

# Penerapan Data Mining Dalam Mengelompokkan Produksi Daging Sapi di Pulau Sumatera Menggunakan Algoritma K-Means Clustering

Asril Andi Novany<sup>1</sup>, M.Safii<sup>2</sup>

<sup>1</sup>STIKOM Tunas Bangsa Pematang Siantar

<sup>2</sup>AMIK Tunas Bangsa Pematang Siantar

E-mail: <sup>1</sup>[asrilandinovany09@gmail.com](mailto:asrilandinovany09@gmail.com), <sup>2</sup>[m.safii@amiktunasbangsa.ac.id](mailto:m.safii@amiktunasbangsa.ac.id)

## Abstrak

Daging sapi merupakan salah satu daging yang memiliki sumber protein yang tinggi. Daging sapi bukan merupakan makanan yang asing dan sering dijumpai dalam kehidupan sehari-hari. Namun produksi daging sapi di pulau sumatera masih belum merata karena masih terdapat provinsi yang rendah memproduksi daging sapi. Penelitian ini membahas tentang penerapan Data Mining pada produksi daging sapi yang ada di pulau sumatera, dengan menggunakan K-Means Clustering. Sumber data penelitian ini dikumpulkan berdasarkan dokumen-dokumen pertanian yang dihasilkan oleh Badan Pusat Statistika Nasional. Data yang digunakan pada penelitian ini adalah data dari tahun 2016-2020 yang terdiri dari 10 Provinsi. Variable yang digunakan (1) jumlah produksi daging sapi. Data akan diolah dengan melakukan 2 Cluster yaitu cluster produksi tertinggi dan cluster produksi terendah. Centroid data untuk cluster tingkat produksi tertinggi 110457.86 dan centroid data untuk cluster terendah 8088.28. sehingga diperoleh nilai berdasarkan indeks produksi daging sapi yang ada pada provinsi di pulau sumatera dengan 5 provinsi tingkat produksi tertinggi berada pada provinsi Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, dan Lampung, sedangkan 5 provinsi tingkat produksi terendah berada pada provinsi Riau, Jambi, Bengkulu, Kepulauan Bangka Belitung, dan Kepulauan Riau. Hal ini dapat menjadi masukan kepada pemerintah, provinsi yang menjadi perhatian lebih pada populasi daging ayam berdasarkan cluster yang telah dilakukan.

**Kata Kunci:** Clustering, K-Means, Data Mining, Produksi, Daging Sapi

## Abstract

Beef is one of the meats that has a high protein source. Beef is not a foreign food and is often found in everyday life. However, beef production on the island of Sumatra is still not evenly distributed because there are still provinces that are low in beef production. This study discusses the application of Data Mining on beef production on the island of Sumatra, using K-Means Clustering. Sources of data for this study were collected based on agricultural documents produced by the National Statistics Agency. The data used in this study is data from 2016-2020 which consists of 10 provinces. Variables used (1) the amount of beef production. The data will be processed by doing 2 clusters, namely the highest production cluster and the lowest production cluster. The data centroid for the highest production level cluster is 110457.86 and the data centroid for the lowest cluster is 8088.28. so that the value is obtained based on the beef production index in the provinces on the island of Sumatra with the 5 provinces with the highest production levels being in the provinces of Aceh, North Sumatra, West Sumatra, South Sumatra, and Lampung, while the 5 provinces with the lowest production levels are in the provinces of Riau, Jambi, Bengkulu, Bangka Belitung Islands, and Riau Islands. This can be input to the government, provinces that are more concerned with the population of chicken meat based on the clusters that have been carried out.

