

## Penerapan K-Means *Clustering* untuk Mengelompokkan Pelanggan Berdasarkan Data Penjualan Ayam

<sup>1</sup>Florida Ivonia Manek, <sup>2</sup>Sutan Faisal, <sup>3</sup>Bayu Priyatna  
<sup>1</sup>lf15.floridamanek@ubpkarawang.ac.id, <sup>2</sup>sutan.faisal@ubpkarawang.ac.id, <sup>3</sup>bayu.priyatna@ubpkarawang.ac.id  
Universitas Buana Perjuangan  
Karawang, Indonesia

### ABSTRAK

Suatu kegiatan pemasaran yang berhubungan langsung dengan pelanggan disebut dengan penjualan. Topik yang berkaitan dengan penjualan pada penelitian ini adalah penjualan ayam. Hal ini dikarenakan pada umumnya ayam sangat dibutuhkan masyarakat sebagai sumber protein, sehingga dari waktu ke waktu proses penjualan selalu mengalami peningkatan. Sehingga data yang diperoleh perusahaan pun menjadi semakin meningkat. Masalah ini menyebabkan perusahaan sulit dalam menentukan pelanggan untuk dijadikan sebagai pelanggan yang diprioritaskan oleh perusahaan. Penelitian ini membahas penerapan algoritme K-Means untuk mengelompokkan pelanggan sesuai data penjualan ayam. Data diperoleh dari dokumen-dokumen keterangan peternak yang ada pada perusahaan. Data yang ada digunakan untuk mengelompokkan pelanggan sesuai dengan tingkat penjualan terendah, sedang dan tertinggi. Hasil yang diperoleh dari pengolahan data mampu memberikan suatu informasi penting bagi perusahaan untuk mengetahui dan mengelompokkan pelanggan yang harus diprioritaskan.

**Kata kunci** — *Clustering*, Data Penjualan, K-Means.

### I. PENDAHULUAN

Data penjualan ayam yang diperoleh perusahaan setiap hari mengalami perubahan, karena perubahan yang terjadi maka perusahaan sulit dalam mengelompokkan pelanggan. Salah satu cara untuk mengatasi permasalahan tersebut adalah dengan menggunakan teknik komputasi algoritme K-Means. Hal tersebut dilakukan karena K-Means terbukti mampu menyelesaikan beberapa masalah oleh beberapa peneliti [1, 2, 3, 4, 5].

Penelitian yang dilakukan oleh Metisen dan Sari [1] tentang pengelompokkan produk swalayan dengan algoritme K-Means berhasil mengelompokkan produk penjualan menjadi dua *cluster* yaitu *cluster* penjualan tinggi dan rendah. Nilai *cluster* dihitung berdasarkan harga pokok, jumlah barang terjual dan stok barang. Sehingga swalayan dapat mengetahui barang yang laris terjual dan tidak laris agar tidak terjadi penumpukan stok barang. Selanjutnya pengelompokkan dengan K-Means juga dilakukan oleh Lubis,dkk [2] pada program Sistem Database Pemasarakatan (SDP) lanjutan pegawai. Penelitian tersebut berhasil mengelompokkan program SDP menjadi tiga *cluster* yaitu *cluster* lolos, hampir lolos dan tidak lolos. Penentuan *cluster* didasari dengan kriteria tertentu diantaranya kedisiplinan pegawai, target kerja pegawai dan kepatuhan program SDP. Sehingga pihak pelaksana program SDP dapat mengambil keputusan dalam penentuan pegawai yang layak mengikuti program SDP lanjutan. Di samping itu, penelitian oleh Khusnu, dkk [3] pada pengelompokkan kualitas beras dengan algoritme K-Means mampu menghasilkan tiga *cluster* yaitu *cluster* 1 (5,89333;2,05), *cluster* 2 (6,28199;2,546), dan *cluster* 3 (6,96583;2,999167) serta dihasilkan validasi sebesar 92,82%. Nilai *cluster* dihitung berdasarkan warna dan ukuran beras. Penelitian tersebut dapat dijadikan sebagai acuan dalam klasterisasi kualitas beras. Kemudian penelitian tentang produk *online shop* oleh Muningsih dan Kuswati [4] menggunakan algoritme K-Means untuk menentukan stok barang. *Clustering* pada penelitian tersebut diperoleh berdasarkan kode produk, jumlah transaksi, volume penjualan dan rata-rata penjualan. Sehingga pelaku bisnis dapat menampilkan kategori produk yang diinginkan dalam jangka waktu tertentu. Kemudian penelitian oleh Prasetyo,dkk [5] mengelompokkan calon mahasiswa baru dengan algoritme K-Means mampu menghasilkan empat *cluster* diantaranya *cluster* 0 (15%), *cluster* 1 (33%), *cluster* 2 (30%), dan *cluster* 3 (22%). *Cluster* tersebut diperoleh berdasarkan minat siswa. Sehingga penelitian ini dapat menjadi acuan bagi pihak universitas dalam melakukan strategi promosi ke calon mahasiswa baru.

