Within-cluster (*SSW*) untuk mengetahui kohesi dalam sebuah *cluster* ke-*i*.
2. Menghitung Sum of Square Between-cluster (*SSB*) untuk mengetahui separasi antar *cluster*.
3. Mencari Rasio (Ratio) untuk mengetahui nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*.
4. Menghitung *DBI* untuk mengetahui nilai indeks Davies Bouldin. Semakin kecil nilai *DBI* yang diperoleh (bukan angka negatif), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering*.



Gambar 4.2 Diagram Alir K-Means Clustering

BAB V

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

A. Pengumpulan Data Awal

Pengumpulan data dilakukan di Badan Pusat Statistik Data ekspor ikan segar/dingin hasil tangkap memiliki empat buah indikator (atribut), yaitu negara, tahun, berat bersih (ton), dan nilai FOB (US Dollar).

Tabel 5.1 menjelaskan tabel ekspor ikan segar/dingin hasil tangkap menurut negara tujuan utama tahun 2019-2020.

**Tabel 5.1 Ekspor Ikan Segar/Dingin Hasil Tangkap
menurut Negara Tujuan Utama, 2019-2020**

Negara tujuan	2019	2020
Berat bersih: Ton		
Jepang	2.265,4	3.025,9
Malaysia	28.359,	30.225,

	8	9
	15.452,	15.591,
Singapura	2	1
Amerika		
Serikat	171,0	101,6
Taiwan	2.311,3	2.448,9
Arab Saudi	697,8	262,1
Hongkong	2.124,3	2.519,9
Tiongkok	1.874,1	2.617,4
Australia	320,8	271,7
Uni Emirat		
Arab	144,5	58,6
Lainnya	185,5	571,2
	53.906,	57.694,
Jumlah	7	3

Nilai FOB : 000 US\$		
	12.148,	15.173,
Jepang	8	8
	50.970,	47.183,
Malaysia	4	7

Singapura	32.241,4	34.762,9
Amerika Serikat	1.641,5	703,0
Taiwan	8.983,1	7.674,0
Arab Saudi	2.506,0	944,2
Hongkong	7.071,8	7.539,6
Tiongkok	6.590,3	7.808,0
Australia	2.064,8	2.241,6
Uni Emirat Arab	471,2	163,8
Lainnya	398,8	1.588,1
Jumlah	125.08	125.78
	8,1	2,7

Sumber:BPS

B. Pengolahan Awal (*Preprocessing*)

Langkah pertama pengolahan awal, yaitu melakukan proses cleaning pada data. Integrasi data perlu dilakukan untuk menggabungkan/penyatuan data yang terpisah ke dalam database baru. Selanjutnya melakukan seleksi data dengan proses

meminimalkan jumlah data yang akan digunakan. Setelah melakukan seleksi data, selanjutnya mentransformasi data atau menggabungkan data ke dalam bentuk normalisasi.

Hasil pengolahan awal data ekspor ikan segar/dingin hasil tangkap menurut negara tujuan utama tahun 2019-2020 dijelaskan pada Tabel 5.2 dan Tabel 5.3.

Tabel 5.2 Data ekspor ikan segar/dingin hasil tangkap tahun 2019

Negara	Berat Bersih (ton)	FOB (dollar)
Jepang	2.265,4	12.148,8
Malaysia	28.359,8	50.970,4
Singapura	15.452,2	32.241,4
Amerika Serikat	171,0	1.641,5
Taiwan	2.311,3	8.983,1
Arab Saudi	697,8	2.506,0
Hongkong	2.124,3	7.071,8
Tiongkok	1.874,1	6.590,3
Australia	320,8	2.064,8
Uni Emirat Arab	144,5	471,2
Lainnya	185,5	398,8

Tabel 5.3 Data ekspor ikan segar/dingin hasil tangkap tahun 2020

Negara	Berat Bersih (ton)	FOB (dollar)
Jepang	3.025,9	15.173,8
Malaysia	30.225,9	47.183,7
Singapura	15.591,1	34.762,9
Amerika Serikat	101,6	703,0
Taiwan	2.448,9	7.674,0
Arab Saudi	262,1	944,2
Hongkong	2.519,9	7.539,6
Tiongkok	2.617,4	7.808,0
Australia	271,7	2.241,6
Uni Emirat Arab	58,6	163,8
Lainnya	571,2	1.588,1

Metode K-Means dimulai dengan memilih jumlah *cluster* sebanyak K buah. Selanjutnya, K buah data diambil secara acak dari dataset sebagai *centroid* yang mewakili suatu *cluster*.

Centroid merupakan pusat atau titik tengah dari suatu *cluster*. Semua data kemudian dihitung jaraknya terhadap setiap *centroid* dan setiap data akan menjadi anggota dari sebuah *cluster* yang diwakili oleh *centroid* yang memiliki jarak terdekat dengan data tersebut.

