

Analisis Clustering Paket Data Internet di Indonesia Menggunakan Metode K-Means

I Putu Ananta Wijaya^{a1}, I Gede Santi Astawa^{a2}

^aProgram Studi Informatika, Universitas Udayana
Badung, Indonesia

¹anantaw81@gmail.com

²santi.astawa@unud.ac.id

Abstract

Disruption of technology and the impact of the recent pandemic has made people access the internet longer than before, especially with smartphones. Before accessing the internet, users must purchase a data package provided by an internet service provider. Various data packages are provided by internet service providers. Starting from network coverage, speed, number of quotas, active period, to tariffs are the choices of each operator. These various offerings often make users confused because they have to adjust to economic conditions. The existence of knowledge analysis in the database, grouping data packets can be done using the k-means method. K-means groups data by iterating and creating groups based on the closest distance of the data to the center point. K-means is very widely used because of its simplicity. Before clustering, the data will go through a preprocessing process. The end result is four clusters that have their own characteristics. For example, cluster four has the characteristics of a small quota with a long active period, which is suitable for the typical community who only wants to stay connected to the internet for communication.

Keywords: K-means, Internet Quota, Clustering, Internet Provider Indonesia, Data Mining

1. Pendahuluan

Internet menghubungkan banyak orang di seluruh dunia. Di Indonesia, 19,5% masyarakatnya menghabiskan waktu lebih dari 8 jam sehari untuk mengakses internet. Hal ini dibuktikan dengan hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) periode 2019-kuartal II/2020. Ditambah lagi, pandemi Covid-19 yang mendisrupsi cara kerja masyarakat menjadi online, membuat internet seakan-akan menjadi kebutuhan primer. We Are Social menyebutkan terjadi peningkatan sebesar 2,1 juta pengguna internet pada Januari 2022 dibandingkan tahun sebelumnya. Per Januari 2022, terdapat 205 juta pengguna internet di Indonesia, yaitu 73,7% dari populasi Indonesia yang telah menggunakan internet. We Are Social juga menyebutkan 94,1% pengguna internet Indonesia mengakses internet menggunakan handphone[2].

Salah satu cara mengakses internet adalah melalui penyedia layanan internet. Banyak penyedia layanan internet di Indonesia. Sebelum mengakses internet, pengguna harus membeli paket data yang disediakan oleh penyedia layanan internet. Akibat kebutuhan internet meningkat pesat, kebutuhan paket data atau kuota internet pun semakin besar. Beragam paket data pun disediakan oleh penyedia layanan internet. Mulai dari cakupan jaringan, kecepatan, jumlah kuota, masa aktif, hingga tarif menjadi pilihan masing-masing operator (penyedia layanan internet). Dengan beragamnya skema tarif internet yang ditawarkan dan penyedia layanan internet yang ada, terkadang pelanggan justru bingung memilih paket data yang ada.

Salah satu langkah analisis dalam penemuan pengetahuan di dalam basis data adalah data mining. Pengetahuan yang dimaksud adalah pola data atau relasi antar data yang tidak diketahui sebelumnya. Dapat dikatakan, data mining adalah teknik analisis data untuk mendapatkan informasi yang tersembunyi dari data yang kompleks dalam jumlah besar. Salah satu tugas data mining berdasarkan fungsionalitasnya adalah *clustering* (klasterisasi)[3]. Clustering merupakan metode analisis yang bersifat tanpa arahan (*unsupervised*) yang bertujuan mengelompokkan data sesuai dengan ukuran kemiripannya. Salah satu algoritma pengelompokkan data adalah k-

means. K-means digunakan karena memiliki kelebihan yaitu implementasi algoritma yang mudah, waktu komputasi yang relatif cepat dan telah digunakan secara luas untuk menyelesaikan berbagai persoalan komputasi[4]. Implementasi algoritma k-means dimulai dengan memilih k objek data secara acak sebagai titik pusat (*centroid*) awal, memasukkan setiap objek yang bukan *centroid* ke klaster terdekat berdasarkan ukuran jarak, memperbaharui setiap *centroid* berdasarkan rata-rata dari objek yang ada di setiap klaster, iterasi ulang hingga *centroid* stabil [5].

