

Pengelompokan Data pada Website Lelang Breeder Koi Berdasarkan Ukuran Menggunakan Metode K-Means Clustering

Nanda Fatkhur Rohman¹ Indyah Hartami Santi² Wahyu Dwi Puspitasari³

Fakultas Teknologi Informasi, Universitas Islam Balitar, Kota Blitar, Provinsi Jawa Timur, Indonesia^{1,2,3}

Email: nandafatkhur@gmail.com¹ indyahartamisanti@gmail.com²
pushpitasari23@gmail.com³

Abstrak

Website lelang Breeder Koi adalah salah satu yang mengalami masalah tersebut. *Website* lelang Breeder Koi merupakan salah satu *website* lelang ikan koi yang mulai aktif pada 2019. Permasalahan pada *website* Breeder Koi adalah kesulitan admin yang akan melakukan pemeriksaan data lelang dari ukuran ikan koi tertentu. Hal ini dimaksudkan agar saat proses penampungan di akuarium karantina sementara sebelum dikirim ke pembeli sesuai dengan ukuran ikan koi. Pada proses penampungan ikan sementara Breeder Koi masih melakukan pemilahan secara manual. Masalah tersebut semakin bertambah karena pada bulan November 2023 data lelang mengalami kenaikan. Dengan metode *k-means clustering*, dapat mengelompokkan data lelang ikan koi pada *website* Breeder Koi berdasarkan ukuran ikan koi yang di lelang dengan membagi menjadi 4 *cluster* yang menyesuaikan dengan ukuran akuarium karantina di tempat penampungan Breeder Koi yaitu 4 ukuran untuk ikan koi berukuran kecil, sedang, besar, dan sangat besar. Untuk pengujian pada penelitian ini menggunakan *Davies Bouldin Index* (DBI) yaitu dengan menghitung jarak kedekatan antar data dalam setiap *cluster*nya. Pada penelitian ini mengelompokkan 245 data lelang dengan centroid (titik pusat) akhir 18.06666667, 30.44262295, 34.87037037, dan 41.69411765 Dengan hasil, *cluster* 1 (kecil) sebanyak 45 data, *cluster* 2 (sedang) sebanyak 61 data, *cluster* 3 (besar) sebanyak 62 data, dan *cluster* 4 (sangat besar) 77 data. Pengujian *Davies Bouldin Index* pada penelitian ini mendekati angka 0 dengan nilai sebesar 0.5329667491112356 dimana angka ini sudah termasuk baik terhadap setiap *cluster* hasil dari proses *clustering*/pengelompokan data lelang ikan koi yang telah dilakukan. Berdasarkan hasil diatas, pengelompokan data lelang ikan koi menggunakan algoritma *k-means clustering* bisa dikatakan optimal.

Kata Kunci: Pengelompokan Data, Website Lelang, *K-means Clustering*



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Teknologi informasi pada era sekarang banyak dimanfaatkan di berbagai bidang. Salah satu yang sedang marak adalah penggunaan *website* untuk sarana promosi, jual beli, profil perusahaan, dan lain-lain. Seperti halnya pada bidang perikanan, *website* digunakan sebagai sarana jual beli dan lelang ikan. Akhir-akhir ini, ikan koi banyak diminati di Indonesia. Dalam dunia koi pun juga mengikuti kemajuan teknologi. Untuk tetap bertahan pada era ini diperlukan inovasi berupa pemasaran ikan koi melalui *website* (Nalendra dkk., 2022). Pastinya, dengan mengikuti perkembangan teknologi dunia koi pun juga akan terus diminati seiring berjalannya waktu. *Website* lelang ikan koi ini tentunya banyak sekali data ikan koi yang dilelang dan membuat admin *website* lelang kerepotan dengan data yang tidak teratur. *Website* lelang Breeder Koi adalah salah satu yang mengalami masalah tersebut. *Website* lelang Breeder Koi merupakan salah satu *website* lelang ikan koi terbesar di Indonesia dengan lebih dari 23.000 anggota didalamnya. *Website* ini, mulai aktif pada 2019. *Website* ini memberi sarana untuk penggemar dan penjual ikan koi untuk menawarkan dan membeli ikan koi dengan sistem lelang. Permasalahan pada *website* Breeder Koi adalah kesulitan admin yang akan melakukan pemeriksaan data lelang dari ukuran ikan koi tertentu, dikarenakan data tidak teratur karena

hanyaurut berdasarkan waktu lelangnya. Hal ini dimaksudkan agar saat proses penampungan di akuarium karantina sementara sebelum dikirim ke pembeli sesuai dengan ukuran ikan koi dalam satu hari tersebut. Karena jika tidak sesuai misal ikan koi kecil bercampur dengan ikan koi besar dalam satu akuarium ikan koi bisa stress dan bahkan menyebabkan kematian. Pada proses penampungan ikan sementara Breeder Koi masih melakukan pemilahan secara manual yaitu ketika ikan dari berbagai penjual yang akan dikirim ke pembeli terkumpul di penampungan Breeder Koi, ikan akan diukur menggunakan bak meteran yang kemudian dimasukkan ke akuarium sesuai dengan ukuran ikan tersebut. Cara tersebut kurang efisien dengan jumlah ikan yang banyak. Masalah tersebut semakin bertambah karena pada bulan Juni 2024 data lelang mengalami kenaikan sebesar 10% dengan total 5.798 data yang pada bulan sebelumnya yaitu Mei 2024 total ada 5.256 data. Berikut grafik data ikan koi yang dilelang pada Januari-Juni 2024 pada gambar 1 dibawah ini:



Gambar 1. Grafik Data Lelang

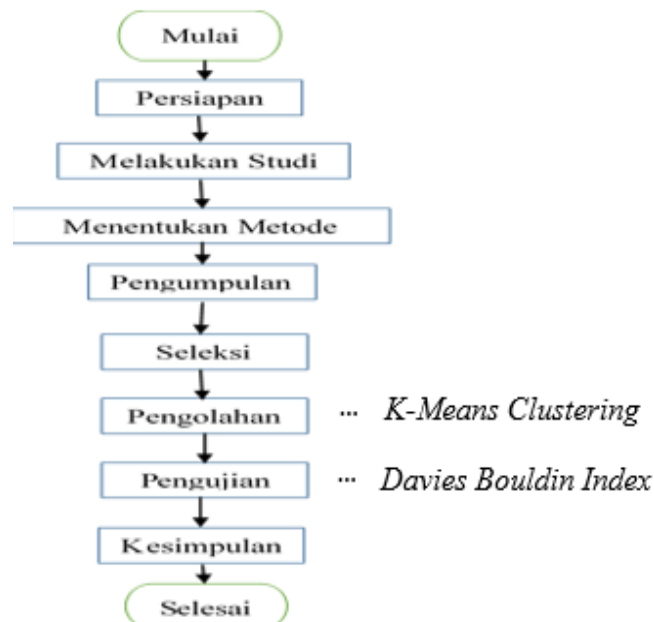
Karena masalah tersebut, diperlukan pengelompokan data lelang ikan koi yang ada untuk mempermudah admin dalam melakukan pencarian. Tentunya dibutuhkan sebuah algoritma untuk melakukan pengelompokan data, yang dimana akan menggunakan metode *clustering*. Dengan menggunakan algoritma *K-Means Clustering*, kalsifikasi data dilakukan dengan mengelompokkan data-data dengan tingkat kemiripan tinggi atau jarak yang berdekatan sehingga tercipta kelompok data yang karakteristiknya berbeda dengan kelompok data lainnya. Penelitian sebelumnya yang berjudul “Klasifikasikasi Jenis Ikan Koi Menggunakan Ekstraksi Warna HSV” oleh Hutagalung dan Aqshal pada tahun 2023, dilakukan klasifikasi untuk menentukan kelompok jenis ikan koi berdasarkan warna melalui gambar. Penulis menggunakan metode ekstraksi warna *Hue Saturation Value* (HSV) untuk tiga jenis ikan koi berbeda yaitu *Kohaku*, *Sanke*, dan *Showa*. Hasil dari penelitian ini adalah dengan total data 189 citra yang dibagi dua yaitu data latih 150 dan data uji 39 diperoleh nilai data latih sebesar 100%, dan nilai data uji 48,71 % (Hutagalung & Aqshal, 2023). Klasifikasi menggunakan metode HSV digunakan untuk menentukan jenis ikan koi yang ada pada gambar yang diinputkan. Dengan berdasarkan 3 jenis ikan koi yang sudah ditentukan sebelumnya, sistem akan menentukan masuk dalam jenis ikan koi manakah gambar yang diinputkan. Penelitian berjudul “Pengenalan Varietas Ikan Koi Berdasarkan Foto Menggunakan *Simple Linear Iterative Clustering Superpixel Segmentation* dan *Convolutional Neural*” oleh Hermawan, Zaeni, Wibawa, Gunawan, Kristian, dan Darmawan pada tahun 2021 menggunakan *simple linear iterative clustering* untuk segmentasi varietas jenis ikan koi menggunakan *simple linear iterative clustering superpixel segmentation* dan *convolutional neural network* (CNN) untuk menentukan jenis ikan koi berdasarkan foto yang diinput. *K-Means Clustering* bertujuan untuk mengurangi fungsi yang dibentuk selama proses pengelompokan, seminimal mungkin. Tujuan tersebut dilakukan dengan cara meminimalisasikan variasi data yang ada didalam *cluster* dan memaksimalkan variasi data yang ada di *cluster* lainnya (Pratiwi & Suwarna, 2023).