Tahap terakhir adalah penghitungan ulang nilai *centroid* yang diperoleh dari nilai rata-rata dari

setiap *cluster* yang ada. Proses pemilihan keanggotaan *cluster* dan penghitungan ulang *centroid* dilakukan terus menerus dan berhenti jika keanggotaan *cluster* tidak mengalami perubahan atau jumlah perulangan yang dilakukan telah melampaui suatu nilai batas tertentu.

C. Penerapan algoritma K-Means untuk pengelompokan data Ekspor Ikan Segar atau Dingin Hasil Tangkap

Langkah pertama (iterasi 1) menerapkan *K-Means clustering*, yakni penentuan jumlah klaster yang hendak dibuat. Jika jumlah klaster yang diinginkan adalah tiga buah (tinggi, menengah, rendah), maka nilai K diset 3. Sesuai dengan nilai K , dipilih tiga data secara acak sebagai *centroid*. Data *centroid* Ekspor Ikan Segar atau Dingin Hasil Tangkap Tahun 2019 diambil dari data kedua, kelima, dan kesepuluh dari Tabel 5.2. Sedangkan data *centroid* Ekspor Ikan Segar atau Dingin Hasil Tangkap Tahun 2020 diambil dari data ketiga, satu, dan keempat dari Tabel 5.3. Angka *centroid* awal bisa diamati pada Tabel 5.4.

Tabel 5.4. Nilai Awal *Centroid* Ekspor Ikan Segar atau Dingin Hasil Tangkap Tahun 2019 (a) dan Tahun 2020 (b)

	Centroid	Berat Bersih (ton)	FOB (dollar)
Data 2	1	28.359,8	50.970,4
Data 5	2	2.311,3	8.983,1
Data 10	3	144,5	471,2

(a)

	Centroid	Berat Bersih (ton)	FOB (dollar)
Data 3	1	15.591,1	34.762,9
Data 1	2	3.025,9	15.173,8
Data 4	3	101,6	703,0

(b)

Langkah berikutnya menghitung jarak antar data terhadap semua *centroid* yang ada. Penilaian jarak antara suatu data dengan suatu *centroid* disimbolkan dengan $d(x_i, c_i)$, menggunakan metode *Euclidean*. Ilustrasi kalkulasi jarak data pertama pada tiap kluster buat Tahun 2019 sebagai berikut

$$\begin{aligned}
d(x1,c1) &= \sqrt{(2.265,4 - 28.359,8)^2 + (12.148,8 - 50.970,4)^2} = 46.776,4 \\
d(x1,c2) &= \sqrt{(2.265,4 - 2.311,3)^2 + (12.148,8 - 8.983,1)^2} = 3.166,0 \\
d(x1,c3) &= \sqrt{(2.265,4 - 144,5)^2 + (12.148,8 - 471,2)^2} = 11.868,6 \\
d(x2,c1) &= \sqrt{(28.359,8 - 28.359,8)^2 + (50.970,4 - 50.970,4)^2} = 0,0 \\
d(x2,c2) &= \sqrt{(28.359,8 - 2.311,3)^2 + (50.970,4 - 8.983,1)^2} = 49.411,1 \\
d(x2,c3) &= \sqrt{(28.359,8 - 144,5)^2 + (50.970,4 - 471,2)^2} = 57.847,0 \\
d(x3,c1) &= \sqrt{(14.452,2 - 28.359,8)^2 + (32.241,4 - 50.970,4)^2} = 22.746,0 \\
d(x3,c2) &= \sqrt{(14.452,2 - 2.311,3)^2 + (32.241,4 - 8.983,1)^2} = 26.713,9 \\
d(x3,c3) &= \sqrt{(14.452,2 - 144,5)^2 + (32.241,4 - 471,2)^2} = 35.265,7 \\
d(x4,c1) &= \sqrt{(171 - 28.359,8)^2 + (1.641,5 - 50.970,4)^2} = 56.815,0 \\
d(x4,c2) &= \sqrt{(171 - 2.311,3)^2 + (1.641,5 - 8.983,1)^2} = 7.647,2 \\
d(x4,c3) &= \sqrt{(171 - 144,5)^2 + (1.641,5 - 471,2)^2} = 1.170,6 \\
d(x5,c1) &= \sqrt{(2.311,3 - 28.359,8)^2 + (8.983,1 - 50.970,4)^2} = 49.411,1 \\
d(x5,c2) &= \sqrt{(2.311,3 - 2.311,3)^2 + (8.983,1 - 8.983,1)^2} = 0,0 \\
d(x5,c3) &= \sqrt{(2.311,3 - 144,5)^2 + (8.983,1 - 471,2)^2} = 8.783,4 \\
d(x6,c1) &= \sqrt{(697,8 - 28.359,8)^2 + (2.506 - 50.970,4)^2} = 55.803,1 \\
d(x6,c2) &= \sqrt{(697,8 - 2.311,3)^2 + (2.506 - 8.983,1)^2} = 6.675,0 \\
d(x6,c3) &= \sqrt{(697,8 - 144,5)^2 + (2.506 - 471,2)^2} = 2.108,7 \\
d(x7,c1) &= \sqrt{(2.124,3 - 28.359,8)^2 + (7.071,8 - 50.970,4)^2} = 51.140,9 \\
d(x7,c2) &= \sqrt{(2.124,3 - 2.311,3)^2 + (7.071,8 - 8.983,1)^2} = 1.920,4 \\
d(x7,c3) &= \sqrt{(2.124,3 - 144,5)^2 + (7.071,8 - 471,2)^2} = 6.891,1 \\
d(x8,c1) &= \sqrt{(1.874,1 - 28.359,8)^2 + (6.590,3 - 50.970,4)^2} = 51.682,5
\end{aligned}$$