Penelitian tentang *clustering* sebelumnya pernah dilakukan oleh Y. Darmi, dkk(2022) menggunakan klasterisasi dalam mengelompokkan penjualan produk menjadi dua klaster, yaitu produk yang laku dan tidak laku. Selain itu, disimpulkan semakin banyak data di input, maka *centroid* yang dapat terbentuk semakin banyak[6]. Z. Mustakim, dkk(2021) dalam mengelompokkan kualitas manajemen pendidikan sekolah di Indonesia, dimana mengelompokkan provinsi dengan hasilnya terdapat tiga klaster dengan karakteristik berupa kategori pendidikan yang baik, sedang, dan kurang[7]. Penelitian kali ini mengelompokkan paket data internet menggunakan metode k-means. Hasil dari penelitian ini berupa beberapa klaster dengan karakteristiknya masing-masing. Karakteristik ini bisa digunakan untuk menentukan paket data yang sesuai dengan kebutuhan pembeli.

2. Research Methods

2.1 Pengumpulan Data

Penelitian ini bekerja dengan pendekatan kuantitatif dan menggunakan data sekunder yang didapat melalui internet, dengan mengakses website resmi penyedia layanan internet. Data yang didapatkan adalah tarif paket data penyedia layanan internet di Indonesia khususnya untuk pengguna mobile yaitu Tri, Indosat Ooredoo, XL, Axis, Telkomsel, dan Smartfren. Karakteristik data dapat dilihat pada Tabel 1.

Tabel 1. Karakteristik Data Paket Data Internet

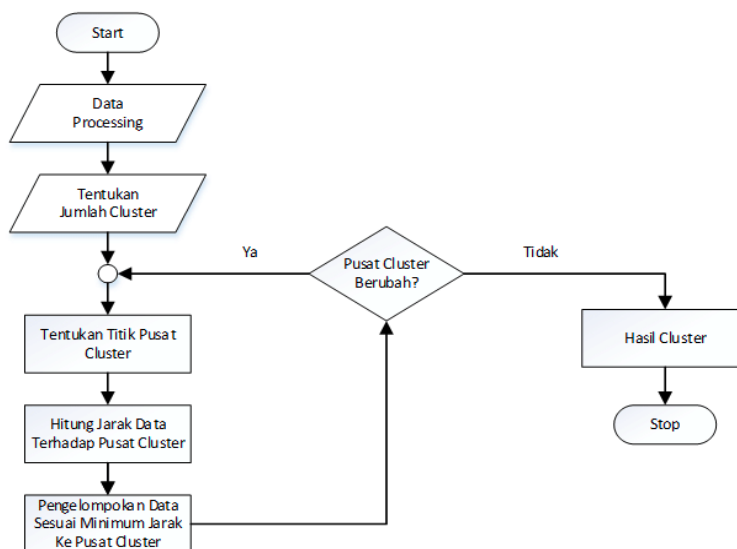
Provider	Nama Paket	Besar Kuota (GB)	Masa Aktif (Hari)	Pembagian Waktu Kuota (Ya/Tidak)	Harga	4G only (Ya/Tidak)	Pembagian Kuota ke App Khusus (Tiktok, Facebook, Webex, dll)
3	Happy 2GB 1 Hari	2	1	Tidak	4000	Tidak	Tidak
3	Happy 7GB 1 Hari	7	1	Tidak	10000	Tidak	Tidak
3	Happy 7GB 3 Hari	7	3	Tidak	15000	Tidak	Tidak
3	Happy 11GB 30 Hari	11	30	Tidak	30000	Tidak	Tidak
XL	Xtra On 17GB PB	17	mengikuti masa kartu	Tidak	88946	Tidak	Tidak
XL	Xtra On 20GB PB	20	mengikuti masa kartu	Tidak	93862	Tidak	Tidak
Telkomsel	Telkomsel GB/30 Hari 4	4	30	Tidak	50300	Tidak	Tidak
Telkomsel	Telkomsel GB/30 Hari 5	5	30	Tidak	56550	Tidak	Tidak
Axis	Bronet 3 GB 60 Hari	3	60	Tidak	32485	Tidak	Tidak
...

Axis	Bronet 5 GB 60 Hari	5	60	Tidak	48465	Tidak	Tidak
Indosat	Data Pure 100 MB 30 Hari	0.1	30	Tidak	1200	Tidak	Tidak
Smartfren	Unlimited Nonstop 6 GB 30 Hari	6	30	Tidak	29500	Tidak	Ya

Total data yang didapatkan sebanyak 233 baris dengan kolom data berupa nama provider, nama paket, besar kuota, masa aktif, pembagian waktu kuota, harga kuota, cakupan paket (4G saja/tidak), dan pembagian kuota ke aplikasi lain. Data yang diberikan adalah harga paket data provider internet Indonesia per September 2022.

