# BAB III
# METODE PENELITIAN

## 3.1 Objek Penelitian

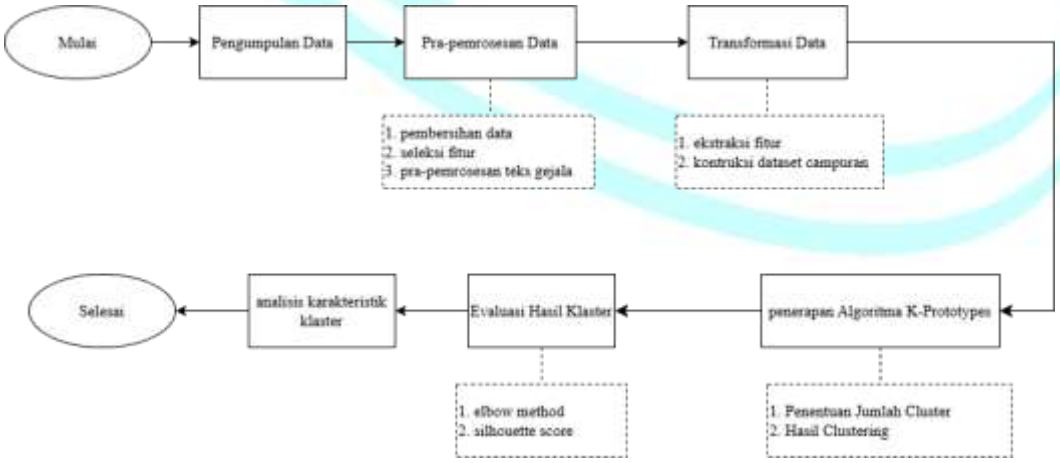
Objek penelitian ini adalah pola gejala penyakit hewan ternak yang tercatat dalam data kegiatan pengobatan tahun 2024 yang diperoleh dari Dinas Ketahanan Pangan dan Pertanian Kabupaten Karawang pada tanggal 9 Oktober 2025. Data tersebut merupakan data sekunder yang bersumber dari pencatatan kegiatan pemeriksaan dan pengobatan hewan ternak yang dilakukan oleh petugas lapangan.

Penelitian ini difokuskan pada identifikasi pola kesamaan gejala penyakit pada hewan ternak ruminansia, meliputi sapi, kambing, dan domba. Analisis dilakukan terbatas pada data gejala yang tersedia dalam dataset, tanpa melibatkan verifikasi laboratorium maupun pemeriksaan klinis tambahan, sehingga hasil penelitian sepenuhnya didasarkan pada informasi yang tercatat dalam data pengobatan.

Data yang dianalisis berasal dari satu periode pencatatan, yaitu tahun 2024. Ruang lingkup penelitian mengacu pada rumusan masalah yang telah ditetapkan, meliputi tahap pra-pemrosesan data, penerapan algoritma K-Prototypes, serta analisis karakteristik setiap *cluster* yang terbentuk sebagai hasil pengelompokan data penyakit hewan ternak.

## 3.2 Prosedur Penelitian

Prosedur penelitian ini bertujuan untuk memastikan bahwa setiap tahapan penelitian dilaksanakan secara sistematis dan terstruktur.




Gambar 3. 1 Prosedur Penelitian

Gambar 3.1 menampilkan prosedur penelitian yang dilakukan dalam analisis pola gejala penyakit hewan ternak. Alur ini disusun untuk memastikan setiap langkah penelitian berjalan teratur, mulai dari pengumpulan data hingga proses klastering. Penelitian ini terdiri dari enam tahapan yang dijelaskan sebagai berikut.

### 3.2.1 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan memperoleh dataset resmi dari Dinas Ketahanan Pangan dan Pertanian Kabupaten Karawang yang berisi catatan kegiatan pemeriksaan dan pengobatan hewan ternak selama tahun 2024.