$$\begin{aligned}
d(x8,c2) &= \sqrt{(1.874,1 - 2.311,3)^2 + (6.590,3 - 8.983,1)^2} = 2.432,4 \\
d(x8,c3) &= \sqrt{(1.874,1 - 144,5)^2 + (6.590,3 - 471,2)^2} = 6.358,8 \\
d(x9,c1) &= \sqrt{(320,8 - 28.359,8)^2 + (2.064,8 - 50.970,4)^2} = 56.373,2 \\
d(x9,c2) &= \sqrt{(320,8 - 2.311,3)^2 + (2.064,8 - 8.983,1)^2} = 7.199,0 \\
d(x9,c3) &= \sqrt{(320,8 - 144,5)^2 + (2.064,8 - 471,2)^2} = 1.603,3 \\
d(x10,c1) &= \sqrt{(144,5 - 28.359,8)^2 + (471,2 - 50.970,4)^2} = 57.847,0 \\
d(x10,c2) &= \sqrt{(144,5 - 2.311,3)^2 + (471,2 - 8.983,1)^2} = 8.783,4 \\
d(x10,c3) &= \sqrt{(144,5 - 144,5)^2 + (471,2 - 471,2)^2} = 0,0 \\
d(x11,c1) &= \sqrt{(185,5 - 28.359,8)^2 + (398,8 - 50.970,4)^2} = 57.890,2 \\
d(x11,c2) &= \sqrt{(185,5 - 2.311,3)^2 + (398,8 - 8.983,1)^2} = 8.843,6 \\
d(x11,c3) &= \sqrt{(185,5 - 144,5)^2 + (398,8 - 471,2)^2} = 83,2
\end{aligned}$$

Perhitungan jarak data kesatu pada tiap-tiap kluster untuk Tahun 2020, yaitu

$$\underline{\underline{d}}(x1,c1) = \sqrt{(3.025,9 - 15.591,1)^2 + (15.173,8 - 34.762,9)^2} = 23.272,7\text{¶}$$

$$\underline{\underline{d}}(x1,c2) = \sqrt{(3.025,9 - 3.025,9)^2 + (15.173,8 - 15.173,8)^2} = 0,0\text{¶}$$

$$\underline{\underline{d}}(x1,c3) = \sqrt{(3.025,9 - 101,6)^2 + (15.173,8 - 703,0)^2} = 14.763,3\text{¶}$$

$$\underline{\underline{d}}(x2,c1) = \sqrt{(30.225,9 - 15.591,1)^2 + (47.183,7 - 34.762,9)^2} = 19.195,1\text{¶}$$

$$\underline{\underline{d}}(x2,c2) = \sqrt{(30.225,9 - 3.025,9)^2 + (47.183,7 - 15.173,8)^2} = 42.005,6\text{¶}$$

$$\underline{\underline{d}}(x2,c3) = \sqrt{(30.225,9 - 101,6)^2 + (47.183,7 - 703,0)^2} = 55.388,9\text{¶}$$

$$\underline{\underline{d}}(x3,c1) = \sqrt{(15.591,1 - 15.591,1)^2 + (34.762,9 - 34.762,9)^2} = 0,0\text{¶}$$

$$\underline{\underline{d}}(x3,c2) = \sqrt{(15.591,1 - 3.025,9)^2 + (34.762,9 - 15.173,8)^2} = 23.272,7\text{¶}$$

$$\underline{\underline{d}}(x3,c3) = \sqrt{(15.591,1 - 101,6)^2 + (34.762,9 - 703,0)^2} = 37.416,6\text{¶}$$

$$\underline{\underline{d}}(x4,c1) = \sqrt{(101,6 - 15.591,1)^2 + (703,0 - 34.762,9)^2} = 37.416,6\text{¶}$$

$$\underline{\underline{d}}(x4,c2) = \sqrt{(101,6 - 3.025,9)^2 + (703,0 - 15.173,8)^2} = 14.763,3\text{¶}$$

$$\underline{\underline{d}}(x4,c3) = \sqrt{(101,6 - 101,6)^2 + (703,0 - 703,0)^2} = 0,0\text{¶}$$

