

Penerapan Metode K-Means Clustering Untuk Menentukan Pola Penjualan Kue Pada Alfaza Bakery

Riki Yohanes Hendriyanto¹⁾ Lely Hiryanto²⁾

^{1) 2)} Teknik Informatika, FTI, Universitas Tarumanagara
Jl. Letjen S Parman no 1, Jakarta 11440 Indonesia

¹⁾email : riki.535170079@stu.untar.ac.id ²⁾email : lelyh@fti.untar.ac.id

ABSTRACT

The home industry is a type of small-scale business activity that is often found in villages and around houses, both in urban and rural areas. Starting from an association of the same people who studied to pursue the field of making cakes and bread and who then wanted to expand the sales area and create jobs for residents. the obstacle faced is ignorance of the products that are purchased the most and in which areas certain products run out the fastest, it is necessary to do data mining analysis using the clustering method. The K-Means method is a data clustering method using observation based on the similarity of the objects studied. A cluster is a collection of data that has similarities in its members or is different from other groups, clusters are used to minimize variation within a cluster and maximize variation between clusters, in other words data that has attribute similarities between one another and attribute differences to other clusters, determines the right cluster by using the elbow method which can maximize the quality of clusters so that the clusters are more varied. The results of testing this study with the elbow method obtained the right number of 4 clusters, then the clustering results with the most sales were obtained in cluster 3, cluster 1 with moderate sales, cluster 0 with few sales and cluster 2 with the least sales.

Key words

Clustering, Data Mining, K-Means, Metode Elbow, Penjualan

1. Pendahuluan

Industri rumahan adalah jenis kegiatan usaha berskala kecil yang sering ditemukan pada perkampungan dan sekitar rumah baik di wilayah kota maupun pedesaan. Industri rumahan biasanya memiliki tenaga kerja sebanyak 5-19 orang pekerja, industri rumahan ini melibatkan warga sekitar sebagai pekerjanya dengan tujuan untuk membuka lapangan pekerjaan bagi warga sekitar dan manfaat industri rumah memberikan kontribusi dan perkembangan perekonomian di suatu daerah.

Bermula dari sebuah perkumpulan paguyuban yang sama sama belajar menekuni di bidang pembuat kue dan

roti yang berlokasi di Jawa Timur, dalam paguyuban tersebut masing masing anggota akan dibiayai oleh investor jika ingin membuka cabang usaha sendiri. Rahmat Sulaeman mencoba untuk merintis usaha baru dengan nama Alfaza Bakery pada tahun 2018 dengan tujuan untuk memperluas area penjualan dan membuka lapangan pekerjaan bagi warga sekitar. Alfaza Bakery memproduksi 8 jenis kue, di antaranya 7 kue basah dengan berbagai rasa seperti coklat, pisang, stroberi, nanas, vanilla, blueberry dan mangga, dan 1 jenis kue kering, dimana diantaranya: produk Tawar Casino yang merupakan roti tawar, Spesial Jumbo kombinasi berbagai rasa seperti coklat, nanas, navila dan blueberry, produk Spesial 9.000 merupakan kombinasi berbagai rasa seperti coklat, stroberi, dan bluberi, produk Spesial 7.000 merupakan kombinasi rasa vanilla dan mangga, produk Pisang Keju, produk Roti Kering, produk Roti Cup dengan isian 1 rasa, dan produk Kecil yang merupakan kue dengan 1 rasa.

Sistem produksi dari Alfaza Bakery yaitu sehari memproduksi sehari pengiriman, sistem pengiriman ke warung – warung dilakukan seminggu sekali, jika ada produk yang tidak terjual maka akan diambil kembali dan diganti dengan yang baru. Alfaza Bakery masih menggunakan sistem penjualan konvensional, dimana transaksi jual beli dilakukan dengan pencatatan nota manual, pemilik usaha ini merasa bahwa sistem penjualan yang seperti ini memiliki beberapa kekurangan seperti sistem pendataan transaksi secara manual, proses pendataan masih dilakukan secara manual dapat memungkinkan kesalahan pencatatan. Selain itu kendala yang dihadapi adalah ketidaktahuan terhadap produk yang paling sering dibeli dan di area mana saja produk tertentu yang paling cepat habisnya, selama ini Alfaza Bakery hanya mengandalkan nota dan hanya perkiraan saja sehingga target penjualan tidak efisien dalam menentukan pendistribusian produk.