| | col | hari tanggal | jenis ternak | jumlah ternak | kecamatan | gejala | diagnosa |
|---|---|---|---|---|---|---|---|
| 0 | NaN | 2024-01-04 00:00:00 | Kambing PE | 20 | Tirtajaya | kurus, buduk | MALNUTRISI |
| 1 | NaN | 2024-01-16 00:00:00 | sapi | 1 | Karawang Timur | tidak nafsu makan | MALNUTRISI |
| 2 | NaN | 2024-01-19 00:00:00 | Domba | 28 | pangkalan | Krang nfsu makan | MALNUTRISI |
| 3 | NaN | HARI/TANGGAL | JENIS TERNAK | JUMLAH TERNAK | KECAMATAN | GEJALA | DIAGNOSA |
| 4 | NaN | 2024-02-07 00:00:00 | SAPI | 1 | Cikampek | Demam,Batuk,Lendir | PMK |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 459 | NaN | NaN | sapi | 1 | majalaya | Luka borok | ABSES |
| 460 | NaN | 2024-12-31 00:00:00 | Sapi | 1 | purwasari | lsd | LSD |
| 461 | NaN | NaN | sapi | 1 | purwasari | Bekak sapi,tdk mau makan | GANGGUAN PENCERNAAN |
| 462 | NaN | NaN | sapi | 1 | majalaya | Demam,tidak mau makan | GANGGUAN PENCERNAAN |
| 463 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

464 rows × 7 columns

Gambar 3. 2 Pengumpulan Dataset Kegiatan Pengobatan Hewan Ternak Tahun 2024

Gambar 3.2 menampilkan hasil pengumpulan dataset kegiatan pengobatan hewan ternak tahun 2024 yang digunakan dalam penelitian ini. Dataset tersebut diterima dalam format Microsoft Excel dan digunakan sebagai sumber data utama dalam penelitian. Dataset terdiri dari 464 baris data pemeriksaan dengan 7 kolom atribut, yang mencakup *col* kolom tambahan sebagai data kosong, informasi tanggal pemeriksaan, jenis ternak, jumlah ternak, lokasi administrasi (kecamatan), gejala klinis yang diamati oleh petugas lapangan, serta diagnosa awal penyakit. Data tersebut merepresentasikan kondisi lapangan yang dicatat secara rutin oleh instansi dan mencerminkan karakteristik data kesehatan hewan ternak.

### 3.2.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas data agar layak digunakan dalam proses analisis klastering. Pada kondisi awal, dataset masih memiliki beberapa permasalahan, seperti posisi header yang tidak konsisten, keberadaan kolom kosong, kolom duplikat, serta baris yang tidak relevan. Oleh karena itu, dilakukan serangkaian tahapan pembersihan dan penyesuaian sebagai berikut:

#### 3.2.2.1 Pembersihan data

Tahap ini berfokus untuk menangani inkonsistensi struktur data. Langkah yang dilakukan meliputi penyesuaian header kolom berdasarkan kata kunci atribut (gejala, diagnosis, jenis ternak), serta penghapusan (*drop rows*) yang terindikasi sebagai data kosong (*null*), duplikat atau data rekapitulasi yang tidak relevan dengan kasus pemeriksaan.

Tabel 3. 1 Contoh Sebelum Pembersihan Data

| ***col*** | **hari/tanggal** | **jenis ternak** | **jumlah ternak** | **kecamatan** | **gejala** | **diagnosa** |
|---|---|---|---|---|---|---|
| *NaN* | 2021-01-05 | kambing | 1 | cikampek | *null* | Malnutrisi |
| *NaN* | 2024-01-04 | sapi | 2 | klari | Demam, Pincang kaki belakang | PMK |
| *NaN* | *NaN* | Domba | 5 | telukjambe | *null* | Abses |

Tabel 3.1 menunjukkan kondisi data sebelum dilakukan proses pembersihan. Pada tahap ini, data masih mengandung beberapa permasalahan, seperti nilai kosong (*null*) pada kolom tertentu, penulisan header yang belum konsisten, serta baris data yang tidak lengkap sehingga belum dapat digunakan secara langsung dalam proses analisis.

Tabel 3. 2 Contoh Sesudah Pembersihan Data

| **hari_tanggal** | **jenis_ternak** | **jumlah_ ternak** | **kecamatan** | **gejala** | **diagnosa** |
|---|---|---|---|---|---|
| 2024-01-04 | sapi | 2 | klari | Demam, Pincang | PMK |

| | | | | | |
|---|---|---|---|---|---|
| | | | | kaki belakang | |

Tabel 3.2 menampilkan kondisi data setelah dilakukan proses pembersihan. Baris data yang mengandung nilai kosong (*null*) atau tidak relevan dihapus agar struktur kolom disesuaikan agar konsisten. Hasil dari tahap ini adalah data yang lebih bersih dan siap digunakan pada tahap seleksi fitur dan proses analisis selanjutnya.