$$\underline{\underline{d}}(x5,c1) = \sqrt{(2.448,9 - 15.591,1)^2 + (7.674,0 - 34.762,9)^2} = 30.108,6\text{¶}$$

$$\underline{\underline{d}}(x5,c2) = \sqrt{(2.448,9 - 3.025,9)^2 + (7.674,0 - 15.173,8)^2} = 7.522,0\text{¶}$$

$$\underline{\underline{d}}(x5,c3) = \sqrt{(2.448,9 - 101,6)^2 + (7.674,0 - 703,0)^2} = 7.355,6\text{¶}$$

$$\underline{\underline{d}}(x6,c1) = \sqrt{(262,1 - 15.591,1)^2 + (944,2 - 34.762,9)^2} = 37.130,6\text{¶}$$

$$\underline{\underline{d}}(x6,c2) = \sqrt{(262,1 - 3.025,9)^2 + (944,2 - 15.173,8)^2} = 14.495,5\text{¶}$$

$$\underline{\underline{d}}(x6,c3) = \sqrt{(262,1 - 101,6)^2 + (944,2 - 703,0)^2} = 289,7\text{¶}$$

$$\begin{aligned}
d(x7,c2) &= \sqrt{(2.519,9 - 3.025,9)^2 + (7.539,6 - 15.173,8)^2} = 7.651,0 \\
d(x7,c3) &= \sqrt{(2.519,9 - 101,6)^2 + (7.539,6 - 703,0)^2} = 7.251,7 \\
d(x8,c1) &= \sqrt{(2.617,4 - 15.591,1)^2 + (7.808,0 - 34.762,9)^2} = 29.914,6 \\
d(x8,c2) &= \sqrt{(2.617,4 - 3.025,9)^2 + (7.808,0 - 15.173,8)^2} = 7.377,1 \\
d(x8,c3) &= \sqrt{(2.617,4 - 101,6)^2 + (7.808,0 - 703,0)^2} = 7.537,3 \\
d(x9,c1) &= \sqrt{(271,7 - 15.591,1)^2 + (2.241,6 - 34.762,9)^2} = 35.948,8 \\
d(x9,c2) &= \sqrt{(271,7 - 3.025,9)^2 + (2.241,6 - 15.173,8)^2} = 13.222,2 \\
d(x9,c3) &= \sqrt{(271,7 - 101,6)^2 + (2.241,6 - 703,0)^2} = 1.548,0 \\
d(x10,c1) &= \sqrt{(58,6 - 15.591,1)^2 + (163,8 - 34.762,9)^2} = 37.925,7 \\
d(x10,c2) &= \sqrt{(58,6 - 3.025,9)^2 + (163,8 - 15.173,8)^2} = 15.300,5 \\
d(x10,c3) &= \sqrt{(58,6 - 101,6)^2 + (163,8 - 703,0)^2} = 540,9 \\
d(x11,c1) &= \sqrt{(571,2 - 15.591,1)^2 + (1.588,1 - 34.762,9)^2} = 36.416,5 \\
d(x11,c2) &= \sqrt{(571,2 - 3.025,9)^2 + (1.588,1 - 15.173,8)^2} = 13.805,7 \\
d(x11,c3) &= \sqrt{(571,2 - 101,6)^2 + (1.588,1 - 703,0)^2} = 1.002,0
\end{aligned}$$

Perincian jarak secara lengkap disajikan pada Tabel 5.5, yakni jarak dari setiap data ke *centroid* yang telah dipilih sebelumnya.

**Tabel 5.5 Selisih masing-masing data ke *centroid* awal
 buat Ekspor Ikan Segar atau Dingin Hasil Tangkap
 Tahun 2019 (a) dan Tahun 2020 (b)**

Jarak ke centroid		
1	2	3
46.776,4	3.166,0	11.868,6
0,0	49.411,1	57.847,0
22.746,0	26.713,9	35.265,7
56.815,0	7.647,2	1.170,6
49.411,1	0,0	8.783,4
55.803,1	6.675,0	2.108,7
51.140,9	1.920,4	6.891,1
51.682,5	2.432,4	6.358,8
56.373,2	7.199,0	1.603,3
57.847,0	8.783,4	0,0
57.890,2	8.843,6	83,2

(a)

Jarak ke centroid		
1	2	3
23.272,7	0,0	14.763,3
19.195,1	42.005,6	55.388,9
0,0	23.272,7	37.416,6
37.416,6	14.763,3	0,0
30.108,6	7.522,0	7.355,6
37.130,6	14.495,5	289,7
30.198,7	7.651,0	7.251,7
29.914,6	7.377,1	7.537,3
35.948,8	13.222,2	1.548,0
37.925,7	15.300,5	540,9
36.416,5	13.805,7	1.002,0

(b)

Setelah jarak antara data dan *centroid* awal dihitung, selanjutnya dilakukan penentuan keanggotaan *cluster* berdasarkan jarak *centroid* terdekat. Bagian yang diarsir pada Tabel 5.5 menunjukkan jarak terdekat dari data ke *centroid*. Berdasarkan informasi tersebut, tabel keanggotaan *cluster* dapat dibentuk. Tabel keanggotaan cluster ditampilkan pada Tabel 5.6.