K-Means Clustering dapat melakukan pengelompokan dengan baik karena mampu mengolah banyak data dan pemrosesan dengan cepat tanpa bergantung pada besar kecilnya

jumlah data. Untuk pengujian pada penelitian ini menggunakan *Davies Bouldin Index* (DBI) yaitu dengan menghitung jarak kedekatan antar data dalam setiap clusternya. Sehingga penulis tertarik untuk mengimplementasikan algoritma *K-Means Clustering* dan pengujian DBI pada pengelompokan data lelang ikan koi dari *website* Breeder Koi. Pada penelitian ini, penulis menggunakan metode *k-means clustering* untuk mengelompokkan data lelang ikan koi pada *website* Breeder Koi berdasarkan ukuran ikan koi yang di lelang dengan membagi menjadi 4 *cluster* yang menyesuaikan dengan ukuran akuarium karantina di tempat penampungan Breeder Koi yaitu 4 ukuran untuk ikan koi berukuran kecil, sedang, besar, dan sangat besar dengan menggunakan pengujian *Davies Bouldin Index* untuk menguji kedekatan antar data dalam setiap clusternya yang diharapkan dapat memudahkan admin lelang *website* Breeder Koi dalam melakukan pengecekan data lelang ikan koi. Untuk itu penelitian dilaksanakan dengan judul “Pengelompokan Data Pada Website Lelang Breeder Koi Berdasarkan Ukuran Menggunakan Metode K-Means Clustering”.

METODE PENELITIAN

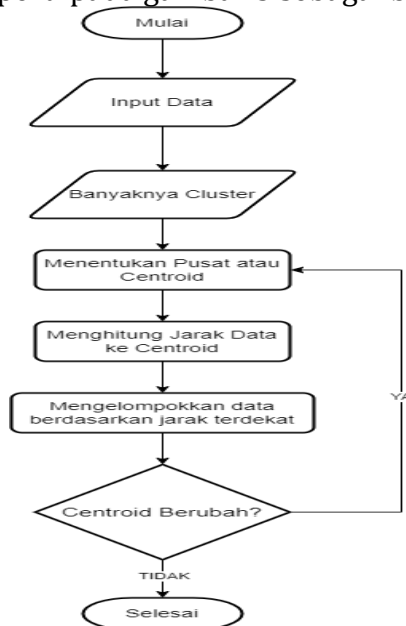
Penelitian ini dilakukan dari November 2023 hingga Juli 2024. Pengambilan data pada penelitian ini dilakukan di *website* Breeder Koi dengan alamat [link breederkoi.com](http://breederkoi.com) dan pengolahan data dilakukan di laboratorium komputer fakultas teknologi dan informasi Universitas Islam Balitar Blitar. Penelitian dengan judul “Pengelompokan Data Pada Website Lelang Breeder Koi Berdasarkan Ukuran Menggunakan Metode K-Means Clustering” ini menggunakan jenis penelitian deskriptif. Teknik pengumpulan data dalam penelitian ini dilakukan dengan cara wawancara dan studi literatur. Adapun tahapan-tahapan penelitian berdasarkan latar belakang, rumusan masalah, tujuan penelitian, dan studi literatur pada gambar 2 sebagai berikut:



Gambar 2. Tahapan Penelitian

1. Persiapan. Tahap ini merupakan tahap melakukan identifikasi masalah terkait *website* Breeder Koi. Menyusun rencana penelitian dan batasan masalah.
2. Melakukan Studi Pustaka. Di dalam studi pustaka dilakukan pemahaman tentang algoritma *K-Means Clustering* pada penelitian yang dahulu pernah dilakukan dengan cara mengumpulkan teori dari buku, jurnal artikel yang membahas tentang *Ikan Koi* dan *K-Means Clustering*.

3. Menentukan Metode Data Mining. Langkah selanjutnya adalah menentukan metode yang sesuai untuk digunakan dalam teknik pengelompokan, setelah melakukan pengamatan peneliti memilih menggunakan algoritma *K-Means Clustering*.
4. Pengumpulan Data. Pada tahap ini peneliti memperoleh data lelang ikan koi pada bulan Juni 2024 yang dikirim oleh admin *website* dari arsip Breeder Koi.
5. Seleksi Data. Data yang diperoleh perlu diolah untuk digunakan sesuai dengan yang dibutuhkan. Pada penelitian ini, dari data bulan Juni diambil data satu hari yang paling tinggi jumlah datanya yaitu pada 25 Juni 2024.
6. Pengolahan Data. Tahap pengolahan data yang sudah menjadi *dataset* akan diolah dengan menggunakan Algoritma *K-Means Clustering*, yaitu dengan menghitung jarak antar data berdasarkan nilai *centroid* yang ditentukan. Berikut alur diagram *flowchart K-Means Clustering* yang ditunjukkan seperti pada gambar 3 sebagai berikut:



Gambar 3. Flowchart Algoritma K-Means Clustering

7. Pengujian Data. Data hasil dari pengelompokan menggunakan *K-Means* akan diuji menggunakan metode *Davies Bouldin Index* (DBI), yaitu dengan menghitung jarak antar data dalam satu cluster. Alur tahapan pengujian *Davies Bouldin Index* dalam bagan *flowchart* pada gambar 4 berikut:



Gambar 4. Flowchart Pengujian DBI

8. Kesimpulan. Setelah diketahui hasil dari pengujian maka dapat ditarik kesimpulan yang mengacu pada rumusan masalah dan tujuan penelitian. Dan saran yang digunakan untuk mengembangkan penelitian selanjutnya serta, sebagai bahan untuk meningkatkan kualitas penelitian.