**3.2.2.2 Seleksi Fitur**

Tahap seleksi fitur dilakukan untuk menentukan atribut yang paling relevan dalam pembentukan klaster penyakit ternak. Pada tahap ini, dilakukan normalisasi penamaan kolom untuk menjaga konsistensi format, serta penghapusan atribut yang bersifat identitas atau administratif dan tidak merepresentasikan karakteristik penyakit, seperti nomor kasus, nama pemilik, dan desa. Selain itu, atribut jumlah ternak tidak disertakan dalam proses klastering karena tidak merepresentasikan kondisi klinis penyakit, sehingga atribut yang dipertahankan hanya meliputi jenis ternak, kecamatan, gejala, dan diagnosa.

Tabel 3. 3 Contoh Sebelum Seleksi Fitur

| hari_tanggal | jenis_ternak | jumlah_ternak | kecamatan | gejala | diagnosa |
|---|---|---|---|---|---|
| 2024-01-04 | sapi | 2 | Klari | Demam, Pincang kaki belakang | PMK |
| 2024-01-19 | kambing | 4 | Cikampek | batuk2, lendir hidung, Nafsu makan turun | Pneumonia |
| 2024-02-07 | domba | 3 | Telukjambe | diare encer, lemas, bulu kusam | Malnutrisi |

Tabel 3.3 menunjukkan contoh data sebelum dilakukan proses seleksi fitur. Pada tahap ini, dataset masih memuat beberapa atribut, seperti tanggal kejadian dan jumlah ternak, yang belum sepenuhnya merepresentasikan karakteristik klinis penyakit ternak. Selain itu, kolom gejala masih disajikan dalam bentuk teks mentah (*free text*), sebagaimana dicatat pada saat pemeriksaan di lapangan.

Tabel 3. 4 Contoh Sesudah Seleksi Fitur

| jenis_ternak | kecamatan | gejala | diagnosa |
|---|---|---|---|
| Sapi | Klari | Demam, Pincang kaki belakang | PMK |
| kambing | cikampek | batuk2, lendir hidung, Nafsu makan turun | pneumonia |
| domba | telukjambe | diare encer, lemas, bulu kusam | Malnutrisi |

Tabel 3.4 menampilkan contoh data setelah dilakukan seleksi fitur. Pada tahap ini, atribut yang tidak relevan terhadap pembentukan pola penyakit, seperti tanggal kejadian dan jumlah ternak, dihapus. Atribut yang dipertahankan meliputi jenis ternak dan kecamatan sebagai atribut kategorikal, serta gejala dan diagnosa sebagai representasi kondisi klinis penyakit. Kolom gejala tetap dipertahankan dalam bentuk teks mentah karena akan diproses lebih lanjut pada tahap pra-pemrosesan teks gejala sebelum dilakukan analisis klastering.

### 3.2.2.3 Pra-pemrosesan Teks Gejala

Dikarenakan atribut "gejala" berupa data teks bebas (*free-text*), maka dilakukan tahap pra-pemrosesan teks untuk mengurangi variasi penulisan dan meningkatkan konsistensi data. Proses ini meliputi empat tahapan yaitu:

*a. Case Folding*

Tahap *case folding* bertujuan untuk menyeragamkan format huruf dengan mengubah seluruh teks gejala menjadi huruf kecil (*lowercase*). Proses ini dilakukan untuk menghindari perbedaan makna akibat variasi penggunaan huruf besar dan kecil.

Tabel 3. 5 Contoh *Case Folding*

| Sebelum | Sesudah |
|---|---|
| Demam, Pincang kaki belakang, tdk mau makan | demam, pincang kaki belakang, tdk mau makan |
| Luka Borok Pada Kaki Depan | luka borok pada kaki depan |

Tabel 3.5 menunjukkan proses *case folding* dengan mengubah seluruh teks gejala menjadi huruf kecil untuk menghindari perbedaan makna akibat variasi kapitalisasi.

*b. Cleaning*

Tahap *cleaning* dilakukan untuk menghilangkan *noise* pada data teks, seperti tanda baca, angka, simbol khusus, dan spasi berlebih yang tidak memiliki kontribusi terhadap makna gejala penyakit.