Metode *K-Means* adalah salah satu metode data clustering non hirarki, yang berfungsi untuk mempartisi data kedalam bentuk satu atau kelompok, dengan cara observasi berdasarkan kemiripan objek yang diteliti. *Cluster* adalah suatu kumpulan data yang memiliki kemiripan pada anggotanya atau berbeda dengan kelompok lain, *cluster* digunakan untuk meminimalisasi variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*, dengan kata lain data yang memiliki kesamaan atribut antara satu sama lain dan perbedaan atribut terhadap *cluster* lainnya[1]

Dalam penelitian yang dilakukan oleh Angga Saputra[2], Penelitian ini menggunakan metode *K-Means Clustering* dan *Simple Additive Weighting*, tujuan dari penelitian ini adalah untuk mencari tempat wisata kuliner dengan jarak terdekat yang berada di Jakarta. Dengan metode *K-Means Clustering* dapat mengelompokkannya tempat kuliner berdasarkan jarak terdekat, kemudian dilakukan perhitungan menggunakan metode *Simple Additive Weighting* dengan menambah kategori lebih spesifik seperti fasilitas yang dimiliki oleh tempat kuliner tersebut di antaranya fasilitas lengkap (wifi, AC, musik, kontak listrik dan parkir), fasilitas cukup (AC dan parkir) dan fasilitas kurang (AC, parkir), dari ketiga cluster tersebut dipilih cluster dengan fasilitas terlengkap. Berikut ini adalah hasil dari perhitungan.

Tabel 1. Hasil perhitungan

Alternatif	Urutan
C15	0.97
C10	0.95
C3	0.90
C8	0.90
C1	0.85
C9	0.83
C7	0.83
C6	0.83
C11	0.79
C4	0.78
C14	0.75
C12	0.73
C13	0.72
C2	0.70
C5	0.70

Maka didapatkan tempat wisata kuliner dengan nilai tertinggi yaitu C15 dengan nilai 0.97. Rancangan yang saat ini sedang dilakukan akan menggunakan metode *K-Means Clustering* untuk mengelompokkan data dan untuk menentukan *cluster* yang tepat dengan menggunakan metode *elbow* yang dapat memaksimalkan kualitas *cluster* sehingga *cluster* lebih bervariasi, tujuannya adalah untuk melihat kualitas seberapa baik objek data dalam penempatan di suatu *cluster*.

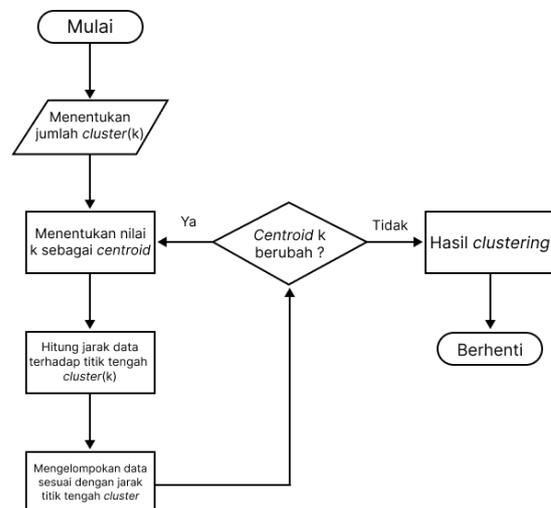
2. Metode Penelitian

Perancangan sistem yang akan dibuat menggunakan metode *K-Means Clustering*. Sistem ini dibuat untuk menentukan pola penjualan dengan cara mengelompokkan data penjualan. Akan dilakukan secara sistematis menjelaskan proses kerja penelitian yang

dilaksanakan. Dalam melakukan penelitian, membutuhkan data, proses pengumpulan data dan analisis data menggunakan metode *K-Means Clustering*.