HASIL PENELITIAN DAN PEMBAHASAN

Pengumpulan Data

Data yang digunakan disini adalah data sekunder yang diperoleh dari Website Lelang Breeder Koi berupa Pelelang, Jenis Koi, dan Ukuran. Dari hasil pengumpulan data diperoleh data bulan Juni 2024 sejumlah 5798 yang ditampilkan pada tabel 1 berikut:

Tabel 1. Data Lelang Satu Bulan

No	Pelelang	Jenis Koi	Ukuran (cm)
1	Sambalado Koi	Hi Utsuri	35
2	Sambalado Koi	Shiro Utsuri	38
3	Sambalado Koi	Showa	39
4	Sambalado Koi	Showa	41
5	Roziquil Fahmi	Chagoi	19
...			
51	Master Koi	Showa	25
52	Master Koi	Sanke	45
53	Master Koi	Kohaku	35
...			
293	Minan Koi	Showa	21
294	Minan Koi	Goshiki	23
295	Minan Koi	Shusui	34
...			
5796	Basuki Koi	Goshiki	33
5797	Basuki Koi	Soragoi	46
5798	Basuki Koi	Platinum Kumpay	45

Tabel 1 diatas adalah data lelang pada bulan Juni 2024 yang berjumlah 5798. Tabel yang berwarna hijau adalah data lelang pada tanggal 25 Juni 2024 yang akan digunakan pada penelitian ini.

Seleksi Data

Berdasarkan tabel 1 diatas, diambil data yang paling tinggi jumlah datanya dalam satu hari di bulan Juni 2024 karena dalam proses pemilahan penampungannya adalah ikan koi yang dilelang dalam satu hari tersebut yaitu pada tanggal 25 Juni 2024. Adapun data lelang pada 25 Juni 2024 dapat dilihat pada tabel 2 berikut:

Tabel 2. Data Lelang

No	Pelelang	Jenis Koi	Ukuran (cm)
1	Master Koi	Showa	25
2	Master Koi	Sanke	45
3	Master Koi	Kohaku	35
4	Mas Bren	Goshiki	23
5	Mas Bren	Goshiki	25
6	Mas Bren	Hi Utsuri	37
7	Mas Bren	Shiro Utsuri	40

8	Mas Bren	Showa	21
9	Risky Koi Blitar	Showa	17
10	Risky Koi Blitar	Kohaku	25
...			
236	Fatro	Karasu Kumpay	41
237	Sahrul Koi Farm	Goshiki	40
238	Sahrul Koi Farm	Hi Utsuri	35
239	Sahrul Koi Farm	Shiro Utsuri	38
240	Sahrul Koi Farm	Showa	39
241	Sahrul Koi Farm	Shiro Utsuri	41
242	Sahrul Koi Farm	Showa	15
243	Minan Koi	Showa	21
244	Minan Koi	Goshiki	23
245	Minan Koi	Shusui	34

Perhitungan *K-Means Clustering*

1. Input Data. Tahap ini memasukkan data yang digunakan untuk proses perhitungan menggunakan metode *K-Means Clustering*. Data yang digunakan dapat dilihat pada tabel 2 yang sudah disajikan diatas.
2. Banyaknya *Cluster*. Penentuan *cluster* berfungsi untuk membagi data yang ada. Berdasarkan pada tabel 2 yang sudah disajikan, jumlah *cluster* yang digunakan adalah 4 *cluster* yaitu kecil, sedang, besar, dan sangat besar.
3. Menentukan *Centroid*. *Centroid*/titik pusat digunakan untuk acuan pembagian *cluster*. Berikut *centroid* setiap *cluster* yang ditentukan secara acak disajikan pada tabel 3 dibawah ini:

Tabel 3. Centroid Cluster

Centroid 1	23.70114943
Centroid 2	31
Centroid 3	33.02
Centroid 4	40.59259259

4. Menghitung Jarak Data ke Centroid. Tahap ini setiap data akan dihitung jaraknya dengan *centroid* yang sudah ditentukan dengan *euclidean distance* menggunakan rumus pada persamaan dengan hasil sebagai berikut:

$$d = \sqrt{(25-23.70114943)^2} = 1.29885057$$

$$d = \sqrt{(45-23.70114943)^2} = 21.29885057$$

$$d = \sqrt{(35-23.70114943)^2} = 11.29885057$$

$$d = \sqrt{(23-23.70114943)^2} = 0.70114943$$

$$d = \sqrt{(25-23.70114943)^2} = 1.29885057$$

Adapun hasil lengkap perhitungan jarak menggunakan ditampilkan pada tabel 4 sebagai berikut:

Tabel 4. Perhitungan Jarak

No	Ukuran	Jarak C1	Jarak C2	Jarak C3	Jarak C4
1	25	1.29885057	6	8.02	15.59259
2	45	21.29885057	14	11.98	4.407407
3	35	11.29885057	4	1.98	5.592593
4	23	0.70114943	8	10.02	17.59259

5	25	1.29885057	6	8.02	15.59259
...					
241	41	17.29885057	10	7.98	0.407407
242	15	8.70114943	16	18.02	25.59259
243	21	2.70114943	10	12.02	19.59259
244	23	0.70114943	8	10.02	17.59259
245	34	10.29885057	3	0.98	6.592593

5. Mengelompokkan Data Berdasarkan Jarak Terdekat. Berdasarkan hasil perhitungan jarak, data dapat dikelompokkan berdasarkan jarak terdekatnya (jarak *minimum*). Hasil pengelompokan terdapat pada tabel 5 dibawah ini:

Tabel 5. Hasil Clustering

No	Ukuran	Jarak Minimum	Cluster
1	25	1.298851	1
2	45	4.407407	4
3	35	1.98	3
4	23	0.701149	1
5	25	1.298851	1
...			
241	41	0.407407	4
242	15	8.701149	1
243	21	2.701149	1
244	23	0.701149	1
245	34	0.98	3

Perhitungan dan hasil hitung manual diatas diterapkan menggunakan *Google Colab* dengan bahasa pemrograman *Python* dengan perintah:

```
X = df [['Ukuran (cm)']]
kmeans = KMeans(n_clusters=4, random_state=0)
kmeans.fit(X)
df ['Cluster'] = kmeans.fit_predict(X)
```

Adapun hasil *running Google Colabs* dari perintah program diatas dapat dilihat pada gambar 5 berikut:

	Ukuran (cm)	Cluster
0	25	0
1	45	3
2	35	2
3	23	0
4	25	0
..	...	...
240	41	3
241	15	0
242	21	0
243	23	0
244	34	2

Gambar 5. Hasil Perhitungan Clustering

6. Centroid Berubah/Iterasi Berulang. Tahap ini iterasi berulang dilakukan dengan menentukan *centroid* baru (C_k) dengan rumus pada persamaan sebagai berikut:

$$\frac{1}{nk1} \sum d1 = \frac{nk}{d1} = \frac{913}{49} = 18,63265$$

$$\frac{1}{nk2} \sum d2 = \frac{nk}{d2} = \frac{1726}{56} = 30,82456$$

Persamaan diatas nk adalah hasil penjumlahan seluruh data dalam satu *cluster* dan d adalah jumlah data dalam satu *cluster*. Adapun nilai lengkap dari *centroid* baru dengan detail pada tabel 6 sebagai berikut:

Tabel 6. Centroid Iterasi Kedua

Centroid 1	18.63265
Centroid 2	30.82456
Centroid 3	34.2619
Centroid 4	41.1134

Seperti sebelumnya, setelah menentukan *centroid* baru dilakukan penghitungan jarak menggunakan *euclidean distance* dengan rumus yang sama pada iterasi pertama:

$$d = \sqrt{(25-18.63265)^2} = 6.367347$$

$$d = \sqrt{(45-18.63265)^2} = 26.36735$$

$$d = \sqrt{(35-18.63265)^2} = 26.36735$$

$$d = \sqrt{(23-18.63265)^2} = 4.367347$$

$$d = \sqrt{(25-18.63265)^2} = 6.367347$$

Adapun hasil lengkap penghitungan jarak yang disajikan pada tabel 7 sebagai berikut:

Tabel 7. Hasil Penghitungan Jarak Iterasi Kedua

No	Ukuran	C1	C2	C3	C4
1	25	6.367347	5.824561	9.261905	16.1134
2	45	26.36735	14.17544	10.7381	3.886598
3	35	16.36735	4.175439	0.738095	6.113402
4	23	4.367347	7.824561	11.2619	18.1134
5	25	6.367347	5.824561	9.261905	16.1134
...					
241	41	22.36735	10.17544	6.738095	0.113402
242	15	3.632653	15.82456	19.2619	26.1134
243	21	2.367347	9.824561	13.2619	20.1134
244	23	4.367347	7.824561	11.2619	18.1134
245	34	15.36735	3.175439	0.261905	7.113402

Hasil dari pengelompokan data pada iterasi kedua terdapat pada tabel 4.8 dibawah ini:

Tabel 8. Hasil Pengelompoksn Data Iterasi Kedua

No	Ukuran	Cluster Lama	Cluster Baru	Keterangan
1	25	1	2	berubah
2	45	4	4	sama
3	35	3	3	sama
4	23	1	1	sama
5	25	1	2	berubah
...				
241	41	4	4	sama
242	15	1	1	sama
243	21	1	1	sama
244	23	1	1	sama
245	34	3	3	sama

Adapun perintah program *clustering* menggunakan *Google Colab* iterasi kedua dengan *centroid* baru:

```
for i in range(max_iter):
distances=np.sqrt (((data.values centroids[:,np.newaxis])**2). sum(axis=2))
labels = np.argmin(distances, axis=0)
```

```

new_centroids = np.array([data.values[labels == j].mean(axis=0) if np.any(labels == j) else
centroids[j] for j in range(n_clusters)])
sort_idx = np.argsort(new_centroids, axis=0).flatten()
sorted_centroids = new_centroids[sort_idx]
relabeled_clusters = np.zeros_like(labels)
for new_label, old_label in enumerate(sort_idx):
relabeled_clusters[labels == old_label] = new_label

print(f"Iteration {i+1}")
print(data.assign(Cluster=relabeled_clusters))
print("Centroids:", sorted_centroids, "\n")
if np.all(np.abs(new_centroids - centroids) < tol):
centroids = sorted_centroids
labels = relabeled_clusters
break
centroids = sorted_centroids
labels = relabeled_clusters
return labels, centroids, i+1

```

Adapun *centroid* baru pada iterasi ketiga ditampilkan pada tabel 9 sebagai berikut:

Tabel 9. Centroid Iterasi Ketiga

Centroid 1	18.06666667
Centroid 2	30.44262295
Centroid 3	34.87037037
Centroid 4	41.69411765

Perhitungan jarak iterasi ketiga dengan persamaan seperti dibawah ini:

$$\begin{aligned}
 d &= \sqrt{(25-18.06666667)^2} = 6.933333 \\
 d &= \sqrt{(45-18.06666667)^2} = 26.933333 \\
 d &= \sqrt{(35-18.06666667)^2} = 16.933333 \\
 d &= \sqrt{(38-18.06666667)^2} = 19.933333 \\
 d &= \sqrt{(39-18.06666667)^2} = 20.933333
 \end{aligned}$$

Hasil dari pengelompokan data iterasi ketiga terdapat pada tabel 10 dibawah ini:

Tabel 10. Hasil Pengelompoksn Data Iterasi Ketiga

No	Ukuran	Cluster Lama	Cluster Baru	Keterangan
1	25	2	2	sama
2	45	4	4	sama
3	35	3	3	sama
4	23	1	1	sama
5	25	2	2	sama
...				
239	38	3	4	berubah
240	39	4	4	sama
241	41	4	4	sama
242	15	1	1	sama
243	21	1	1	sama
244	23	1	1	sama
245	34	3	3	sama

Terdapat perubahan data pada *cluster* di iterasi ketiga, sehingga dilakukan iterasi keempat dengan menentukan centroid baru lagi menggunakan persamaan:

$$\frac{1}{nk1} \sum d1 = \frac{nk}{d1} = \frac{813}{45} = 18.06667$$

$$\frac{1}{nk2} \sum d2 = \frac{nk}{d2} = \frac{1857}{61} = 30.44262$$

Adapun *centroid* baru pada iterasi keempat yang ditampilkan pada tabel 11 sebagai berikut dibawah ini:

Tabel 11. Centroid Iterasi Keempat

Centroid 1	18.06667
Centroid 2	30.44262
Centroid 3	35.27419
Centroid 4	42.07792

Hasil dari pengelompokan data iterasi keempat terdapat pada tabel 12 dibawah ini:

Tabel 12. Hasil Pengelompoksn Data Iterasi Keempat

No	Ukuran	Cluster Lama	Cluster Baru	Keterangan
1	25	2	2	sama
2	45	4	4	sama
3	35	3	3	sama
4	23	1	1	sama
5	25	2	2	sama
...				
241	41	4	4	sama
242	15	1	1	sama
243	21	1	1	sama
244	23	1	1	sama
245	34	3	3	sama

Adapun hasil *clustering* menggunakan *Google Colab* iterasi ketiga dengan *centroid* baru pada gambar 6 dibawah ini:

	Ukuran (cm)	Cluster
0	25	1
1	45	3
2	35	2
3	23	0
4	25	1
..	...	...
240	41	3
241	15	0
242	21	0
243	23	0
244	34	2

Gambar 6. Perhitungan Iterasi Keempat

Data *clustering* pada iterasi keempat sama dengan iterasi ketiga, sehingga iterasi berulang berhenti. Berikut hasil data setiap cluster dari perhitungan diatas pada tabel 13 dibawah ini:

Tabel 13. Hasil Jumlah Data

Jumlah Data Tiap Cluster	
Cluster	Jumlah Data
1	45
2	61
3	62
4	77

Pengujian *Davies Bouldin Index*

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

Pada tahap ini menghitung nilai SSW yaitu rata-rata jarak antar data pada setiap *cluster* dengan rumus pada persamaan:

$$(4.9333+2.9333+1.0666+\dots+4.9333):45 = 2.340741$$

$$(5.4426+5.4426+5.4426+\dots+1.5573):61 = 1.404372$$

$$(0.2741+1.7258+2.2741+\dots+1.2741):62 = 1.629032$$

$$(2.9220+2.0779+3.9220+\dots+1.0779):77 = 2.28504$$

Adapun hasil dari SSW (*Sum of Square Within-cluster*) ditampilkan pada tabel 14. sebagai berikut dibawah ini:

Tabel 14. Hasil Nilai SSW

Cluster	Nilai SSW
1	2.340741
2	1.404372
3	1.629032
4	2.28504

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

Pada tahap ini menghitung nilai SSB yaitu jarak antar *centroid*/pusat data pada *cluster* dengan rumus:

$$c2-c1= 30.44262-18.06667= 12.37596$$

$$c3-c1= 35.27419-18.06667= 17.20753$$

$$c3-c1= 42.07792-18.06667= 24.01126$$

Adapun hasil dari nilai SSB (*Sum of Square Between-cluster*) disajikan pada tabel berikut 15 dibawah ini:

Tabel 15. Hasil Nilai SSB

SSB	1	2	3	4
1	0	12.37596	17.20753	24.01126
2	12.37596	0	4.831571	11.6353
3	17.20753	4.831571	0	6.803729
4	24.01126	11.6353	6.803729	0

3. Menentukan nilai Rasio

Pada tahap ini menghitung nilai rasio yaitu hasil bagi SSW dan SSB dengan rumus:

$$(2.34074+1.40437):12.37596 = 0.302612$$

$$(2.34074+1.62903):17.20753 = 0.2307$$

$$(2.34074+2.28504):24.01126 = 0.192651$$

Adapun hasil lengkap dari nilai *rasio* dan nilai *rasio maksimum* pada tabel 16 sebagai berikut:

Tabel 16. Hasil Nilai Rasio

Rasio	1	2	3	4	Rasio Maksimum
1	0	0.302612	0.2307	0.192651	0.302612

2	0.302612	0	0.62783	0.317088	0.62783
3	0.2307	0.62783	0	0.575283	0.62783
4	0.192651	0.317088	0.575283	0	0.575283

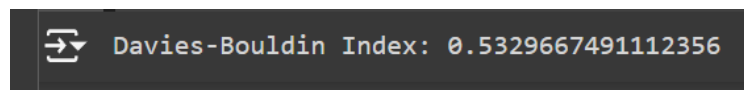
4. Menentukan nilai DBI (*Davies Bouldin Index*)

Pada tahap ini menghitung nilai DBI yang diperoleh dari rata-rata *rasio maksimum* dengan menggunakan rumus sebagai berikut dibawah ini:

$$DBI = \frac{1}{4} (0.302612 + 0.62783 + 0.62783 + 0.575283)$$

$$= 0.5329667491112356$$

Dengan rumus diatas hasil nilai DBI (*Davies Bouldin Index*) yaitu sebesar 0.5329667491112356. Berikut hasil penghitungan DBI menggunakan *Python* yang ditampilkan pada gambar 7 berikut:



Gambar 7. Hasil *Davies Bouldin Index*

Pembahasan

Proses penghitungan diatas membagi data lelang ikan koi menjadi 4 kelompok yaitu kecil, sedang, besar, dan sangat besar. Setelah menentukan jumlah *cluster*, proses selanjutnya menentukan centroid setiap *cluster*. Penentuan *centroid* ini dilakukan secara acak namun berurutan dari yang kecil ke yang besar. Centroid yang ditentukan bisa dilihat pada tabel 3 Berikutnya, adalah penghitungan jarak antara data dengan *centroid* yang sudah ditentukan sebelumnya. Penghitungan jarak ini bertujuan untuk memaksimalkan kemiripan. Kelompok objek dibentuk sehingga objek dalam kelompok memiliki kesamaan yang tinggi (Sianturi, dkk., 2019). Sehingga terbentuk *cluster* yang datanya memiliki tingkat kemiripan yang tinggi. Berikutnya, adalah mengelompokkan data berdasarkan jaraknya dengan centroid yang sudah ditentukan. Tahap selanjutnya adalah iterasi berulang untuk memastikan *cluster*. Yaitu mengulangi tahap penghitungan dari menentukan *centroid* baru hingga pengelompokannya. Karena data *cluster* masih ada perubahan, maka dilakukan iterasi ulang hingga 4 iterasi dan akhirnya data *cluster* yang dihasilkan tidak berubah/sama. Pada iterasi kedua, ketiga, dan keempat *centroid* baru dihasilkan dari persamaan dengan hasil *centroid* yang bisa dilihat pada tabel 6, tabel 9, dan tabel 11. Hasil akhir *clustering* untuk data lelang Breeder Koi ini yaitu *cluster* 1 (kecil) sebanyak 45 data, *cluster* 2 (sedang) sebanyak 61 data, *cluster* 3 (besar) sebanyak 62 data, dan *cluster* 4 (sangat besar) 77 data.

