

# Analisis *Clustering* Pengelompokan Penjualan Paket Data Menggunakan Metode K-Means

Dimas Galang Ramadhan<sup>1</sup>, Indri Prihatini<sup>2</sup>, Febri Liantoni<sup>3</sup>

<sup>1,2,3</sup> Pendidikan Teknik Informatika dan Komputer, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sebelas Maret

Diterima 01 April 2021

Disetujui 13 Juni 2021

**Abstract**— At present, the COVID-19 pandemic requires all activities based in the network, starting from work, college, school, and everything based on the network. Certain *provider* users will experience excessive data plan usage. This also affects a counter that sells data packages, which must provide several data package services following current conditions. This research was conducted to analyze the grouping of sales of data packages that are often purchased by customers in a counter by using the K-Means method. The K-Means method is used because the K-Means algorithm is not influenced by the order of the objects used, this is proven when the writer tries to determine the initial *cluster* center randomly from one of the objects in the first calculation. sales of data packages at a counter. Variables used include Price, Active period, and several data packages. The K-Means *Cluster* Analysis algorithm is basically applied to the problem of understanding consumer needs, identifying the types of data package products that are often purchased. The K-Means algorithm can be used to describe the characteristics of each group by summarizing a large number of objects so that it is easier.

**Index terms**—K-Means, *Clustering*, Data Mining

## I. PENDAHULUAN

Salah satu langkah dalam knowledge discovery databases adalah Data Mining. Data Mining merupakan teknik analisis data untuk mendapatkan informasi yang tersembunyi dari data yang kompleks dalam jumlah besar. Salah satu teknik analisa Data Mining adalah *Cluster* Analysis (Analisis Kelompok) atau dikenal sebagai *Clustering*. *Clustering* merupakan sebuah metode analisis dengan tujuan mengelompokkan data berdasarkan karakteristik objek pada suatu wilayah yang sama. Algoritma K-Means digunakan dalam mengembangkan metode *Clustering*, algoritma ini merupakan algoritma pengelompokan data secara nonhierarki (sekatan) yang membagi data kedalam dua atau lebih kelompok (*cluster*) dengan karakteristik sama.

Pada penelitian sebelumnya metode *clustering* digunakan untuk mengelompokkan daun berdasarkan fitur moment invarian. Pada penelitian ini metode *clustering* yang digunakan yaitu *Hierarchical Clustering* dan K-Means [1], [2].

Konter merupakan salah satu bidang usaha penjualan pulsa dan paket data. Dimana untuk situasi pandemi sekarang ini paket data banyak sekali dicari untuk kebutuhan aktivitas dalam jaringan, penggunaan *provider* dan jumlah paket data yang sering digunakan konsumen membuat bidang usaha ini harus menentukan jenis *provider*, jumlah paket data dan harganya yang akan diperbanyak stoknya berdasarkan penggunaan pelanggan. Oleh karena itu diperlukan data yang cukup agar bisa dianalisis lebih lanjut dan menghasilkan sebuah informasi yang tepat.

Berdasarkan hal tersebut diharapkan sebuah konter dapat menentukan suatu pengelompokan untuk menentukan *provider*, jumlah paket data, serta harga yang akan di perbanyak stoknya. Penulis juga mengupayakan agar bisa memberikan rangsangan kepada pengolah data supaya bisa mengumpulkan informasi lebih banyak yang dibutuhkan oleh sebuah konter, sehingga bisa menghasilkan pilihan yang tepat untuk menjawab kebutuhan dari sebuah Konter. Hasil penelitian ini akan memberikan informasi pilihan paket favorit yang diminati.

## II. METODE PENELITIAN

### A. *Clustering*

*Clustering* adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya [2], [3]. *Cluster* adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam *cluster* yang sama dan disimilari terhadap objek-objek yang berbeda *cluster* [4][1].

Data dari *clustering* memiliki tujuan, yaitu untuk meminimalisasikan objektif function yang diatur dalam proses *clustering*, dimana pada umumnya berusaha meminimalisasikan variasi dalam suatu *cluster* dan antar *cluster*.

### B. Algoritma Klasterisasi K-Means

Metode algoritma K-Means digunakan pada penelitian ini untuk menentukan kelompok (*cluster*) yang paling baik untuk pemilihan *provider*, jumlah paket data, serta harga yang akan diperbanyak di Konter Plus Ultra Pabelan. Sehingga peluang untuk menambah

penghasilan penjualan semakin banyak dengan menambah stok *provider* yang paling banyak dibeli.