Teknik-teknik tersebut [1,2,3,4,5] terbukti mampu memberi berbagai macam solusi untuk mengatasi permasalahan yang ada. Salah satunya dengan menggunakan teknik komputasi dalam *clustering* data penjualan ayam untuk mengetahui tingkat penjualan rendah, sedang dan tinggi. Pengelompokan ini bertujuan untuk memberikan informasi penting bagi perusahaan dalam menentukan pelanggan yang akan diprioritaskan perusahaan.

Detail penelitian ini dijelaskan pada bagian II tentang Algoritme K-Means, bagian III tentang Implementasi algoritme K-Means, dan bagian IV tentang Hasil dan pembahasan serta bagian V tentang kesimpulan.

## II. ALGORITME K-MEANS

Algoritme K-Means merupakan suatu metode paling sederhana dalam pengelompokan dokumen yang awalnya mengambil sebagian dari banyaknya komponen data untuk dijadikan sebagai pusat *cluster* awal [6]. Algoritme K-Means dikenalkan oleh J.B Macqueen pada tahun 1976 [7]. Prinsip utama K-Means adalah menyusun *k prototype* atau pusat massa (*centroid*) dari sekumpulan data berdimensi *n*. Menurut [8] Algoritme K-Means dikenal sebagai suatu metode pengelompokan yang paling terkenal diberbagai bidang karena sangat mudah diimplementasi dan mampu mengelompokkan data dalam jumlah yang banyak. Algoritme K-Means juga sangat akurat dalam pengelompokan data [9]. Tahapan dalam perhitungan algoritme K-Means yaitu tentukan banyaknya *cluster* berdasarkan *cluster* data yang akan dibentuk. Selanjutnya tentukan *centroid* atau pusat *cluster* awal secara acak. Kemudian hitung setiap data dan kelompokkan ke *cluster* terdekat berdasarkan jarak antara data dengan pusat *cluster* awal. Lalu Hitung kembali *centroid* dengan keanggotaan *cluster* yang sekarang. Selanjutnya pada setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses Perhitungan selesai. Pengelompokan dalam perhitungan jarak data dari masing-masing *centroid* ditunjukkan pada persamaan 1

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Keterangan :

$D(i,j)$  = Jarak data ke *i* ke pusat *cluster* *j*

$x_{ki}$  = Data ke *i* pada atribut data ke *k*

Setelah memperoleh hasil perhitungan iterasi, maka perhitunganaan pusat *cluster* baru (iterasi 2) berdasarkan pengelompokan data pada iterasi 1 dapat dilihat pada persamaan 2

$$T(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum n} \quad (2)$$

Keterangan :