Tabel 3. 6 Contoh *Cleaning*

| Sebelum | Sesudah |
|---|---|
| batuk2, lendir/hidung, nafsu makan turun bulu rontok, kulit kemerahan, gatal | batuk lendir hidung nafsu makan turun bulu rontok kulit kemerahan gatal |

Tabel 3.6 menunjukkan proses *cleaning*, yaitu penghapusan tanda baca, angka, dan simbol khusus yang tidak berkontribusi terhadap makna klinis gejala penyakit.

*c. Stopword Removal*

Tahap *stopword removal* bertujuan untuk menghapus kata-kata umum yang sering muncul tetapi tidak memberikan informasi klinis tambahan terhadap kondisi penyakit, seperti kata penghubung dan kata keterangan umum.

Tabel 3. 7 Contoh *Stopword Removal*

| Sebelum | Sesudah |
|---|---|
| nafsu makan turun ada lemas pincang pada kaki | nafsu makan turun lemas pincang kaki |

Tabel 3.7 menunjukkan proses *stopword removal*, yaitu penghapusan kata-kata umum yang tidak memiliki makna spesifik terhadap kondisi penyakit, sehingga teks gejala menjadi lebih ringkas dan informatif.

*d. Term Standardization*

Tahap *term standardization* bertujuan untuk menyeragamkan istilah gejala yang memiliki makna sama atau serupa (sinonim), sehingga dikenali sebagai satu entitas yang sama.

Tabel 3. 8 Contoh *Term Standardization*

| Sebelum | Sesudah |
|---|---|
| tdk mau makan pincang kaki belakang | nafsu makan turun pincang |

Tabel 3.8 menunjukkan proses *term standardization*, yaitu penyeragaman istilah gejala yang memiliki makna sama agar direpresentasikan sebagai satu entitas yang konsisten dalam proses analisis klastering.

### 3.2.3 Transformasi Data

Setelah melalui tahap pra-pemrosesan, data perlu ditransformasikan ke dalam bentuk yang dapat diproses oleh algoritma klastering. Oleh karena itu, pada tahap ini dilakukan transformasi data untuk mengonversi data teks dan data kategorikal ke dalam representasi yang sesuai dengan karakteristik algoritma K-Prototypes. Proses transformasi difokuskan pada dua aspek utama yaitu:

#### 3.2.3.1 Ekstraksi Fitur

Data teks gejala yang telah melalui tahap pra-pemrosesan diekstraksi menjadi fitur numerik menggunakan pendekatan biner (*binary encoding*). Pada tahap ini, kata kunci gejala dominan, seperti demam, pincang, batuk, sesak napas, diare, diidentifikasi sebagai fitur yang merepresentasikan keberadaan gejala pada setiap data.

Tabel 3. 9 Contoh Hasil Ekstraksi Fitur

| gejala | demam | pincang | batuk | sesak_nafas | diare |
|---|---|---|---|---|---|
| demam pincang luka mulut | 1 | 1 | 0 | 0 | 0 |
| batuk lendir hidung sesak | 0 | 0 | 1 | 1 | 0 |
| diare encer bulu kusam | 0 | 0 | 0 | 0 | 1 |

Nilai 1 menunjukkan keberadaan gejala pada data, sedangkan nilai 0 menunjukkan gejala tidak ditemukan. Pendekatan biner digunakan untuk merepresentasikan keberadaan gejala secara sederhana dan sesuai dengan karakteristik algoritma K-Prototypes.

#### 3.2.3.2 Kontruksi Dataset Campuran

Fitur numerik hasil ekstraksi gejala selanjutnya digabungkan dengan atribut kategorikal lainnya, yaitu jenis ternak, kecamatan, dan diagnosis awal.

Tabel 3. 10 Contoh Dataset Campuran

| jenis_ ternak | kecamatan | diagnosis | demam | pincang | batuk | sesak_ nafas | diar e |
|---|---|---|---|---|---|---|---|
| sapi | klari | PMK | 1 | 1 | 0 | 0 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| kambing | cikampek | Pneumonia | 0 | 0 | 1 | 1 | 0 |
| domba | telukjambe | Diare | 0 | 0 | 0 | 0 | 1 |

Dataset campuran ini terdiri atas atribut kategorikal dan numerik, sehingga sesuai dengan karakteristik algoritma K-Prototypes dan siap digunakan pada tahap pemodelan klaster.