Metode K-Means merupakan metode pengelompokan data berdasarkan titik pusat *cluster* (centroid) [5]. Metode K-Means bertujuan untuk mengelompokan data dengan karakteristik yang sama dalam satu *cluster*.

Langkah awal proses *clusterisasi* data dengan menggunakan metode K-Means adalah penentuan titik pusat awal  $c_j$ , pada umumnya penentuan titik awal centroid didapatkan secara acak [5], [6]. Banyaknya centroid  $c_j$  yang didapatkan sesuai dengan jumlah *cluster* yang ditentukan di awal. Setelah titik pusat awal terbentuk kemudian langkah selanjutnya menghitung jarak tiap data  $x_i$  dengan titik pusat awal ke- $j$  sampai  $k$ , dinotasikan dengan  $d(x_i, c_j)$ .

Kelanjutan dari jarak data dengan titik pusat tersebut dicari jarak terdekat sehingga data akan dikelompokkan berdasarkan centroid terdekat [7]. Langkah selanjutnya yaitu memperbarui titik centroid dengan cara menghitung rata-rata jarak dari seluruh data dengan centroid. Kemudian kembali lagi ke langkah awal. Iterasi ini dilakukan berulang-ulang hingga mendapatkan titik centroid sudah tidak berubah (konstan).

Metode K-Means digunakan tergantung dari data yang ada dan konklusi yang ingin dicapai dengan aturan sebagai berikut : (a) Jumlah *Cluster* perlu diinputkan. (b) Hanya memiliki atribut numeric [8].

Metode K-Means awalnya mengambil beberapa komponen dari populasi sebagai pusat titik awal atau pusat *cluster* awal [9]. Pada tahapan ini pusat *cluster* ditentukan secara acak serta menandai komponen tersebut ke dalam salah satu pusat *cluster* yang telah didefinisikan, berdasarkan hasil dari jarak minimum data dengan titik pusat *cluster*. Titik pusat awal (*cluster*) selanjutnya dihitung kembali hingga semua komponen data dikelompokkan kedalam masing-masing pusat *cluster* sehingga terbentuk posisi pusat *cluster* yang baru.

Terdapat 2 proses pada algoritma K-Means yaitu proses pencarian anggota dari tiap-tiap *cluster* dan proses pendeteksian lokasi pusat tiap *cluster*. Cara kerja algoritma K-Means (a) Menentukan  $k$  sebagai jumlah *cluster* yang ingin dibentuk. (b) Bagkitkan  $k$  titik pusat *cluster* awal secara acak. (c) Hitung jarak masing –masing data terhadap titik pusat *cluster*. (d) Pilih setiap data dengan titik pusat *cluster* terdekat. (e) Tentkan titik pusat *cluster* baru dengan menghitung rata-rata data yang terletak pada titik awal *cluster* yang sama. (f) Kembali pada langkah ke-3 jika posisi titik awal *cluster* lama dengan yang baru masih berbeda.

### III. HASIL DAN PEMBAHASAN

Pada penelitian ini, terdapat 40 data yang digunakan dalam uji coba, dari 40 hasil data tersebut disusun pada tabel yang terdiri dari nama pembeli, besar kuota, masa aktif, dan harga.