$T(i)$  = *Centroid* data

$x_1$  = Data ke *i* pada atribut data ke *k*

$\sum n$  = Jumlah Data

## III. IMPLEMENTASI ALGORITME K-MEANS UNTUK DATA PENJUALAN

### A. Data Penelitian.

Penelitian ini menggunakan data penjualan ayam broiler yang dimiliki PT Surya Unggas Mandiri. Data penjualan terdiri dari 136 data dan beberapa atribut ntara lain atribut tanggal, SPA, nama pelanggan, total ayam, rata-rata, berat (kg) dan jumlah kematian. Atribut yang digunakan dalam mengelompokkan data penjualan ayam yaitu atribut nama pelanggan, frekuensi dan jumlah ekor yang dibeli. Hal ini dikarenakan informasi yang terkandung didalamnya sudah mewakili informasi yang dibutuhkan untuk di jadikan indikator penentu dalam proses pengelompokkan menggunakan algoritme K-Means. Data yang telah diakumulasi ditunjukkan pada Tabel 1.

Tabel 1 Data Penjualan Ayam

No	SPA	Frekuensi	Total Ekor yang diambil
1	P001	19	15051
2	P002	10	7700
3	P003	9	6610

No	SPA	Frekuensi	Total Ekor yang diambil
4	P004	3	2435
5	P005	9	5550
6	P006	21	14951
7	P007	1	600
8	P008	11	15996
9	P009	1	1175
10	P010	1	488
.			
.			
.			
22	P022	1	650

### B. Tahap Implementasi Algoritme K-Means

Tahap implementasi algoritme K-Means dilakukan dengan perhitungan data menggunakan tiga *cluster* yaitu rendah, sedang dan tinggi. Berikutnya menentukan tiga nilai awal yang dipilih secara acak (data ke-4,8, dan 16). Setelah menentukan nilai awal maka melakukan perhitungan pada setiap data *centroid* agar menghasilkan jarak dan pengelompokkan ini dilakukan secara manual. Tahapan perhitungan data penjualan ayam broiler secara manual menggunakan *centroid* iterasi 1 berikut contoh perhitungannya jarak data dengan *cluster* terendah dengan menggunakan persamaan 1.

$$D(x_1, T_0) = \sqrt{(a_1 - c_{0a})^2 + (b_1 - c_{0b})^2} = \sqrt{((19 - 1)^2 + (15051 - 322)^2)} = \mathbf{14729.010}$$

$$D(x_2, T_0) = \sqrt{(a_2 - c_{0a})^2 + (b_2 - c_{0b})^2} = \sqrt{((10 - 1)^2 + (7700 - 322)^2)} = \mathbf{7378.005489}$$

Selanjutnya tahapan perhitungan data penjualan ayam broiler secara manual menggunakan *centroid* pada iterasi 2. berikut contoh perhitungannya jarak data ke *cluster* sedang dengan menggunakan persamaan 1.

$$D(x_1, T_0) = \sqrt{(a_1 - c_{0a})^2 + (b_1 - c_{0b})^2} = \sqrt{(19 - 1.125)^2 + (15051 - 809.875)^2} = \mathbf{14241.13622}$$

$$D(x_2, T_0) = \sqrt{(a_2 - c_{0a})^2 + (b_2 - c_{0b})^2} = \sqrt{(10 - 1.125)^2 + (7700 - 809.875)^2} = \mathbf{6890.130716}$$

Kemudian tahapan perhitungan data penjualan ayam broiler secara manual menggunakan *centroid* pada iterasi 3. berikut contoh perhitungannya jarak data ke *cluster* tinggi dengan menggunakan persamaan 1.

$$D(x_1, T_0) = \sqrt{(a_1 - c_{0a})^2 + (b_1 - c_{0b})^2} = \sqrt{(19 - 1.583333)^2 + (15051 - 1213.583)^2} = \mathbf{13837.42796}$$

$$D(x_2, T_0) = \sqrt{(a_2 - c_{0a})^2 + (b_2 - c_{0b})^2} = \sqrt{(10 - 1.583333)^2 + (7700 - 1213.583)^2} = \mathbf{6486.422461}$$

Proses K-Means akan terus beriterasi sampai pengelompokkan data sama dengan pengelompokkan data iterasi sebelumnya. proses akan terus melakukan iterasi sampai data pada iterasi terakhir samadengan iterasi sebelumnya. Setelah memperoleh hasil perhitungan iterasi, maka perhitungaan pusat *cluster* baru (iterasi 2) berdasarkan pengelompokkan data pada iterasi 1 dengan menggunakan persamaan 2.