Tabel 5.6 Daftar Anggota Cluster Ekspor Ikan Segar atau Dingin Hasil Tangkap Tahun 2019 dan Tahun 2020 Setelah Iterasi 1

Anggota Cluster		
Cluster 1	Cluster 2	Cluster 3
Malaysia	Jepang	Amerika Serikat
Singapura	Taiwan	Arab Saudi
	Hongkong	Australia
	Tiongkok	Uni Emirat Arab
		Lainnya

Untuk memastikan tidak ada perubahan pada anggota *cluster*, maka perlu ditentukan *centroid* baru. Perhitungan *centroid* baru menggunakan hasil anggota *cluster* yang diperoleh sebelumnya. *Centroid* baru masing-masing *cluster* dihitung dengan rumus

$$Centroid \text{ baru} = \frac{\text{anggota cluster } (x, y, z)}{\text{jumlah anggota cluster}} \quad (9)$$

Tahun 2019

Berat Bersih:

$$C1 = (28.359,8 + 15.452,2) / 2 = 21.906$$

$$C2 = (2.265,4 + 2.311,3 + 2.124,3 + 1.874,1) / 4 = 2.143,8$$

$$C3 = (171+697,8+320,8+144,5+185,5) / 5 = 303,9$$

FOB:

$$C1 = (50.970,4+32.241,4) / 2 = 41.605,9$$

$$C2 = (12.148,8+8.983,1+7.071,8+6.590,3) / 4 = 8.698,5$$

$$C3 = (1.641,5+2506+2.064,8+471,2+398,8) / 5 = 1.416,5$$

Tahun 2020

Berat Bersih:

$$C1 = (30.225,9 + 15.591,1) / 2 = 22.908,5$$

$$C2 = (3.025,9 + 2.488,9 + 2.519,9 + 2.617,4) / 4 = 2.653$$

$$C3 = (101,6+262,1+271,7+58,6+571,2) / 5 = 253$$

FOB:

$$C1 = (47.183,7+34.762,9) / 2 = 40.973,3$$

$$C2 = (15.173,8+7.674,0+7.539,6+7.808,0) / 4 = 9.548,9$$

$$C3 = (703,0+944,2+2.241,6+163,8+1.588,1) / 5 = 1.128,1$$

Hasil perhitungan *centroid* baru untuk cluster 1, 2, dan 3 dijelaskan pada Tabel 5.7.

Tabel 5.7 *Centroid* baru klaster 1, 2, dan 3 Tahun 2019

(a) dan Tahun 2020 (b)

<u>Centroid</u>	<u>Berat Bersih (ton)</u>	<u>FOB (dollar)</u>
1	21.906	41.605,9
2	2.143,8	8.698,5
3	303,9	1.416,5

(a)

Centroid	Berat Bersih (ton)	FOB (dollar)
1	22.908,5	40.973,3
2	2.653	9.548,9
3	253	1.128,1

(b)

Iterasi 2 menghitung kembali jarak *Euclidean Distance* dari masing-masing data ke *centroid* baru, lewat cara yang sama seperti iterasi 1. Luaran kalkulasi jarak yang didapat bisa diamati pada Tabel 5.8.

Tabel 5.8 Jarak dari setiap data ke centroid baru Tahun 2019 (a) dan Tahun 2020 (b)

Jarak ke centroid baru		
1	2	3
35.404,4	3.452,4	10.910,1
11.373,0	49.741,3	56.944,9
11.373,0	27.044,1	34.346,0
45.492,5	7.327,6	261,4
38.055,2	330,2	7.828,4
44.481,3	6.359,1	1.158,6
39.798,5	1.626,8	5.941,1
40.340,7	2.125,4	5.406,9
45.049,1	6.879,6	648,6
46.536,3	8.466,7	958,6
46.581,2	8.527,6	1.024,5

(a)

Jarak ke centroid baru		
1	2	3
32.572,0	5.637,3	14.316,7
9.597,6	46.654,5	54.949,9
9.597,6	28.339,8	36.966,9
46.280,1	9.206,5	451,3
39.082,5	1.885,9	6.904,4
45.991,2	8.930,7	184,2
39.160,0	2.013,7	6.800,4
38.880,1	1.741,2	7.086,0
44.861,7	7.685,5	1.113,6
46.771,1	9.737,1	983,7
45.278,6	8.228,5	559,3

(b)

Berikutnya jarak tiap objek data ke *centroid* baru dibandingkan. Dari perbandingan tersebut diketahui, bahwa anggota klaster aktual bisa diamati pada Tabel 5.9.

