**Metode**

Dalam analisis ini, metode yang digunakan adalah klasifikasi K-Means dengan pendekatan data mining. Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar [1]. K-Means merupakan metode analisis data pada data mining dimana proses pemodelan tanpa *supervise* dan merupakan salah satu metode yang mengelompokkan data secara partisipasi. Pada metode k-means data dikelompokkan menjadi beberapa kelompok dimana setiap kelompok mempunyai karakteristik yang mirip atau sama dengan lainnya namun dengan kelompok lainnya memiliki karakteristik yang berbeda. Sebuah cluster adalah suatu kumpulan data yang mirip dengan lainnya atau kemiripan data pada kelompok lainnya. Metode ini meminimalisasi perbedaan antar data di dalam satu cluster. Serta memaksimalkan perbedaan dengan cluster yang lain [2]. Adapun langkah-langkah metode k-means yaitu sebagai berikut:

1. Tentukan jumlah klaster

Metode yang digunakan adalah metode gap statistic(k). metode tersebut bertujuan untuk menentukan jumlah klaster lebih konstan dan baik dibandingkan metode pengukuran lainnya.

1. Menentukan centroid (titik pusat) awal.
2. Hitung jarak antara data dengan pusat klaster dengan menggunakan euclidian distance.

Adapun perhitungan Euclidian distance adalah sebagai berikut:

(1)

dengan

= Jarak objek antara objek i dan j

= Dimensi data

= Koordinat dari objek i pada dimensi k

= Koordinat dari objek j pada dimensi k

1. Kelompokkan setiap data terhadap jarak pada titik pusat centroid terdekat.
2. Perbarui nilai titik centroid

Adapun perhitungan untuk menghitung titik centroid adalah sebagai berikut:

(2)

dengan

= titik centroid dari klaster ke-k.

= banyaknya data pada klaster ke-k.

= data ke-I pada klaster ke-k.

1. Ulangi langkah 3, 4, dan 5 dengan memakai pusat klaster yang baru. Apabila pusat klaster tidak mengalami perubahan lagi, maka proses klasifikasi dengan metode K-means dihentikan karena sudah konvergen.

Dalam melakukan validasi hasil klaster yang telah dilakukan, maka digunakan metode *silhouette coefficient*. Validasi cluster berguna untuk mengukur tingkat kebaikan dari cluster yang terbentuk, salah satu internal validation index adalah Silhouette Coefficient[3].

**Data dan Sumber Data**

Data yang digunakan dalam analisis ini yaitu data produksi daging sapi dan kambing menurut provinsi (Ton) tahun 2021. Sumber data berasal dari website Badan Pusat Statistik (BPS), yaitu sebagai berikut:

1. Produksi daging sapi menurut provinsi (Ton) : <https://www.bps.go.id/indicator/24/480/1/produksi-daging-sapi-menurut-provinsi.html>
2. Produksi daging kambing menurut provinsi (Ton) : <https://www.bps.go.id/indicator/24/482/1/produksi-daging-kambing-menurut-provinsi.html>

Proses pengolahan menggunakan bantuan R-Studio dan MS-Excel. Adapun hasil proses pengolahan data dapat diakses melalui: [aryasepta7901/-2022-Tugas-Akhir-Statistik-Produksi (github.com)](https://github.com/aryasepta7901/-2022-Tugas-Akhir-Statistik-Produksi)

**Referensi**

[1] Setiawan, R. (2017). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik Lp3i Jakarta). *Jurnal Lentera Ict*, *3*(1), 76-92.

[2] Rendy Novianto, R. N., & Leonard Goeirmanto, L. G. (2019). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menganalisa Bisnis Perusahaan Asuransi. *Jatisi (Jurnal Teknik Informatika Dan Sistem Informasi)*, *6*(1).

[3] Azuri, D. F., & Pontoh, R. S. (2016). Pengelompokkan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means. Jatinagor: Prosiding Seminar Nasional FMIPA Universitas Padjadjaran.