2.1 K-Means Clustering

Metode *K-Means Clustering* adalah salah satu metode data *clustering* non hirarki, yang berfungsi untuk mempartisi data kedalam bentuk satu atau kelompok, dengan cara observasi berdasarkan kemiripan objek yang diteliti. *Cluster* adalah suatu kumpulan data yang memiliki kemiripan pada anggotanya atau berbeda dengan kelompok lain, *cluster* digunakan untuk meminimalisasi variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*, dengan kata lain data yang memiliki kesamaan atribut antara satu sama lain dan perbedaan atribut terhadap *cluster* lainnya.[1] Berikut ini merupakan tahapan dari metode *K-Means Clustering*.



Gambar 1. Tahapan metode K-Means Clustering

1. Mempersiapkan data yang akan diproses ke dalam algoritma *K-Means*.
2. Menentukan jumlah *cluster* (k), *cluster* merupakan suatu kelompok yang berisikan kumpulan data, tiap *cluster* memiliki keunikan tersendiri sehingga dapat dibedakan masing-masing *cluster*. Menentukan jumlah *cluster* dapat disesuaikan dengan kebutuhan data yang ada, sebagai contoh pada Alfaza Bakery ini ingin mengetahui produk apa saja yang paling laku terjual, sehingga jumlah *cluster* ada 3, yaitu penjualan paling banyak, penjualan sedang dan penjualan paling sedikit.

- Setelah menentukan jumlah *cluster* (k), menentukan nilai k sebagai titik tengah atau *centroid* dari *cluster*, biasanya dilakukan secara acak.
- Hitung jarak data dengan *centroid* menggunakan rumus *Euclidean* (persamaan I).

Euclidean (persamaan I).

Distance:

$$d(x_i, u_j) = \sqrt{\sum_{i=1}^N (x_i - u_j)^2} \quad (1)$$

Keterangan :

$d(x_i, u_j)$ = Jarak objek antara nilai data dengan titik tengah *cluster*.

x_i = Nilai data.

u_j = Titik tengah *cluster* ke- j .

N = Jumlah data.

- Selanjutnya, mengelompokkan data yang sudah dihitung ke masing masing *cluster* yang sudah di tentukan dengan keunikan tersendiri dan kedekatan data dengan titik tengah *cluster*.
- Hitung titik tengah *cluster* ke- j yang baru berdasarkan data yang berada di *cluster* ke- j

$$u_j(t + 1) = \frac{1}{N_{s_j}} \sum_{i=1}^{s_j} x_i \quad (2)$$

Keterangan :

$u_j(t + 1)$ = Titik tengah baru yang didapatkan dari perhitungan iterasi pertama.

N_{s_j} = Banyaknya data pada s_j .

u_j = Merupakan titik tengah *cluster* ke- j .

Dimana nilai $u_j(t+1)$ merupakan titik tengah baru pada iterasi pertama dan N_{s_j} merupakan banyaknya data pada *cluster* s_j .

- Kemudian lanjut ke iterasi ke 3 sampai 4 atau hingga tiap data tidak lagi memiliki perubahan pada titik tengah *cluster*.
- Perhitungan data telah selesai, dan hasil sudah sesuai dengan yang diinginkan.

Perhitungan data penjualan Alfaza Bakery dalam menentukan pola penjualan, data diambil di tahun 2021 pada bulan Januari di daerah Jakarta. Atribut dari data adalah jumlah produk yang terjual.

Tabel 2. Data penjualan bulan Januari 2021

Januari - Februari 2021 Jakarta	
Nama Produk	Januari
TAWAR CASINO	128
SPECIAL JUMBO	141
SPECIAL 9.000	110
SPECIAL 7.000	148
PISANG KEJU	137
ROTI KERING	99
ROTI CUP	135
KECIL	107

Setelah data tersedia, langkah awal algoritma k-means menentukan *cluster* terlebih dahulu sebagai berikut :

- Menentukan jumlah *cluster*, dimana jumlahnya adalah 3.
- Inisialisasi titik tengah secara acak berdasarkan data yang tersedia, berikut tabel titik tengah *cluster*:

Tabel 3. Titik tengah (centroid)

C1	C2	C3
128	141	110

- Menghitung semua jarak data dengan titik tengah *cluster* dengan rumus di persamaan (1):