**Keywords:** Clustering, K-Means, Data Mining, Production, Beef

## 1. Pendahuluan

Salah satu subsektor pertanian merupakan peternakan, yang produksinya antara lain, daging, telur, serta susu. Daging ialah sumber protein hewani bermutu besar serta wajib disantap oleh kanak-kanak dan orang berumur. Zona pertanian Indonesia merupakan sector penyusun PDB, melalui kontribusi terhadap PDB pada tahun 2013 sebesar Rp 2.769.053,00 milyar, setelah itu sebesar Rp 2.909.181,50 milyar pada tahun 2014 (BPS, ditangani oleh kementerian pertanian). [1] Pemenuhan kebutuhan pangan ialah salah satu tujuan pembangunan ekonomi disemua Negeri. Sumber pangan sangat bermacam-macam, namun pangan dengan kandungan protein besar memiliki harga yang relative lebih mahal ketimbang dengan sumber pangan berkarbohidrat. Salah satu sumber protein bisa diperoleh dari daging sapi. Oleh sebab itu, hasil ternak merupakan komoditas yang bernilai jual tinggi. Dengan meningkatnya pendapatan masyarakat serta terwujudnya kesadaran pangan yang bergizi, permintaan komoditas yang bernilai jual tinggi ini pula terus menjadi bertambah. [2]

Clustering merupakan tatacara analisis data yang bertujuan untuk mengelompokkan data dengan karakteristik yang sama ke dalam zona yang sama. Salah satu metode yang digunakan untuk mengembangkan metode clustering adalah metode K-Means, yaitu metode pengelompokkan data non hierarki yang bertujuan untuk membagi data menjadi dua atau lebih kelompok (Cluster) dengan karakteristik yang sama dan membaginya menjadi satu kelompok yang sama. [3] Data yang digunakan dalam penelitian ini berdasarkan dokumen-dokumen yang dihasilkan oleh Badan Pusat Statistika melalui situs <https://www.bps.go.id> . dalam hal ini peneliti mengangkat topik produksi daging sapi di pulau sumatera dimana proses metode yang dilakukan adalah clustering. Hasil dari cluster dapat menjadi masukan bagi pemerintah agar provinsi yang masuk kedalam cluster rendah mendapat perhatian lebih. Proses cluster dibagi menjadi 2 (dua) cluster yakni produksi daging sapi terendah dan produksi daging sapi tertinggi.

## 2. Tinjauan Pustaka

### 2.1. Daging

Daging adalah bahan makanan yang nilai gizinya sangat besar sebab kaya akan protein, lemak, mineral serta zat lain yang dibutuhkan tubuh manusia. [4]

### 2.2. Data Mining

Di dalam bukunya Joko Suntoro, Data Mining Algoritma dan Implementasi Dengan Pemrograman PHP, 2019 [5] ada beberapa pengertian data mining menurut para ahli, Yaitu:

1. Menurut John Naisbitt dalam Larose (Larose, 2005), kita terbenam di dalam data, namun kita kekurangan informasi dan pengetahuan. Hal tersebut menjadi gambaran apa yang telah terjadi pada hari ini bahwa skala volume data yang jumlahnya sangat besar tersebut hanya menjadi sampah di memori penyimpanan saja apabila tidak dikelola menjadi informasi atau pengetahuan, diperlukan teknik/metode yang dinamakan dengan data mining.
2. Menurut (Witten, Ian H. Frank, 2011), Definisi adalah proses ekstraksi suatu data (sebelumnya tidak diketahui, bersifat implisit dan dianggap tidak berguna) menjadi informasi atau pengetahuan atau pola dari data yang jumlahnya besar.

### 2.3. Clustering

Analisis Clustering/Pengelompokan merupakan proses membagi sekelompok data menjadi beberapa kelompok, kesamaan satu kelompok data lebih besar daripada kelompok data lainnya. Potensi clustering adalah dapat digunakan untuk menentukan struktur data, yang selanjutnya dapat digunakan dalam berbagai aplikasi seperti klasifikasi, pengolahan citra, serta pengenalan pola.[6]

## 2.4. Algoritma K-Means

Algoritma K-Means merupakan metode data pengelompokan non hierarki yang mencoba membagi data yang ada menjadi satu atau lebih cluster atau grup, untuk mengelompokkan data dengan karakteristik yang sama, dan mengelompokkan data dengan karakteristik berbeda ke dalam group lain.[7] Secara umum algoritma k-means adalah:

1. Menentukan banyaknya *cluster* (*k*).
2. Menentukan *centroid*.
3. Apakah *centroid*-nya berubah ?
  - a) Jika ya, hitung jarak data dari *centroid*.
  - b) Jika tidak, selesai.
4. Mengelompokkan data berdasarkan jarak yang terdekat.