Tabel 1. Data uji coba pembelian paket data

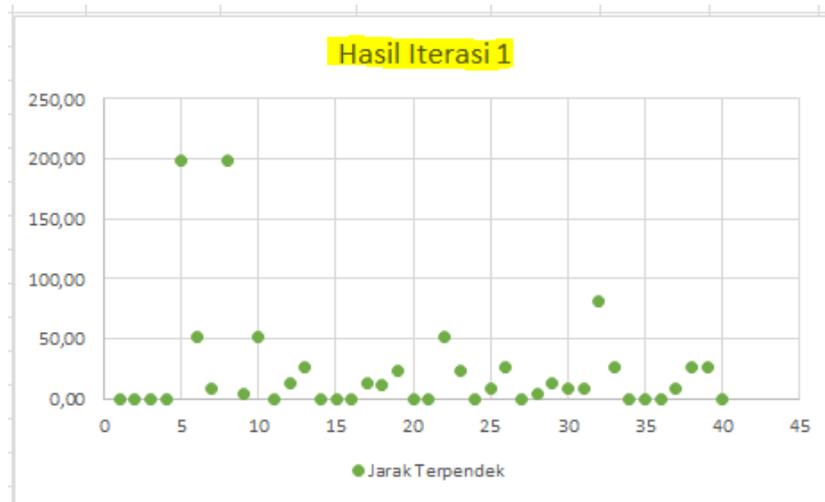
No	Nama Pembeli	Besar Kuota	Masa Aktif	Harga	No	Nama Pembeli	Besar Kuota	Masa Aktif	Harga
1	Irene	2,75	7	10	21	Carol	6	7	35
2	Liam	10	30	50	22	Anna	25	30	100
3	Austin	10	30	50	23	Blake	2	30	15
4	Tracey	2,75	7	10	24	Christopher	2,75	7	10
5	Mary	70	30	239	25	Brian	10	30	59
6	Tracey	25	30	100	26	Rose	8	30	76
7	Nicholas	10	30	59	27	Felicity	6	7	35
8	Christopher	70	30	239	28	Justin	3	7	14
9	Deirdre	3	7	14	29	Richard	2	30	39
10	Austin	25	30	100	30	Luke	10	30	59
11	Victor	10	30	50	31	Theresa	10	30	59
12	Samantha	2	30	39	32	Ava	30	30	129
13	Ryan	8	30	76	33	Harry	8	30	76
14	James	10	30	50	34	Amanda	10	30	50
15	Brian	6	7	35	35	Evan	10	30	50
16	Thomas	10	30	50	36	Stewart	2,75	7	10
17	Dan	2	30	39	37	Hannah	10	30	59
18	Joseph	3	30	41	38	Victor	8	30	76
19	Ryan	2	30	15	39	Nicola	18	30	75

20	Mary	6	7	35	40	Olivia	6	7	35
----	------	---	---	----	----	--------	---	---	----

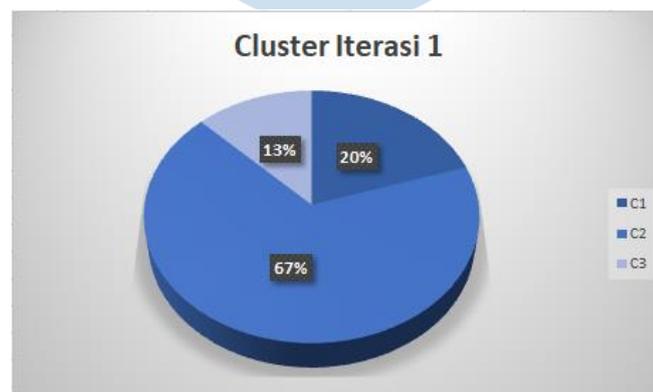
Dari 40 data tersebut kemudian ditentukan 3 *cluster* yang digunakan sebagai acuan dalam iterasi awal pada metode k-means yang digunakan. *Cluster* ini mewakili besaran paket yang umum tersedia, yaitu paket banyak, sedang dan sedikit. 3 *cluster* tersebut seperti ditunjukkan pada Tabel 2. Berdasarkan proses *clustering* yang dilakukan pada iterasi 1, maka diperoleh data iterasi 1 yang ditunjukkan dengan grafik iterasi 1 seperti pada Gambar 1 dan Gambar 2.

Tabel 2. Penentuan *Cluster* untuk Iterasi 1

No	Cluster	Besar Kuota	Masa Aktif	Harga	Jenis
1	C1	2,75	7	10	Sedikit
2	C2	10	30	50	Banyak
3	C3	6	7	35	Sedang



Gambar 1. Grafik data hasil iterasi 1

Gambar 2. Hasil *Cluster* Iterasi 1

Algoritme pengelompokan K-means dapat ditingkatkan secara signifikan dengan menggunakan teknik inisialisasi yang lebih baik, dan dengan mengulang (memulai kembali) algoritme tersebut [10].

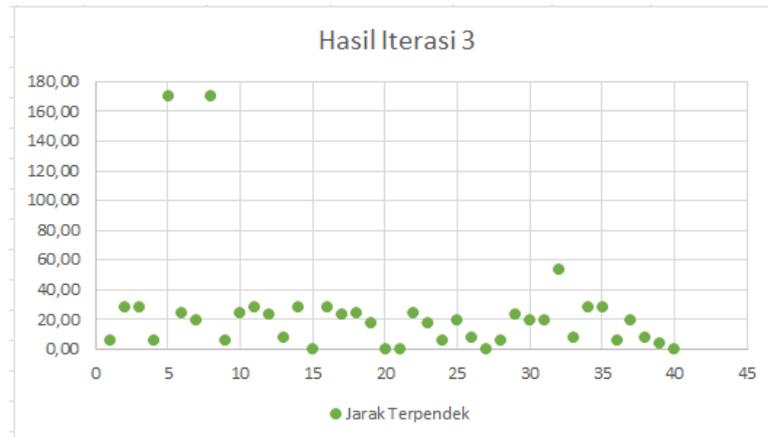
Jika data memiliki *cluster* yang tumpang tindih, k-means dapat meningkatkan hasil teknik inisialisasi. Ketika data memiliki *cluster* yang terpisah dengan baik, kinerja k-means sepenuhnya bergantung pada kebaikan inisialisasi.