Contoh perhitungan *centroid* 1:

$$d_{1,1} = \sqrt{(128 - 128)^2} = 0$$

$$d_{2,1} = \sqrt{(141 - 128)^2} = 13$$

$$d_{3,1} = \sqrt{(101 - 128)^2} = 18$$

$$d_{4,1} = \sqrt{(148 - 128)^2} = 20$$

$$d_{5,1} = \sqrt{(137 - 128)^2} = 9$$

$$d_{6,1} = \sqrt{(99 - 128)^2} = 29$$

$$d_{7,1} = \sqrt{(135 - 128)^2} = 7$$

$$d_{8,1} = \sqrt{(107 - 128)^2} = 21$$

Lakukan hal yang sama seperti perhitungan *centroid* 1 pada *centroid* 2 dan 3 hingga selesai. Berikut ini merupakan tabel perhitungan (iterasi 1):

Tabel 4. Hasil perhitungan jarak data dengan cluster

Data ke -	Atribut	Cluster 1	Cluster 2	Cluster 3	Jarak Terdekat	Keterangan
	Value	128	141	110		
1	128	0	13	18	0	Cluster 1
2	141	13	0	31	0	Cluster 2
3	110	18	31	0	0	Cluster 3
4	148	20	7	38	7	Cluster 2
5	137	9	4	27	4	Cluster 2
6	99	29	42	11	11	Cluster 3
7	135	7	6	25	6	Cluster 2
8	107	21	34	3	3	Cluster 3

- Setelah didapatkan hasil perhitungan iterasi pertama, masing masing data berkelompok sesuai dengan jarak titik tengah cluster.
- Mencari titik tengah baru dengan melihat data yang ada di dalam *cluster*

Cluster 1 = {1} dengan value {(128)}

Cluster 2 = {2,4,5,7} dengan value {(141,148,137,135)}

Cluster 3 = {3,6,8} dengan value {(110,99,107)}

Cara menentukan *cluster* baru dengan menjumlahkan value dan dibagi banyaknya data menggunakan persamaan (2).

Cluster 1

$$value\ 1 = \frac{128}{1} = 128$$

Cluster 2

$$value\ 1 = \frac{141+148+137+135}{4} = 140,25$$

Cluster 3

$$value\ 1 = \frac{110+99+107}{3} = 105,33$$

Berikut ini merupakan hasil perhitungan titik tengah baru.

Tabel 5. Cluster baru

C1	C2	C3
128	140,25	105,33

Perhitungan ulang seperti langkah ke 4 pada iterasi ke 2 hingga tidak kelompok data yang berubah lagi.

- Hasil akhir *cluster* di dapatkan pada iterasi ke 2

Hasil perhitungan *cluster* akhir berhenti di iterasi ke 2 dapat dilihat titik tengah dari *cluster*, C1 dengan nilai cluster 128, C2 140,25 dan C3 105,33.

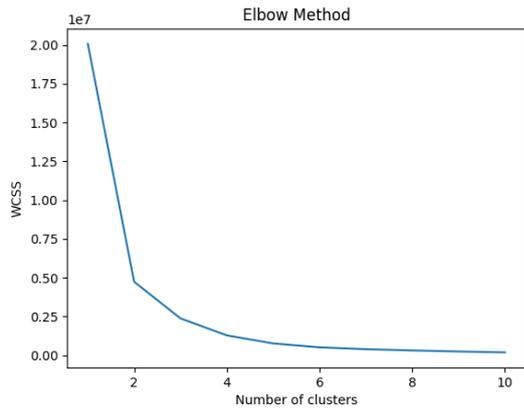
Tabel 6. Hasil akhir cluster

Data ke -	Atribut	Cluster 1	Cluster 2	Cluster 3	Jarak Terdekat	Keterangan
	Value	128	140,25	105,33		
1	128	0	12,25	22,67	0	Cluster 1
2	141	13	0,75	35,67	0,75	Cluster 2
3	110	18	30,25	4,67	4,67	Cluster 3
4	148	20	7,75	42,67	7,75	Cluster 2
5	137	9	3,25	31,67	3,25	Cluster 2
6	99	29	41,25	6,33	6,33	Cluster 3
7	135	7	5,25	29,67	5,25	Cluster 2
8	107	21	33,25	1,67	1,67	Cluster 3

Kesimpulan dari hasil contoh perhitungan tabel iterasi ke 2, bahwa *cluster* 2 merupakan kelompok data dengan produk yang terjual paling banyak sedangkan *cluster* 3 merupakan kelompok dengan penjualan sedang dan *cluster* 1 kelompok dengan penjualan paling sedikit.