## 3. METODE PENELITIAN

### 3.1. Tahap Pengumpulan Data

Dalam penerapan Data Mining produksi daging sapi yang ada pada provinsi di pulau sumatera, diperlukan data terkait tentang hal itu. Sumber data penelitian diperoleh dari data yang dikumpulkan berdasarkan dokumen-dokumen keterangan pertanian yang dihasilkan oleh Direktorat Jenderal Bea dan Cukai melalui situs <https://www.bps.go.id>.

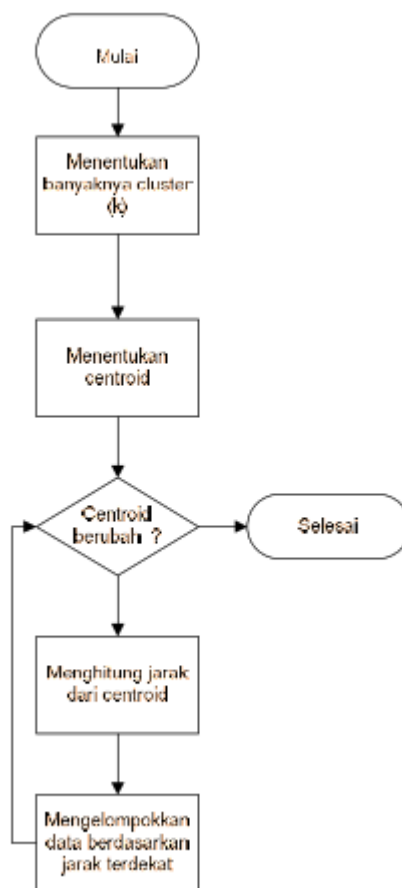
Data yang digunakan dalam penelitian ini adalah data produksi daging sapi yang ada pada provinsi dipulau sumatera dari tahun 2016-2020 yang terdiri dari 10 provinsi. Varibale yang digunakan jumlah produksi daging sapi yang sudah diakumulasikan. Kemudian data akan diolah dengan melakukan clustering produksi daging sapi berdasarkan provinsi yang ada di pulau sumatera dalam 2 cluster yakni cluster tingkat produksi tertinggi dan cluster tingkat produksi terendah.

### 3.2. Tahap Pengolahan Data

Data yang telah diperoleh akan diolah terlebih dahulu untuk dapat di clustering. Dalam tahap sebelumnya, data setiap provinsi akan dijumlah setiap aspeknya sehingga pada tahapan ini sudah diperoleh perhitungan nilai yang akan diproses pada tahap clustering.

### 3.3. Tahap Clustering

Analisis Clustering/Pengelompokan adalah proses membagi sekelompok data menjadi beberapa kelompok, kesamaan satu kelompok data lebih besar daripada kelompok data lainnya. Potensi clustering adalah dapat digunakan untuk menentukan struktur data, yang selanjutnya dapat digunakan dalam berbagai aplikasi seperti klasifikasi, pengolahan citra, dan pengenalan pola. [6] Dalam menentukan cluster berdasarkan data yang telah tersedia, dibutuhkan sebuah flowchart untuk memudahkan dalam menentukan alur perhitungan sebagai alur untuk menemukan hasil dari penerapan cluster terhadap data yang akan diproses. Berikut adalah flowchart dalam menentukan cluster dengan K-Means



Gambar 1. Flowchart K-Means

### 3.4. Tahap Analisis

Pada tahapan ini dilakukan analisis data produksi daging sapi. Data yang diperoleh kemudian akan diolah dengan menggunakan perhitungan bobot dari tiap indeks. Pada tahapan sebelumnya, telah ditentukan akan dicluster ke dalam 2 cluster yaitu cluster tingkat produksi tertinggi dan cluster tingkat produksi terendah, pada tahap ini akan dianalisis hasilnya.