$$T_{10} = \frac{1+1+1+1+1+1+2+1+1}{8} = \mathbf{1.125}$$

$$T_{11} = \frac{10+9+3+9+6+1+6+4+8+2}{10} = \mathbf{5.8}$$

$$T_{12} = \frac{19+21+11+17}{4} = \mathbf{17}$$

$$T_{20} = \frac{600+1.175+488+322+1.259+1.285+700+650}{8} = \mathbf{809.875}$$

$$T_{21} = \frac{7700 + 6.610 + 2.435 + 5.550 + 5.774 + 1.500 + 6.040 + 2.194 + 7.240 + 1.955}{10} = \mathbf{4699.8}$$

$$T_{22} = \frac{15051 + 14951 + 15996 + 19345}{4} = \mathbf{16335.75}$$

Selanjutnya tahapan pada iterasi 1 dan 2 diperoleh *cluster* yang berbeda pada data penjualan ayam broiler sehingga iterasi akan terus berlanjut dan perhitungan pusat *cluster* baru (iterasi 3) berdasarkan pengelompokan data pada iterasi 2 dengan menggunakan persamaan 2.

$$T10 = \frac{3+1+1+1+1+1+4+1+1+2+1+2+1}{12} = 1.583333$$

$$T11 = \frac{10+9+9+6+6+8}{6} = 8$$

$$T12 = \frac{19+21+11+17}{5} = 17$$

$$T20 = \frac{2435+600+1175+488+1500+2194+322+700+1259+1285+700+1955+650}{8} = 1213.583$$

$$T21 = \frac{7700+6610+5550+5774+6040+7240}{6} = 6485.667$$

$$T22 = \frac{15051+14951+15996+19345}{4} = 16335.75$$

Berdasarkan persamaan 1 dan persamaan 2, maka keseluruhan algoritme K-Means ditunjukkan pada algoritme 1 [10].

**Algoritme 1, Algoritme K-Means**

**Input :**

D = {  $x_1, x_2, x_3 \dots x_n$  } // untuk mengatur elemen

C = 3 // jumlah *cluster* yang digunakan

**Output :**

C = 3 // set dari *cluster*

**K-Means Algorithm**

Menentukan pusat *cluster* T1,T2,...Tn

**Repeat**

Menentukan nama barang setiap untuk *cluster* yang memiliki rata-rata ;

Menghitung rata-rata baru untuk masing-masing;

**Until** pengelompokan terpenuhi;

C. Evaluasi

Penelitian ini, menghasilkan pengelompokan pelanggan berdasarkan data penjualan ayam dengan algoritme K-Means yang dilakukan perhitungan manual menggunakan Microsoft Excel. Lalu evaluasi menggunakan teknik komputasi sebagai pembandingan dengan perhitungan manual.

**IV. HASIL DAN PEMBAHASAN**

Perhitungan manual menggunakan Microsoft Excel dilakukan sebanyak tiga pengulangan. Hasil perhitungan manual pengelompokan untuk iterasi 1 ditunjukkan pada Tabel 2.

Tabel 2 Pengelompokan Iterasi 1

No	SPA	Frekuensi	Total Ekor yang diambil	dc0	dc1	dc2
1	P001	19	15051	14729.011	12616.01015	<b>4290</b>
2	P002	10	7700	7378.005489	<b>5265.004653</b>	11596
3	P003	9	6610	6288.005089	<b>4175.004311</b>	12671
4	P004	3	2435	2113.000947	<b>0</b>	16714
5	P005	9	5550	5228.006121	<b>3115.005778</b>	13731
6	P006	21	14951	14629.01367	12516.01294	<b>4378</b>
7	P007	1	600	<b>278</b>	1835.00109	18489
8	P008	11	15996	15674.00319	13561.00236	<b>3313</b>
9	P009	1	1175	<b>853</b>	1260.001587	17914
10	P010	1	488	<b>166</b>	1947.001027	18601
.						
.						
.						
22	P022	1	650	<b>328</b>	1785.00112	18439