Hasil penelitian ini selaras dengan penelitian sebelumnya yaitu penelitian oleh Hutagalung, Syahputra, dan Tanjung pada tahun 2022. Penelitian ini berjudul “Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma *K-Means Clustering*”. Penelitian ini menerapkan *k-means clustering* dengan sampel 120 data dari nilai rapor siswa SMK Raksana 2 Medan, untuk pengelompokan siswa kelas unggulan berdasarkan penguasaan pemrograman yang dikuasai oleh siswa tersebut dengan menginputkan nilai *centroid* dan mencari nilai terdekat. Menghasilkan 3 *cluster* yaitu *cluster* 1 penguasaan pemrograman *android* sebanyak 10 siswa, *cluster* 2 penguasaan pemrograman *web* sebanyak 62 siswa dan *cluster* 3 penguasaan pemrograman *desktop* sebanyak 48 siswa. Tetapi pada penelitian tersebut tidak dilakukan pengujian, sehingga hasil *clustering* belum bisa dipastikan keberhasilannya. Pengujian *cluster* menggunakan metode *davies bouldin index* yaitu menghitung jarak data antar *cluster* yang ada. Seperti pada penelitian oleh Gantara dan Ali pada tahun 2023, semakin rendah nilai DBI, semakin baik klasterisasi yang dihasilkan. Nilai k yang paling optimal dapat dicari dengan

mencari nilai DBI yang mendekati 0. Pada penelitian ini, nilai DBI dari penghitungan yang dilakukan adalah sebesar 0.5329667491112356 dimana angka ini sudah termasuk baik terhadap setiap *cluster* hasil *clustering*/pengelompokan data lelang ikan koi yang telah dilakukan. Berdasarkan hasil diatas, pengelompokan data lelang ikan koi menggunakan algoritma *k-means clustering* bisa dikatakan optimal.

KESIMPULAN

1. Pengelompokan data lelang ikan koi di Breeder Koi dengan menggunakan metode *k-means clustering* dibagi menjadi 4 cluster. Dilakukan 4 iterasi dengan centroid (titik pusat) akhir 18.06666667, 30.44262295, 34.87037037, dan 41.69411765. Hasil dari pengelompokannya adalah *cluster* 1 (kecil) sebanyak 45 data, *cluster* 2 (sedang) sebanyak 61 data, *cluster* 3 (besar) sebanyak 62 data, dan *cluster* 4 (sangat besar) 77 data yang dapat dilihat pada tabel 4.15.
2. Pengujian *davies bouldin index* pada penelitian ini mendekati angka 0 dengan nilai sebesar 0.5329667491112356 dimana angka ini sudah termasuk baik terhadap setiap *cluster* hasil dari proses *clustering*/pengelompokan data lelang ikan koi yang telah dilakukan. Berdasarkan hasil diatas, pengelompokan data lelang ikan koi menggunakan algoritma *k-means clustering* bisa dikatakan optimal.

DAFTAR PUSTAKA

- Fhadli, M., Tempola, F. (2020). Data Mining Dengan Python Untuk Pemula. (n.d.). (n.p.): SPASI MEDIA.
- Gantara, N. P., & Ali, I. (2023). Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station. E-Link: Jurnal Teknik Elektro dan Informatika, 18(1), 12. doi:10.30587/e-link.v18i1.5339
- Hutagalung, D. D., & Aqshal, G. M. (2023). Klasifikasi Jenis Ikan Koi Menggunakan Ekstraksi Warna HSV dan Metode Adaptive Neuro-Fuzzy Inference System (ANFIS). Jurnal ICT: Information Communication & Technology, 23(2), 379-384
- Nalendra, A. K., Fuad, M. N., Mujiono, M., Wahyudi, D., & Utomo, P. B. (2022). Pelatihan Pembuatan Website Profile untuk Peternak Ikan Koi Kota Blitar Berbasis Wordpress. Archive: Jurnal Pengabdian Kepada Masyarakat, 1(2), 8. doi:10.55506/arch.v1i2.38
- Pratiwi, Y., & Suwarna, N. (2023). Penerapan Metode K-Means Clustering Pada Populasi Ayam Ras Petelur Berdasarkan Provinsi. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1). doi:10.36040/jati.v7i1.6457
- Sianturi, F. A., Hasugian, P. M., Simangunsong, A., & Nadeak, B. (2019). DATA MINING: Teori dan Aplikasi Weka (Vol. 1). IOCS Publisher.
- Yuhefizar. (2009). CMM Website Interaktif MCMS Joomla (CMS). (n.d.). (n.p.): Elex Media Komputindo.