## 4. HASIL DAN PEMBAHASAN

Dalam melakukan clustering, data yang diperoleh akan dihitung terlebih dahulu berdasarkan jumlah produksi daging sapi yang ada pada provinsi di pulau sumatera dari tahun 2016-2020. Hasil penjumlahan berdasarkan 1 kriteria penilaian yakni produksi daging sapi yang ditunjukkan pada table 1.

Tabel 1. Data Produksi Daging Sapi yang ada di Provinsi Pada Pulau Sumatera

Provinsi	Tahun					Total
	2016	2017	2018	2019	2020	
Aceh	10400.81	8550.28	11524.28	10416.2	10740.15	51631.72
Sumatera Utara	25571.07	26297.65	15240.33	14153.16	14569.88	95832.09
Sumatera Barat	26440.92	20206.48	20198.94	21589.63	22021.89	110457.86
Riau	9396.29	7700.43	9792.53	8379.1	8611.02	43879.37
Jambi	4386.38	4085.1	4540.06	5026.46	5094.19	23132.19
Sumatera Selatan	17878.73	12666.1	11261.05	11455.31	11615.04	64876.23
Bengkulu	3056.96	2587.24	2471.91	2587.44	3149.1	13852.65

Lampung	12609.07	12998.57	13332.35	14326.31	13522.13	66788.43
Kep. BangkaBelitung	2316.67	2601.43	3065.93	3212.58	3015.31	14211.92
Kep. Riau	2668.8	1470.16	1164.33	1384.08	1400.91	8088.28

Sumber: Badan Pusat Statistika, Url: <https://www.bps.go.id>.

Kemudian data tersebut diakumulasikan menjadi 1 kriteria yaitu jumlah produksi daging sapi seperti yang ditunjukkan pada tabel 2.

Tabel 2. Data Akumulasi Produksi Daging Sapi

Provinsi	Jumlah Produksi Daging Sapi
Aceh	51631.72
Sumatera Utara	95832.09
Sumatera Barat	110457.86
Riau	43879.37
Jambi	23132.19
Sumatera Selatan	64876.23
Bengkulu	13852.65
Lampung	66788.43
Kep. BangkaBelitung	14211.92
Kep. Riau	8088.28

Setelah diakumulasikan maka akan didapatkan nilai dari seluruh produksi daging sapi yang ada di provinsi pada pulau sumatera, selanjutnya kita akan masuk ke tahap clustering dengan menerapkan algoritma K-Means untuk mengcluster data menjadi 2 cluster.

#### 4.1. Centroid Data

Dalam penerapan algoritma K-Means dihasilkan nilai titik tengah atau centroid dari data yang di dapat dengan ketentuan bahwa clusterisasi yang diinginkan adalah 2, penentuan cluster dibagi atas dua bagian yaitu cluster tingkat produksi tertinggi (C1), dan tingkat produksi terendah (C2). Maka nilai titik tengah atau centroid juga terdapat 2 titik. Penentuan titik cluster ini dilakukan dengan mengambil nilai terbesar (Maksimum) untuk tingkat produksi tertinggi (C1) dan mengambil nilai terkecil (Minimum) untuk tingkat produksi terendah (C2). Nilai titik tersebut dapat dilihat pada table 3 dibawah ini.

Tabel 3. Centroid Data Awal

Atribut	Cluster Produksi Tertinggi	Cluster Produksi Terendah
Nilai	110457.86	8088.28

#### 4.2. Clustering Data

Dengan menggunakan centroid tersebut maka data yang telah di dapat, dapat dicluster menjadi 2 cluster. Pada proses cluster dilakukan dengan mengambil jarak terdekat dari setiap data yang diolah. Dari data produksi daging sapi yang ada pada provinsi di pulau sumatera tersebut didapatkan pengelompokan pada iterasi 1 untuk dua cluster tersebut. cluster tingkat produksi tertinggi (C1) berada pada provinsi Sumatera Utara, Sumatera Barat, Jambi, Sumatera Selatan dan Lampung, dan cluster pada tingkat produksi terendah (C2) berada pada provinsi Aceh, Riau, Bengkulu, Kep Bangka Belitung, dan Kep Riau. Proses pencarian data terpendek dan pengelompokan data pada iterasi 1 dapat dilihat pada tabel dibawah ini.