Proses K-Means akan terus beriterasi sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya. Penentuan *centroid* pada iterasi-iterasi tergantung pada nilai atau jarak terkecil. Sehingga pada iterasi 1 dijelaskan bahwa kode P001, frekuensi 19 dan total penjualan sebanyak 15051 memperoleh jarak *centroid* terkecil yaitu 4290. Selanjutnya pada iterasi

kedua dan ketiga memperoleh jarak *centroid* yang sama yaitu 1284.751557, sehingga perhitungan iterasi selesai. Pengelompokan iterasi kedua ditunjukkan pada Tabel 3.

Tabel 3 Pengelompokan Iterasi 2

No	SPA	Frekuensi	Total Ekor yang diambil	dc0	dc1	dc2
1	P001	19	15051	14241.13622	10351.20842	<b>1284.751557</b>
2	P002	10	7700	6890.130716	<b>3000.20294</b>	8635.752837
3	P003	9	6610	5800.130346	<b>1910.20268</b>	9725.75329
4	P004	3	2435	<b>1625.126082</b>	2264.801731	13900.75705
5	P005	9	5550	4740.131542	<b>850.2060221</b>	10785.75297
6	P006	21	14951	14141.13897	10251.21127	<b>1384.755777</b>
7	P007	1	600	<b>209.8750372</b>	4099.80281	15735.75813
8	P008	11	15996	15186.12821	11296.2012	<b>339.802976</b>
9	P009	1	1175	<b>365.1250214</b>	3524.803268	15160.75844
10	P010	1	488	<b>321.8750243</b>	4211.802735	15847.75808
.						
.						
22	P022	1	650	<b>159.8750489</b>	4049.802845	15685.75816

Pada iterasi 1 dan 2 diperoleh *cluster* data yang berbeda pada data penjualan ayam broiler sehingga iterasi akan terus berlanjut dan perhitungannya pusat *cluster* baru (iterasi 3). Hasil perhitungan manual pengelompokan untuk iterasi 3 ditunjukkan pada Tabel 4.

Tabel 4 Pengelompokan Iterasi 3

No	SPA	Frekuensi	Total Ekor yang diambil	dc0	dc1	dc3
1	P001	19	15051	13837.42796	8565.340063	<b>1284.751557</b>
2	P002	10	7700	6486.422461	<b>1214.334647</b>	8635.752837
3	P003	9	6610	5396.422097	<b>124.3370214</b>	9725.75329
4	P004	3	2435	<b>1221.417822</b>	4050.670086	13900.75705
5	P005	9	5550	4336.423342	<b>935.6675344</b>	10785.75297
6	P006	21	14951	13737.43072	8465.342982	<b>1384.755777</b>
7	P007	1	600	<b>613.5832773</b>	5885.671163	15735.75813
8	P008	11	15996	14782.42	9510.333473	<b>339.802976</b>
9	P009	1	1175	<b>38.58740943</b>	5310.671613	15160.75844
10	P010	1	488	<b>725.5832345</b>	5997.671085	15847.75808
.						
.						
22	P022	1	650	<b>563.5833019</b>	5835.671198	15685.75816

Pada iterasi ke-2 dan iterasi ke-3 posisi *cluster* tidak berubah dan titik pusat dari setiap *cluster* sudah tidak berubah atau tidak ada lagi terdapat data yang berpindah dari satu *cluster* ke *cluster* yang lain. Berdasarkan hasil perhitungan dalam mengelompokkan pelanggan menggunakan teknik komputasi menghasilkan tiga *cluster*. Hasil pengelompokan pelanggan ditunjukkan pada Tabel 5.