Tabel 5.9 Anggota Kluster Aktual Ekspor Ikan Segar atau Dingin Tahun 2019 dan Tahun 2020 Setelah Iterasi 2

Anggota Cluster		
Cluster 1	Cluster 2	Cluster 3
Malaysia	Jepang	Amerika Serikat
Singapura	Taiwan	Arab Saudi
	Hongkong	Australia
	Tiongkok	Uni Emirat Arab
		Lainnya

Setelah iterasi kedua selesai, ternyata tidak terdapat perubahan keanggotaan *cluster* antara iterasi pertama dan kedua. Pada kondisi seperti ini, *cluster* dianggap *konvergen* dan iterasi dapat dihentikan. Apabila *cluster* yang terbentuk tidak dapat mencapai kondisi *konvergen*, atau keanggotaan cluster selalu berubah-ubah, maka proses *K-Means* dapat dihentikan setelah melalui n iterasi.

D. Evaluasi

Davies Bouldin Index (DBI) dipakai untuk menilai luaran kluster terbaik metode *K-Means*. Suatu

kluster hendak dikira mempunyai desain *clustering* yang maksimal, yaitu kluster yang mempunyai *DBI* minimum. Langkah-langkah kalkulasi *DBI* sebagai berikut.

1. Menghitung *SSW*

Data *centroid* baru cluster 1, 2, dan 3 dan jarak dari setiap data ke *centroid* baru hasil iterasi terakhir *clustering* Tahun 2019 dan Tahun 2020 digunakan untuk menghitung *SSW*. Persamaan yang dipakai untuk menghasilkan nilai *SSW* adalah

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

Tahun 2019:

$$SSW\ 1 = 1/3(11,373+11,373) = 7,582$$

$$SSW\ 2 = 1/3(3,452+0,33+1,627+2,125) = 2,511$$

$$SSW\ 3 = 1/3(0,261+1,159 + 0,649 + 0,959 + 1,025) = 1,351$$

Tahun 2020:

$$SSW\ 1 = 1/3(9,598+9,598) = 6,399$$

$$SSW\ 2 = 1/3(5,637+1,886+2,014+1,741) = 3,759$$

$$SSW_3 = 1/3(0,451+0,184 +1,114 +0,984 +0,559) = 1,097$$

2. Menghitung *SSB*

Data *centroid* baru cluster 1, 2, dan 3 hasil iterasi terakhir *clustering* Tahun 2019 dan Tahun 2020 digunakan untuk menghitung *SSB*. Persamaan yang dipakai buat memperoleh nilai *SSB*, yaitu

$$SSB_{i,j} = d(c_i, c_j)$$

Tahun 2019:

$$\begin{aligned} SSB_{1,2} = d(c_1, c_2) &= \sqrt{(X_{c1} - X_{c2})^2 + (Y_{c1} - Y_{c2})^2} \\ &= \sqrt{(21,906 - 2,144)^2 + (41,606 - 8,699)^2} \\ &= 38,4 \end{aligned}$$

$$\begin{aligned} SSB_{1,3} = d(c_1, c_3) &= \sqrt{(X_{c1} - X_{c3})^2 + (Y_{c1} - Y_{c3})^2} \\ &= \sqrt{(21,906 - 0,304)^2 + (41,606 - 1,417)^2} \\ &= 45,6 \end{aligned}$$

$$\begin{aligned} SSB_{2,3} = d(c_2, c_3) &= \sqrt{(X_{c2} - X_{c3})^2 + (Y_{c2} - Y_{c3})^2} \\ &= \sqrt{(2,144 - 0,304)^2 + (8,699 - 1,417)^2} \\ &= 7,5 \end{aligned}$$

Hasil matrik SSB

SSB	1	2	3
1	0,0	38,4	45,6
2	38,4	0,0	7,5
3	45,6	7,5	0,0

Tahun 2020:

$$\begin{aligned}SSB_{1,2} = d(c_1, c_2) &= \sqrt{(Xc_1 - Xc_2)^2 + (Yc_1 - Yc_2)^2} \\ &= \sqrt{(22,909 - 2,653)^2 + (40,973 - 9,549)^2} \\ &= 37,4\end{aligned}$$

$$\begin{aligned}SSB_{1,3} = d(c_1, c_3) &= \sqrt{(Xc_1 - Xc_3)^2 + (Yc_1 - Yc_3)^2} \\ &= \sqrt{(22,909 - 0,253)^2 + (40,973 - 1,128)^2} \\ &= 45,8\end{aligned}$$

$$\begin{aligned}SSB_{2,3} = d(c_2, c_3) &= \sqrt{(Xc_2 - Xc_3)^2 + (Yc_2 - Yc_3)^2} \\ &= \sqrt{(2,653 - 0,253)^2 + (9,549 - 1,128)^2} \\ &= 8,8\end{aligned}$$

Hasil matrik SSB

SSB	1	2	3
1	0	37,4	45,8
2	37,4	0	8,8
3	45,8	8,8	0

3. Menghitung R

Data hasil perhitungan SSW dan hasil matrik perhitungan SSB digunakan untuk menghitung R. Rumus yang dipakai buat mendapatkan angka R

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Tahun 2019:

$$R_{1,2} = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = (7,582+2,511)/38,4 = 0,262950$$

$$R_{1,3} = \frac{SSW_1 + SSW_3}{SSB_{1,3}} = (7,582+1,351)/45,6 = 0,195784$$

$$R_{2,3} = \frac{SSW_2 + SSW_3}{SSB_{2,3}} = (2,511+1,351)/7,5 = 0,514233$$

Hasil rasio tertinggi (R max) dari tiap cluster adalah sebagai berikut.