Tabel 4. Perhitungan Jarak Pusat Cluster

Provinsi	Jumlah Produksi Daging Sapi	C1	C2	Jarak Terpendek
Aceh	51631.72	26842.598	19647.932	19647.93214
Sumatera Utara	95832.09	13163.599	40933.155	13163.59917
Sumatera Barat	110457.86	0	45956.162	0
Riau	43879.37	30159.304	16107.055	16107.05473
Jambi	23132.19	39400.353	6934.7527	6934.752664
Sumatera Selatan	64876.23	20519.778	25775.821	20519.77764
Bengkulu	13852.65	43459.869	2759.0841	2759.084075
Lampung	66788.43	20381.998	26347.66	20381.99786
Kep. BangkaBelitung	14211.92	43409.945	3312.031	3312.031026
Kep. Riau	8088.28	45956.162	0	0

Tabel 5. Pengelompokan Data Iterasi 1

Provinsi	C1	C2
Aceh		1
Sumatera Utara	1	
Sumatera Barat	1	
Riau		1
Jambi	1	
Sumatera Selatan	1	
Bengkulu		1
Lampung	1	
Kep. Bangka Belitung		1
Kep. Riau		1

Setelah semua data dikelompokkan pada cluster terdekat kemudian hitung kembali pusat cluster terbaru berdasarkan rata-rata anggota yang ada pada cluster tersebut pada proses iterasi berikutnya. setelah menemukan centroid baru kemudian melakukan Proses pencarian data terpendek dan pengelompokan data kembali pada iterasi berikutnya.

Proses K-Means akan terus beriterasi sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya. Dengan kata lain, proses akan terus melakukan iterasi sampai data pada iterasi terakhir sama dengan iterasi sebelumnya. Pada proses ini, iterasi ke 2 posisi cluster tidak sama dengan iterasi 1 dapat dilihat pada tabel dibawah ini.

Tabel 6. Pengelompokan Data Iterasi 2

Provinsi	C1	C2
Aceh		1
Sumatera Utara	1	
Sumatera Barat	1	
Riau	1	
Jambi		1
Sumatera Selatan	1	
Bengkulu		1
Lampung	1	
Kep. Bangka Belitung		1
Kep. Riau		1

Jika posisi cluster pada iterasi ke 2 mengalami perubahan seperti tabel diatas, maka harus dilanjutkan pada iterasi ke 3 dengan cara yang sama yaitu menghitung kembali pusat cluster terbaru berdasarkan rata-rata anggota yang ada pada cluster tersebut pada proses iterasi berikutnya kemudian mengelompokkannya kembali. Pada proses ini iterasi ke 3 posisi cluster juga tidak sama dengan iterasi ke 2, dapat dilihat pada tabel dibawah ini

**Tabel 7. Pengelompokan Data Iterasi Ke 3**

Provinsi	C1	C2
Aceh	1	
Sumatera Utara	1	
Sumatera Barat	1	
Riau		1
Jambi		1
Sumatera Selatan	1	
Bengkulu		1
Lampung	1	
Kep. Bangka Belitung		1
Kep. Riau		1

Pada iterasi ke 3 diperoleh cluster data produksi daging sapi yang ada di provinsi di pulau sumatera, yang dapat dilihat pada tabel diatas. Proses iterasi tersebut berhenti pada iterasi ke 4, pada iterasi ke 4 akan dilakukan proses mencari nilai titik tengah atau centroid kembali. Setelah mendapatkan nilai centroid, proses sama dilakukan dengan mencari jarak terdekat, proses pencarian jarak terpendek, pengelompokan data pada iteasi ke 4 dapat digambarkan pada tabel berikut:

**Tabel 8. Perhitungan jarak Pusat Cluster Iterasi 4**

Provinsi	Jumlah Produksi Daging Sapi	C1	C2	Jarak Terpendek
Aceh	51631.72	12735.661	13976.562	12735.66139
Sumatera Utara	95832.09	12365.179	35912.53	12365.17889
Sumatera Barat	110457.86	14882.683	40458.309	14882.68275
Riau	43879.37	15708.891	10475.935	10475.93477
Jambi	23132.19	24866.594	1341.4835	1341.483534
Sumatera Selatan	64876.23	6227.3413	20518.083	6227.341297
Bengkulu	13852.65	28859.682	3081.9521	3081.952107
Lampung	66788.43	6888.5795	20688.523	6888.579543
Kep. Bangka Belitung	14211.92	28827.162	3005.7406	3005.740641
Kep. Riau	8088.28	31334.658	5715.9486	5715.948568

**Tabel 9. Pengelompokan Data Iterasi Ke 4**

Provinsi	C1	C2
Aceh	1	
Sumatera Utara	1	
Sumatera Barat	1	
Riau		1
Jambi		1

Sumatera Selatan	1	
Bengkulu		1
Lampung	1	
Kep. Bangka Belitung		1
Kep. Riau		1

### 4.3. Analisa Data

Pada Iterasi 4, data yang dikelompokkan yang dilakukan terhadap 2 cluster dengan iterasi 3 didapatkan hasil yang sama, dari 10 data produksi daging sapi yang ada pada provinsi pulau sumatera dapat diketahui, 5 provinsi cluster tingkat tertinggi berada pada provinsi Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, dan Lampung, sedangkan 5 provinsi tingkat terendah berada pada provinsi Riau, Jambi, Bengkulu, Kep Bangka Belitung, dan Kep Riau.

### 5. Kesimpulan

Berdasarkan hasil penelitian produksi daging sapi yang ada pada provinsi di pulau sumatera dengan menggunakan metode Algoritma K-Means Clustering, dapat menentukan pengelompokan produksi daging sapi dengan menginput data yang digunakan dalam penelitian yaitu tahun, 2016, 2017, 2018, 2019, 2020. Output yang ditampilkan dari data yang digunakan yaitu berupa pengelompokan data berdasarkan algoritma K-Means. Data tersebut diolah menggunakan MS Excel untuk menentukan nilai titik centroid dalam 2 cluster yaitu cluster produksi tertinggi dan cluster produksi terendah. Centroid data untuk cluster tingkat produksi tertinggi 110457.86 dan centroid data untuk cluster terendah 8088.28. sehingga diperoleh nilai berdasarkan indeks-indeks produksi daging sapi yang ada pada provinsi di pulau sumatera dengan 5 provinsi berada pada cluster produksi tertinggi berada pada provinsi Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, dan Lampung, sedangkan 5 provinsi tingkat terendah berada pada provinsi Riau, Jambi, Bengkulu, Kep Bangka Belitung, dan Kep Riau. Dan hasil dari penelitian dapat menjadi masukan kepada pemerintah, provinsi yang menjadi perhatian lebih pada populasi daging ayam berdasarkan cluster yang telah dilakukan.

### Daftar Pustaka

- [1] Fatmawati, Rostin, and J. N. Baso, "Faktor-faktor yang mempengaruhi permintaan daging sapi di Indonesia," *J. Ekon.*, vol. 1, no. April, pp. 128–134, 2016.
- [2] A. Revi, S. Solikhun, and M. Safii, "Jaringan Syaraf Tiruan Dalam Memprediksi Jumlah Produksi Daging Sapi Berdasarkan Provinsi," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 297–304, 2018.
- [3] Y. Darmi and A. Setiawan, "Penerapan metode clustering k-means dalam pengelompokan penjualan produk," *J. Media Infotama Univ. Muhammadiyah Bengkulu*, vol. 12, no. 2, pp. 148–157, 2016.
- [4] N. P. Kurniawan, D. Septinova, and K. Adhianto, "Kualitas Fisik Daging Sapi dari Tempat Pemotongan Hewan Di Bandar Lampung," *J. Ilm. Peternak. Terpadu*, vol. 2, no. 3, pp. 133–137, 2014.
- [5] J. SUNTORO, "DATA MINING : Algoritma dan Implementasi dengan Pemrograman php." p. 179, 2019.
- [6] M. G. Sadewo, A. P. Windarto, and D. Hartama, "Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 60–67, 2017.
- [7] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.