Iterasi pada metode K-Means akan selalu dilakukan, sampai mendapatkan hasil yang optimal.

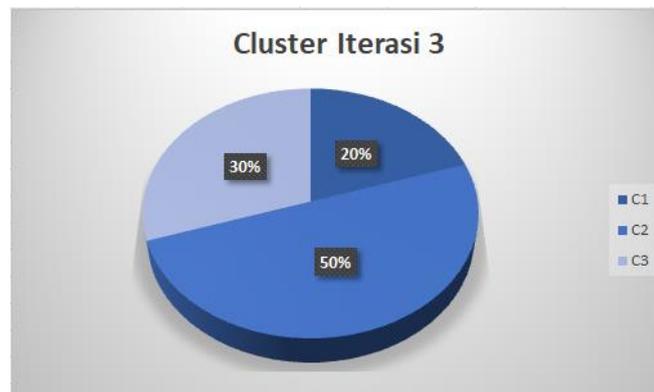
Pada proses metode K-Means pada literasi ke 2, diperoleh hasil yang belum optimal sehingga perlu dilakukan proses *clustering* perhitungan iterasi ke 3, dari hasil iterasi ke 3 diperoleh hasil yang optimal dari data uji coba yang digunakan. Dari hasil iterasi ke 3 ini akan dijadikan rujukan dalam penentuan hasil *cluster* yang dibentuk. Hasil penghitungan keseluruhan dari iterasi ke 3 ditunjukkan pada Tabel 3. Sedangkan grafik hasil iterasi ke 3 dan hasil *cluster* yang terbentuk seperti ditunjukkan pada Gambar 3 dan Gambar 4.

Tabel 3. Hasil penghitungan iterasi ke 3

No	Nama Pembeli	Besar Kuota	Masa Aktif	Harga	C1	C2	C3	Jarak Terpendek
1	Irene	2,75	7	10	6,18	72,36	25,21	6,18
2	Liam	10	30	50	42,15	27,96	27,75	27,75
3	Austin	10	30	50	42,15	27,96	27,75	27,75
4	Tracey	2,75	7	10	6,18	72,36	25,21	6,18
5	Mary	70	30	239	237,17	170,51	215,04	170,51
6	Tracey	25	30	100	92,18	24,47	71,52	24,47
7	Nicholas	10	30	59	50,37	19,23	33,48	19,23
8	Christopher	70	30	239	237,17	170,51	215,04	170,51
9	Deirdre	3	7	14	6,02	68,60	21,21	6,02
10	Austin	25	30	100	92,18	24,47	71,52	24,47
11	Victor	10	30	50	42,15	27,96	27,75	27,75
12	Samantha	2	30	39	31,84	40,74	23,69	23,69
13	Ryan	8	30	76	66,26	7,82	47,05	7,82
14	James	10	30	50	42,15	27,96	27,75	27,75
15	Brian	6	7	35	23,71	49,18	0,00	0,00
16	Thomas	10	30	50	42,15	27,96	27,75	27,75
17	Dan	2	30	39	31,84	40,74	23,69	23,69
18	Joseph	3	30	41	33,53	38,52	23,96	23,96
19	Ryan	2	30	15	17,48	63,86	30,74	17,48
20	Mary	6	7	35	23,71	49,18	0,00	0,00
21	Carol	6	7	35	23,71	49,18	0,00	0,00
22	Anna	25	30	100	92,18	24,47	71,52	24,47
23	Blake	2	30	15	17,48	63,86	30,74	17,48
24	Christopher	2,75	7	10	6,18	72,36	25,21	6,18
25	Brian	10	30	59	50,37	19,23	33,48	19,23
26	Rose	8	30	76	66,26	7,82	47,05	7,82
27	Felicity	6	7	35	23,71	49,18	0,00	0,00
28	Justin	3	7	14	6,02	68,60	21,21	6,02
29	Richard	2	30	39	31,84	40,74	23,69	23,69
30	Luke	10	30	59	50,37	19,23	33,48	19,23
31	Theresa	10	30	59	50,37	19,23	33,48	19,23
32	Ava	30	30	129	121,15	53,57	99,70	53,57
33	Harry	8	30	76	66,26	7,82	47,05	7,82
34	Amanda	10	30	50	42,15	27,96	27,75	27,75
35	Evan	10	30	50	42,15	27,96	27,75	27,75
36	Stewart	2,75	7	10	6,18	72,36	25,21	6,18
37	Hannah	10	30	59	50,37	19,23	33,48	19,23
38	Victor	8	30	76	66,26	7,82	47,05	7,82
39	Nicola	18	30	75	66,87	3,30	47,68	3,30
40	Olivia	6	7	35	23,71	49,18	0,00	0,00