### 3.2.4 Penerapan Algoritma K-Prototypes

Tahap ini merupakan implementasi utama dari metode klastering. Algoritma K-Prototypes dipilih karena kemampuannya menangani data bertipe campuran (*mixed data types*), yaitu data numerik dan kategorikal, dalam satu proses klastering. Karena klastering termasuk dalam *unsupervised learning*, jumlah klaster optimal belum diketahui sejak awal. Oleh karena itu, penerapan algoritma dilakukan melalui serangkaian percobaan dengan variasi jumlah klaster (*k*). Proses penerapan algoritma K-Prototypes dijelaskan dalam bentuk pseudocode sebagai berikut:

Tabel 3. 11 Pseudocode Algoritma K-Prototypes

***Algoritma K-Prototypes***

```
BEGIN

INPUT:
    Dataset campuran (atribut numerik dan kategorikal)
    Rentang jumlah klaster k =  2 sampai k_max

OUTPUT:
    Hasil klaster dan nilai cost function untuk setiap nilai k
FOR setiap nilai k dalam rentang jumlah klaster DO

    INITIALIZE k pusat klaster secara acak dari dataset

    REPEAT
        FOR setiap data dalam dataset DO
            CALCULATE jarak data ke seluruh pusat klaster
            ASSIGN data ke klaster dengan jarak terdekat
        END FOR

        UPDATE pusat klaster:
            CALCULATE mean untuk atribut numerik
            CALCULATE mode untuk atribut kategorikal
```

```
        UNTIL pusat klaster tidak berubah (konvergen)

        CALCULATE nilai cost function
        SAVE hasil klaster dan nilai cost untuk nilai k

    END FOR

    END
```

Tabel 3.11 menggambarkan pseudocode dari alur penerapan algoritma K-Prototypes dalam penelitian ini. Proses klastering dilakukan secara iteratif untuk setiap variasi jumlah klaster (k), dimulai dari inisialisasi pusat klaster, perhitungan jarak campuran, hingga pembaruan pusat klaster. Proses iterasi dihentikan ketika pusat klaster telah mencapai kondisi konvergensi. Nilai cost function yang dihasilkan pada setiap nilai k selanjutnya digunakan sebagai dasar evaluasi dalam penentuan jumlah klaster optimal dan penilaian kualitas klaster.

### 3.2.5 Evaluasi Hasil Klaster

Setelah diperoleh hasil klaster dari tahap penerapan algoritma K-Prototypes, dilakukan evaluasi untuk menentukan jumlah klaster terbaik serta menilai kualitas klaster yang terbentuk. Tahap evaluasi ini dilakukan menggunakan dua metode, yaitu:

#### 3.2.5.1 *Elbow Method*

*Elbow Method* digunakan untuk menentukan jumlah klaster optimal berdasarkan perubahan nilai *cost function* pada setiap percobaan jumlah klaster. Nilai *cost function* dari setiap percobaan dengan variasi jumlah klaster ($k = 2$ sampai $k = 10$) diplot ke dalam grafik. Grafik dianalisis untuk menemukan titik siku (*elbow point*), yaitu titik di mana penurunan nilai *cost function* mulai melandai secara signifikan. Nilai *k* pada titik tersebut dipilih sebagai jumlah klaster yang paling optimal.

#### 3.2.5.2 *Silhouette Score*

*Silhouette Score* digunakan untuk mengevaluasi kualitas klaster yang terbentuk dengan mengukur tingkat kedekatan data terhadap klaster asalnya dibandingkan dengan klaster terdekat lainnya. Nilai Silhouette Score berada pada rentang −1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa klaster yang terbentuk semakin kompak dan terpisah dengan baik, sedangkan nilai yang

mendekati 0 atau bernilai negatif menunjukkan adanya tumpang tindih antar klaster atau struktur klaster yang kurang optimal.

### 3.2.6 Analisis Karakteristik Klaster

Hasil klastering yang diperoleh selanjutnya dianalisis untuk mengidentifikasi karakteristik masing-masing klaster penyakit ternak. Analisis dilakukan dengan menyajikan tabel karakteristik dan grafik distribusi untuk menggambarkan pola pengelompokan penyakit yang terbentuk pada setiap klaster. Setiap klaster diberi label identitas (*Cluster ID*) yang berbeda untuk membedakan satu kelompok penyakit dengan kelompok lainnya.

Selain itu, hasil klastering diekspor kembali ke dalam format *.csv* dengan menambahkan kolom *Cluster ID* pada data rekam medis ternak. Proses ini bertujuan untuk memudahkan analisis lanjutan serta interpretasi karakteristik penyakit pada tahap pembahasan.