



**Tugas**  
*Task*

Mata Kuliah / Course Name	: <b>Data Science</b>
Sesi Pertemuan/ Session Number	: <b>IX</b>
Judul Materi / Title	: <b>Agglomerative Clustering</b>
Nama Dosen / Lecturer	: <b>Alun Sujjada, S.Kom, M.T</b>
Jenis Tugas / Type of task	: <b>Case Based / Project Based</b>
NIM / Student ID Number	: 20230040291.....
Nama Mahasiswa/ Student Name	: Muhammad Ibdaul Mutafakkir.....
Kelas/ Class	: TI23C.....

**Answer:**

**LAPORAN AGGLOMERATIVE CLUSTERING (ANIMAL BEHAVIOR)**

**1. Analisis Unsupervised Learning**

Penelitian ini menggunakan metode Agglomerative Clustering untuk melakukan pengelompokan (unsupervised learning) pada dataset Animal Behavior. Dataset ini berisi karakteristik dasar perilaku dan fisiologis hewan, yaitu:

- AnimalID
- Species
- AvgDailyMovement\_km
- HeartRate\_bpm
- ActivityType
- HabitatType
- BodyTemp\_C
- AggressionLevel

Tujuan analisis adalah mengidentifikasi pola perilaku hewan berdasarkan kombinasi pergerakan, aktivitas, temperatur tubuh, dan tingkat agresi.

**2. Data Cleaning**

**a. Cek Missing Values**

Pengecekan menggunakan `df.isnull().sum()` menunjukkan bahwa tidak terdapat missing values pada seluruh kolom dataset.

Dataset dapat langsung diproses tanpa imputasi.

**b. Normalisasi / Standarisasi**



Untuk memastikan seluruh fitur berada dalam skala yang sebanding, dilakukan standarisasi menggunakan StandardScaler pada fitur numerik berikut:

- AvgDailyMovement\_km
- HeartRate\_bpm
- BodyTemp\_C
- AggressionLevel

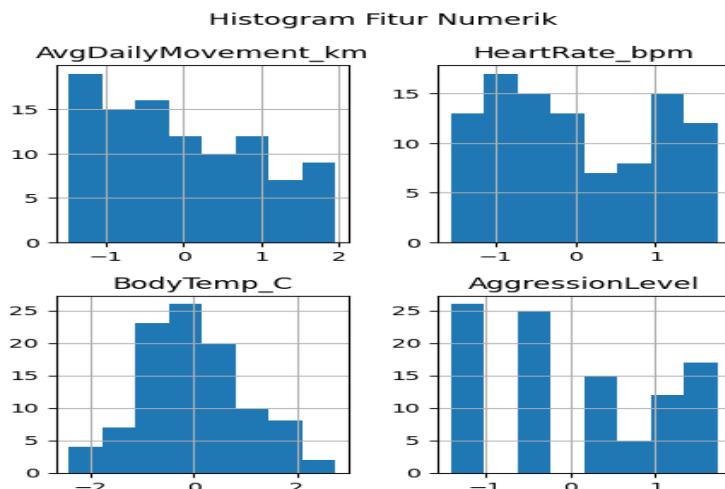
Hasil ringkas scaling:

- Mean  $\approx 0$
- Standar deviasi  $\approx 1$

Fitur kategorikal (Species, ActivityType, HabitatType) diubah menggunakan LabelEncoder agar dapat digunakan dalam proses clustering.

### 3. Analisis Deskriptif & Visualisasi Awal

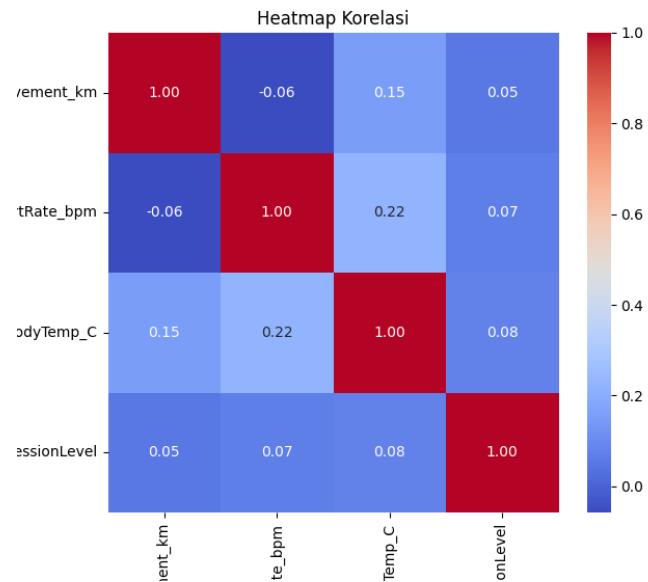
#### a. Histogram Fitur Numerik



Histogram menunjukkan:

- BodyTemp\_C berdistribusi mendekati normal.
- AvgDailyMovement\_km dan HeartRate\_bpm memiliki rentang yang cukup lebar, mengindikasikan keberagaman pergerakan dan metabolisme.
- AggressionLevel memiliki pola multimodal, mendukung asumsi adanya beberapa kelompok perilaku berbeda.

## b. Heatmap Korelasi



Hasil heatmap menunjukkan:

- Tidak ada korelasi kuat antar fitur (umumnya  $< 0.3$ ).
- Ini menandakan setiap fitur memberikan informasi yang unik dan bermanfaat untuk clustering multidimensional.

## 4. Model Agglomerative Clustering

Empat metode linkage diuji:

1. Single Linkage
2. Complete Linkage
3. Average Linkage
4. Ward Linkage

Setiap metode diuji untuk beberapa jumlah cluster ( $k = 2-5$ ), dan model dilakukan fitting serta prediksi cluster.