Gambar 3. Grafik data hasil iterasi 3

Gambar 4. Hasil *Cluster* Iterasi 3

Gambar 4 menunjukkan hasil *cluster* paling optimal yang terbentuk. Dari hasil perhitungan iterasi 3 yang merupakan hasil final dari rangkaian proses K-Means yang sudah dijalankan sebelumnya, dari 40 data yang digunakan diperoleh suatu hasil akhir yang menunjukkan bahwa hasil yang terendah yaitu C1 (*Cluster* 1) dengan kelompok paket data dengan kuota sedikit, lama masa aktif sedang, dan harga paket data terjangkau, hasil sedang yaitu C3 (*Cluster* 3) dengan kelompok paket data dengan kuota sedang, lama masa aktif sedang, dan harga paket data terjangkau, dan hasil yang tertinggi yaitu C2 (*Cluster* 2) dengan kelompok paket data dengan kuota banyak, lama masa aktif sedang, dan harga paket data normal. Berdasarkan hasil uji coba yang diperoleh tersebut menunjukkan bahwa mayoritas penjualan berupa jumlah kuota banyak, masa aktif sedang dan harga normal.

#### IV. SIMPULAN

Pengelompokan paket penjualan paket data dengan 40 data yang telah diuji dapat dianalisis dengan menggunakan metode K-Means *Clustering*. Berdasarkan uji coba yang telah dilakukan diperoleh hasil optimal pada iterasi ke 3 dengan hasil akhir yang menunjukkan bahwa hasil yang terendah yaitu C1 (*Cluster* 1), hasil sedang yaitu C3 (*Cluster* 3), dan hasil yang tertinggi yaitu C2 (*Cluster* 2). Dari hasil tersebut

menunjukkan jika penjualan terbanyak berupa kuota banyak menjadi pilihan favorit.

#### DAFTAR PUSTAKA

- [1] F. Liantoni and L. Cahyani, "Pemanfaatan Hierarchical *Clustering* Untuk Pengelompokan Daun Berdasarkan Fitur Moment Invariant," *J. Ilm. Educat. Univ. Trunojoyo Madura*, vol. 3, no. 2, pp. 91–98, 2017.
- [2] F. Liantoni, N. Ramadijanti, and N. Rosyid Mubtada'i, "Klasifikasi Daun Dengan Centroid Linked *Clustering* Berdasarkan Fitur Bentuk Tepi Daun," *EEPIS Final Proj.*, 2010.
- [3] B. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*. West Sussex, U.K: Wiley: Chichester, 2011.
- [4] A. Al-Wakeel and J. Wu, "K-means Based *Cluster* Analysis of Residential Smart Meter Measurements," *Energy Procedia*, vol. 88, pp. 754–760, Jun. 2016, doi: 10.1016/j.egypro.2016.06.066.
- [5] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Jan. 2008, doi: 10.1007/s10115-007-0114-2.
- [6] Y. Li and H. Wu, "A *Clustering* Method Based on K-Means Algorithm," *Phys. Procedia*, vol. 25, pp. 1104–1109, 2012, doi: 10.1016/j.phpro.2012.03.206.
- [7] Agusta, "K-means Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3, 2007.
- [8] L. Rokach and O. Maimon, "Clustering Methods," in *Data Mining and Knowledge Discovery Handbook*, Springer, Boston, MA, 2005, pp. 321–352.
- [9] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, "Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm," *Comput.*

[10] *Electron. Agric.*, vol. 165, p. 104962, Oct. 2019, doi: 10.1016/j.compag.2019.104962.  
P. Fränti and S. Sieranoja, "How much can k-means be

improved by using better initialization and repeats?," *Pattern Recognit.*, vol. 93, pp. 95–112, Sep. 2019, doi: 10.1016/j.patcog.2019.04.014.