R	1	2	3	R max
1	0	0,262950	0,195784	0,262950
2	0,262950	0	0,514233	0,514233
3	0,195784	0,514233	0	0,514233

Tahun 2020:

$$R_{1,2} = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = (6,399+3,759)/37,4 = 0,2717$$

$$R_{1,3} = \frac{SSW_1 + SSW_3}{SSB_{1,3}} = (6,399+1,097)/45,8 = 0,1635$$

$$R_{2,3} = \frac{SSW_2 + SSW_3}{SSB_{2,3}} = (3,759+1,097)/8,8 = 0,5546$$

Hasil rasio tertinggi (R max) dari tiap cluster adalah sebagai berikut.

R	1	2	3	R max
1	0	0,2717	0,1635	0,2717
2	0,2717	0	0,5546	0,5546
3	0,1635	0,5546	0	0,5546

4. Menghitung DBI

Data hasil perhitungan rasio tertinggi dari tiap cluster digunakan untuk menghitung DBI. Rumus yang dipakai untuk mendapatkan skor DBI adalah

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

Tahun 2019:

$$\begin{aligned} \text{DBI} &= 1/3(0,262950 + 0,514233 + 0,514233) \\ &= 0,430472 \end{aligned}$$

Tahun 2020:

$$\text{DBI} = 1/3(0,2717 + 0,5546 + 0,5546) = 0,4603$$

Karena nilai DBI mendekati nol, maka cluster sudah cukup baik.

BAB VI

PENUTUP

Berdasarkan hasil penelitian dan pembahasan, maka kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut.

Penerapan *K-Means clustering* membagi data tahun 2019 dan tahun 2020 berdasarkan tingkat ekspor sebanyak tiga klaster. Klaster pertama atau tingkat ekspor tinggi mempunyai dua anggota, yaitu Singapura dan Malaysia. Klaster kedua atau tingkat ekspor menengah memiliki empat anggota, yakni Jepang, Taiwan, Hongkong, dan Tiongkok. Sedangkan klaster ketiga atau tingkat ekspor rendah mempunyai lima anggota yang terdiri dari Amerika Serikat, Arab Saudi, Australia, Uni Emirat Arab, dan negara lainnya.

Evaluasi terhadap hasil klaster data menggunakan *Davis Bouldin Index (DBI)*. Karena, nilai *DBI* klaster data tahun 2019 sama dengan 0,430472, maka klaster sudah cukup baik. Demikian pula, nilai *DBI* klaster data tahun 2020 sama dengan 0,4603, maka klaster sudah cukup baik.

Berdasarkan kesimpulan di atas, ada beberapa saran dapat digunakan untuk penelitian selanjutnya, yaitu pengelompokan data ekspor ikan segar/dingin hasil tangkap menurut negara tujuan utama dengan k-means clustering dapat dilakukan untuk big data.

Penelitian selanjutnya bisa menggunakan algoritma clustering yang lain seperti *k-medoids*, *fuzzy c-means*, *hierarchy agglomerative clustering*, dan *DBSCAN (Density Based Spatial Clustering of Applications with Noise)*, serta membandingkan hasilnya dengan k-means clustering.

DAFTAR PUSTAKA

- Adinugroho, Sigit & Yuita Arum Sari. 2018. *Implementasi data mining menggunakan WEKA*. Malang: UB Press.
- Afifi, Wakhid, Dhiya'an Ramadhanty Nastiti, & Qurrotul Aini. 2020. Clustering K-Means Pada Data Ekspor (Studi Kasus: PT. Gaikindo). *Jurnal SIMETRIS*, Vol. 11 No. 1.
- Anjelita, Mawaddah, Agus Perdana Windarto, Anjar Wanto, & Saifullah. 2019. "Analisis metode k-means pada kasus ekspor barang perhiasan dan barang berharga berdasarkan negara tujuan". *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, Juli 2019, Hal. 476-482.
- Arora, Preeti, Deepali, & Shipra Varshney. 2016. Analysis of K-Means and K-Medoids algorithm for big data. *Procedia Computer Science*, 78, 507-512.
- Badan Pusat Statistik. 2021. Ekspor ikan segar/dingin hasil tangkap menurut negara tujuan utama 2012-2020. Jakarta.
- Barakbah, Ali Ridho & Yasushi Kiyoki. 2009. "A pillar algorithm for K-means optimization by distance

maximization for initial centroid designation”. *IEEE Symposium on Computational Intelligence and Data Mining*.