Tabel 5. Hasil Pengelompokan Pelanggan

No	SPA	Jumlah	Keterangan
1	P001	15.051	Cluster 2
2	P002	7.700	Cluster 1
3	P003	6.610	Cluster 1
4	P004	2.435	Cluster 0
5	P005	5.550	Cluster 1
6	P006	14.951	Cluster 2
7	P007	600	Cluster 0
8	P008	15.996	Cluster 2
9	P009	1.175	Cluster 0
10	P010	488	Cluster 0
.			
.			
22	P022	650	Cluster 0

Berdasarkan Tabel 5 dapat disimpulkan bahwa pelanggan dengan *cluster* 2 merupakan pelanggan yang diprioritaskan perusahaan. Hal ini dikarenakan pelanggan pada *cluster* 2 memiliki transaksi yang terus meningkat pada setiap bulan. Tiga pelanggan tertinggi pada *cluster* 2 adalah P001, P006 dan P008.

## V. KESIMPULAN DAN SARAN

Berdasarkan uraian yang sudah dikemukakan, maka dapat ditarik kesimpulan yaitu Metode *clustering* dengan algoritme K-Means dapat digunakan untuk mengelompokkan pelanggan. Berdasarkan data yang diolah algoritme tersebut mampu menghasilkan tiga *cluster*. *Cluster* terdiri dari *cluster* rendah yang menghasilkan dua belas pelanggan, *cluster* sedang menghasilkan enam pelanggan dan *cluster* tinggi menghasilkan empat pelanggan. Penelitian ini berhasil untuk menentukan pelanggan yang akan diprioritaskan oleh perusahaan. Sehingga saran untuk penelitian ini yaitu mengembangkan menjadi aplikasi telepon pintar.

### DAFTAR PUSTAKA

- [1] Agustina, S., Yhudo, D., Santoso, H., Marnasusanto, N., Tirtana, A., & Khusnu, F. (2012). *Clustering* Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means. *Clustering* K-Means, 1–7.
- [2] Asroni, A. R. (2015) 'Penerapan Metode K-Means untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang', *Jurnal Ilmiah Semesta Teknik*, 18(1), pp. 76–82. doi: 10.1038/hdy.2009.180.
- [3] Dwi, N. N. 2012. " Penerapan Algoritme K-Means Untuk *Clustering* Data Anggaran Pendapatan Belanja Daerah Di Kabupaten Xyz". Skripsi. Teknik Informatika. Fakultas Teknik, Universitas Negeri: Gorontalo
- [4] Hamzah, A. (2014) 'Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November 2014 ISSN: 1979-911X', *Snast*, 3(November), pp. 211–216. doi: 1979-911X.
- [5] Ong, J. O. (2013). Implementasi Algoritme K-Means *Clustering* Untuk Menentukan Strategi Marketing. *Jurnal Ilmiah Teknik Industri*, 12(Juni), 10–20. <https://doi.org/10.1016/j.jlumin.2012.02.015>
- [6] Sadewo, M. G., Windarto, A. P., & Hartama, D. (2017). Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means *Clustering*. *Jurnal Nasional Informatika Dan Teknologi Jaringan*, 2(1), 60–67. Retrieved from [https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/164/pdf\\_10](https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/164/pdf_10)
- [7] Suparman. 2017. "Potensi Pengembangan Peternakan Ayam Broiler Di Kecamatan Malunda Kabupaten Majene". Skripsi. Peternakan. Fakultas Sains dan Teknologi, Universitas Islam Negri Alauddin: Makassar.
- [8] MacQueen, J. 1967. *Some methods for clasificationand analysis of multivariate observations. In proceedings of the Futh Berkeley Symposium on mathematical Statistic and Probability, Volume 1 : Stastic. Berkeley, Calif : University of California Press.*
- [9] Ong, J. O. (2013). Implementasi Algoritme K-Means *Clustering* Untuk Menentukan Strategi Marketing. *Jurnal Ilmiah Teknik Industri*, 12(Juni), 10–20. <https://doi.org/10.1016/j.jlumin.2012.02.015>
- [10] W.M.W Mobd, A. H. Beg, T. Herawan and K. F. Rabbi, " MaxD K-Means: A Clustering Algorithm for Auto-generation of centroids and distance of data points in cluster, " *Communication In Computer and Information Science*, pp. 192-199.2012