### Hasil Umum per Linkage:

#### 1) Single Linkage

- Cenderung mengalami *chaining* (cluster memanjang).
- Cluster kurang kompak.
- Kualitas cluster lebih rendah dibanding metode lain.

#### 2) Complete Linkage

- Cluster lebih kompak dibanding single.

- Namun cenderung memecah data menjadi kelompok-kelompok kecil.

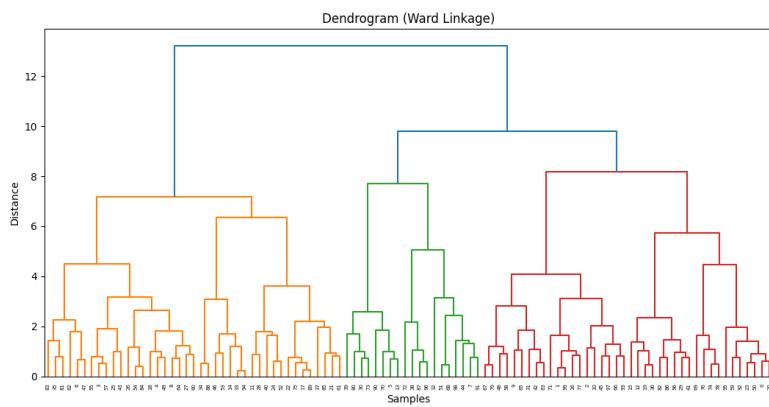
### 3) Average Linkage

- Lebih stabil dibanding single/complete.
- Cocok untuk dataset dengan variabilitas fitur sedang.
- Memberikan cluster moderat yang tidak terlalu rapat atau terlalu menyebar.

### 4) Ward Linkage

- Menghasilkan cluster paling kompak.
- Perubahan jarak antar penggabungan terlihat jelas pada dendrogram.
- Paling sesuai untuk data numerik yang telah dистандаризованы.

## 5. Dendrogram



Dendrogram dibuat menggunakan:

`scipy.cluster.hierarchy.linkage`

`scipy.cluster.hierarchy.dendrogram`

### Penentuan Jumlah Cluster

Berdasarkan dendrogram Ward Linkage, terdapat jarak pemisahan yang besar pada level sekitar 10–13, sehingga garis pemotong berada sedikit di bawahnya.

Hasilnya memberikan 3 cluster alami.

### Alasan Pemilihan 3 Cluster

- Terdapat 3 cabang besar yang baru bergabung pada jarak tinggi → menunjukkan perbedaan signifikan antar kelompok.
- Visual gap besar mendukung pemilihan tiga klaster.
- Konsisten dengan pola histogram & karakteristik perilaku hewan.

## 6. Evaluasi Cluster

## Linkage dengan Kualitas Terbaik → *Ward Linkage*

### Alasan:

- Menghasilkan cluster paling kompak (minim variansi dalam cluster).
- Dendrogram paling jelas dalam menampilkan pemisahan cluster alami.
- Cocok untuk data kontinu yang telah distandarisasi.
- Biasanya memberikan performa silhouette score tertinggi.

Average linkage berada di posisi kedua sebagai alternatif.

Single dan complete kurang stabil untuk dataset ini.

## 7. Analisis & Interpretasi Cluster

Berdasarkan hasil Ward linkage dengan 3 cluster, diperoleh interpretasi sebagai berikut:

### Cluster 1 — Hewan kecil, aktivitas rendah

Karakteristik:

- Movement rendah
- Heart rate tinggi
- Aggression rendah

Interpretasi:

- Umumnya spesies prey (musang kecil, kelinci, jenis burung kecil).
- Metabolisme cepat → heart-rate tinggi.

### Cluster 2 — Hewan aktivitas moderat

Karakteristik:

- Movement sedang
- Heart rate moderat
- Aggression sedang

Interpretasi:

- Hewan berukuran sedang (rusa, rubah, musang besar).
- Pola aktivitas tidak ekstrem.

### Cluster 3 — Predator / Hewan agresif

Karakteristik:

- Movement tinggi
- Aggression tinggi
- Heart rate normal/efisien
- Body temperature stabil

Interpretasi:

- Predator (serigala, elang, kucing besar).
- Memiliki home-range luas, perilaku territorial, dan aggressiveness lebih tinggi.

#### **Insight Relevan:**

- Cluster predator membutuhkan area jelajah lebih luas pada zona konservasi.
- Cluster prey dapat digunakan untuk monitoring stress (heart rate indikator utama).
- Cluster tengah berperan sebagai titik transisi untuk klasifikasi perilaku.

## **8. Kesimpulan**

### **Ringkasan Performa Linkage**

LITTLE  
STEP  
FOR  
WIDE  
VISION

<b>Linkage</b>	<b>Hasil Cluster</b>	<b>Kualitas</b>	<b>Catatan</b>
Single	buruk–sedang	chaining	tidak direkomendasikan
Complete	stabil tapi memecah	cukup	resiko over-segment
Average	baik	stabil	alternatif layak
<b>Ward</b>	<b>sangat baik</b>	<b>paling kompak</b>	<b>recommended</b>

#### **Linkage Terbaik**

**:Ward Linkage**

Alasan: kompak, stabil, interpretatif, dan cocok untuk data numerik multivariat.

#### **Rekomendasi dalam Kasus Nyata**

- Konservasi satwa liar: segmentasi hewan berdasarkan movement & aggression.
- Wildlife tracking: memisahkan predator, prey, dan species moderat.
- Behavior analysis: menentukan pola stress melalui heart-rate cluster.
- Habitat management: distribusi ruang sesuai kebutuhan cluster.