Bates, Andrew & Jugal Kalita. 2016. “Counting clusters in twitter posts”. *Proceedings of the 2nd International Conference on Information Technology for Competitive Strategies*, pp. 85.

Dhuhita, Windha Mega Pradnya. 2015. “Clustering menggunakan metode K-Means untuk menentukan status gizi balita”. *Jurnal Informatika*, Vol. 15, No. 2.

Gaghana, Krizia Gidion. 2022. *Visualisasi clustering data ekspor kopi menurut negara tujuan menggunakan algoritma K-means dan K-medoids*. Universitas Multimedia Nusantara.

Iskandar, Bobby, S. Saifullah, Eka Wirawan, Irfan Sudahri Damanik, & Ilham Syahputra Saragih. 2020. Analisa Perbandingan Pengelompokan Ekspor dan Impor di Indonesia Berdasarkan Bulan menggunakan K-Means. *Prosiding Seminar Nasional Riset dan Information Science (SENARIS) 2020 Vol. 2*, pp. 457-467.

Kementerian Kelautan dan Perikanan. 2021. Peringkat Indonesia Sebagai Eksportir Produk Perikanan Dunia Meningkat di Masa Pandemi. Diakses dari

<https://kkp.go.id/djpdspkp/artikel/33334-peringkat-indonesia-sebagai-eksportir-produk-perikanan-dunia-meningkat-di-masa-pandemi>
tanggal 07/12/2021

Kusnawi. 2022. Evaluasi cluster dengan Davies Bouldin Index (DBI). [Online]. Available. <https://www.youtube.com/watch?v=3UTLjoWSID4&t=5s> [Accessed: 27 Apr 2022].

Nawrin, Sadia, Md R Rahman, Shamim Akhter. 2017. "Exploring K-Means with internal validity indexes for data clustering in traffic management system". *International Journal of Advanced Computer Science and Applications* 8 (3): 264-272.

Pandin, Catherin & Fahrudin. 2019. "Application of data mining for Indonesian products export in South Korea using clustering: Indonesia Trade Promotion Center Busan". *IOP Conf. Series: Materials Science and Engineering* 662 022053.

Sitompul, Bernard Jumadi Dehotman. 2018. *Peningkatan hasil evaluasi clustering Davies-Bouldin Index dengan penentuan titik pusat cluster awal algoritma K-means*. Prodi S2 Teknik Informatika, Fakultas Ilmu Komputer dan Teknologi Informasi, USU, Medan.

- Wani, M. A. & Romana Riyaz. 2017. "A novel point density based validity index for clustering gene expression datasets". *International Journal of Data Mining and Bioinformatics* 17(1): 66–84.
- Windarto, Agus Perdana. 2017. "Penerapan data mining pada ekspor buah-buahan menurut negara tujuan menggunakan K-Means clustering". *Techno.Com: Jurnal Teknologi Informasi*, Vol. 16, No. 4, November 2017 : 348-357.

BIOGRAFI PENULIS

Ekka Pujo Ariesanto Akhmad, S.E., M. Kom.

Ekka Pujo Ariesanto Akhmad lahir di Surabaya 24 Maret 1974, lulus dari SMAN I Surabaya pada tahun 1993, melanjutkan ke Jurusan Manajemen konsentrasi Manajemen Keuangan di Sekolah Tinggi Ilmu Ekonomi Indonesia (STIESIA) Surabaya. Studi S2 diselesaikan di Pasca Sarjana Sekolah Tinggi Teknik Surabaya (STTS), program studi Teknologi Informasi. Sejak tahun 2011 menjadi dosen tetap di program studi Manajemen Pelabuhan dan Logistik Maritim, Fakultas Vokasi Pelayaran, Universitas Hang Tuah. Saat ini aktif meneliti di bidang sistem informasi, pembelajaran berbantuan komputer, dan machine learning.

Budi Priyono, S.Sos., M.M.

Budi Priyono lahir di Pacitan, 25 Maret 1959 lulus dari SMAN I Pacitan pada tahun 1978, melanjutkan ke Jurusan Ilmu Administrasi prodi Ilmu Administrasi Negara di Universitas Wijaya Putra, Surabaya. Studi S2 diselesaikan di Pasca Sarjana program studi Magister Manajemen Universitas Wijaya Putra, Surabaya. Sejak tahun 2017 menjadi dosen tetap di program studi Manajemen Pelabuhan dan Logistik Maritim, Fakultas Vokasi Pelayaran, Universitas Hang Tuah. Saat ini aktif meneliti di bidang manajemen SDM kepelabuhan, manajemen kontainer, dan manajemen bongkar muat dan penumpukkan.

Perkumpulan Rumah Cemerlang Indonesia
Pondok Karisma Residence
Jalan Raflesia VI D:151
Panglayungan, Cipedes Tasikmalaya - 085223186009
<http://rcpress.rcipublisher.org/>