3. Hasil Percobaan

Hasil pengujian dilakukan dengan mengelompokan data yang menghasilkan 4 *cluster*, dengan menggunakan data penjualan. Hasil yang di dapakannya sebagai berikut



Gambar 2. Elbow cluster

Dengan melihat *Elbow* dari data penjualan tersebut dapat untuk menentukan jumlah *cluster* yang terbaik dan memaksimalkan kualitas *cluster* sehingga *cluster* lebih bervariasi, dalam penelitian ini di dapatkan bahwa jumlah *cluster* yang terbaik adalah 4.

Tabel 7. Penjualan masing-masing cluster

Penjualan	Cluster 0	Cluster 1	Cluster 2	Cluster 3
MIN	176	303	87	478
MAX	300	471	175	696

Kategori *cluster* 3 penjualan paling banyak, *cluster* 1 dengan penjualan sedang, *cluster* 0 penjualan sedikit dan *cluster* 2 dengan penjualan paling sedikit.

Tabel 8. Penjualan produk

Penjualan	Cluster 0	Cluster 1	Cluster 2	Cluster 3
MIN	1	1	1	1
MAX	8	8	8	8

Setelah di analisis produk dengan penjualan paling banyak di tiap *cluster* di dapatkan bahwa produk Tawar Casino paling sedikit penjualannya dan produk Roti Kecil paling banyak penjualannya.

Tabel 9. Jumlah data tiap cluster

Jumlah Data	Cluster 0	Cluster 1	Cluster 2	Cluster 3
	237	188	449	86

Dapat di lihat dari tabel di atas bahwa jumlah data tiap *cluster* berbeda beda, jumlah data dalam satu *cluster* yang paling banyak bukan berarti *cluster* tersebut memiliki penjualan paling banyak.

4. Kesimpulan

Kesimpulan harus mengindikasikan secara jelas hasil-hasil yang diperoleh, kelebihan dan kekurangannya, serta kemungkinan pengembangan selanjutnya.

Kesimpulan dapat berupa paragraf, namun sebaiknya berbentuk point-point dengan menggunakan numbering atau bullet.

Berdasarkan penelitian yang telah dilakukan terhadap pengelompokan data tersebut dapat diperoleh kesimpulan sebagai berikut :

1. Dengan menggunakan metode *elbow* dapat memaksimalkan *cluster* yang optimal, jumlah *cluster* yang di dapatkan adalah 4.
2. Jumlah data pada tiap *cluster* bervariasi dapat dilihat bahwa pada *cluster* 0 berjumlah 237 data, *cluster* 1 berjumlah 188 data, *cluster* 2 berjumlah 449 dan *cluster* 3 berjumlah 86. Jumlah data dalam satu *cluster* yang paling banyak bukan berarti *cluster* tersebut memiliki penjualan paling banyak.
3. Dari 5 daerah penjualan, produk yang paling banyak terjual dan paling sedikit penjualannya adalah produk Roti kecil dengan kode produk 8 dan Tawar Casino dengan kode produk 1.
4. Hasil dari analisis *clustering* di dapatkan bahwa *cluster* yang paling banyak penjualannya adalah *cluster* 3, kemudian *cluster* 1 dengan penjualan sedang, *cluster* 0 penjualan sedikit dan *cluster* 2 penjualan paling sedikit.

Berdasarkan beberapa kesimpulan yang terdapat di atas. Penulis berharap semoga penelitian ini bisa sedikit bermanfaat bagi yang menggunakannya dan juga sesuai dengan apa yang ditulis dengan latar belakang

REFERENSI

[1] Sani, Asrul, Penerapan Metode K-Means Clustering Pada Perusahaan, https://www.researchgate.net/publication/326849650_PENERAPAN_METODE_KMEANS_CLUSTERING_PADA_PERUSAHAAN, 30 September 2022.

[2] Angga Saputra "Rekomendasi Lokasi Wisata Kuliner Di Jakarta Menggunakan Metode K-Means Clustering dan Simple Additive Weighting", Jurnal Ilmu Komputer dan Sistem Informasi, Vol.VII Nomor.1, (Februari, 2019), h.14-20.

Riki Yohanes Hendriyanto, saat ini sebagai mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.