2.2 Flowchart Penelitian

Flowchart algoritma k-means dapat dilihat pada gambar 1.



Gambar 1. Flowchart k-means [7]

Algoritma k-means adalah algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K cluster yang sudah ditetapkan di awal. K-means dapat diterapkan pada data yang direpresentasikan dalam r-dimensi ruang[8]. K-means digunakan ketika terdapat data yang belum berlabel (berkategori). Data akan di klusterisasi berdasarkan kesamaan fiturnya. Perhitungan k-means dapat dilakukan dengan langkah-langkah berikut:

1. Data Processing

Pada tahap ini dilaksanakan pembersihan data berupa membersihkan data yang duplikat, data yang tidak relevan dan error, kolom yang tidak diperlukan, data yang tidak konsisten, dan *outliers*. Selain itu, pada tahap ini dilakukan penanganan terhadap data yang hilang, integrasi data, transformasi data, dan reduksi data.

2. Tentukan Jumlah Cluster

Algoritma k-means bergantung pada pencarian jumlah cluster dan label data untuk nilai K yang telah ditentukan sebelumnya. Untuk mendapatkan jumlah cluster dalam data, jalankan k-means untuk nilai K dan membandingkan hasilnya. Setelah itu, dipilih nilai K yang optimal (memberikan kinerja terbaik). Ada berbagai teknik yang tersedia dalam menemukan nilai K, salah satu yang paling umum adalah metode *elbow*.

- Metode *Elbow*

Metode ini memplot nilai fungsi biaya yang dihasilkan oleh nilai K yang berbeda. Secara sederhana, metode ini melakukan perulangan beberapa kali dengan meningkatkan jumlah cluster (misalnya dari 1 hingga 10) dan melakukan plot skor pengelompokan sebagai fungsi dari jumlah cluster. Skor pengelompokan

adalah jumlah kuadrat jarak sampel ke pusat kluster terdekat. *Elbow* adalah titik pada plot yang nilai pengelompokannya jatuh ke bawah dan nilai cluster pada titik plot memberikan nilai optimum.

3. Tentukan Titik Pusat Cluster

Titik pusat cluster dibangkitkan secara acak dari beberapa objek yang tersedia sebanyak jumlah dari k cluster. Untuk menghitung titik pusat cluster ke-i, rumus yang digunakan sebagai berikut.

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad i = 1, 2, 3, \dots n \tag{1}$$

Keterangan

v = titik pusat cluster

x_i = objek ke-i

n = banyaknya jumlah objek yang menjadi anggota cluster

4. Hitung Jarak Data Terhadap Pusat Cluster

Perhitungan jarak data terhadap pusat cluster dilakukan dengan rumus *euclidean distance* :

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; \quad i = 1, 2, 3, \dots n \tag{2}$$

Keterangan

x_i = objek ke-i

y_i = data y ke-i

n = banyaknya objek

5. Pengelompokkan Data sesuai Minimum Jarak Ke Pusat Cluster

Setelah mendapatkan jarak minimum ke pusat cluster, kelompokkan data ke pusat cluster terdekat.

6. Lakukan Iterasi

Lakukan iterasi ke-i, ulangi langkah c - e. Perulangan akan berhenti jika rasio tidak lebih besar dari nilai rasio sebelumnya hingga hasil perhitungan di masing-masing data konvergen.

3. Hasil dan Pembahasan

3.1 Preprocessing Data

Pada tahap ini, dilakukan penghapusan kolom yang tidak digunakan saat klusterisasi. Kolom yang dihapus adalah nama provider, pembagian waktu kuota, cakupan paket (4G saja/tidak), dan pembagian kuota ke aplikasi lain. Sehingga kolom yang tersisa adalah nama paket, harga paket, masa aktif, harga paket. Kolom (fitur) ini sudah cukup menjadi indikator utama pertimbangan pelanggan akan membeli paket data tersebut. Selain itu, terdapat dua baris data yang dihapus karena isinya tidak lengkap. Terdapat nilai yang tidak relevan pada data kolom masa aktif karena terdapat paket yang mengikuti masa aktif, akibatnya ditulis "mengikuti masa aktif". Dilakukan perubahan terhadap data tersebut menjadi 60 hari karena nilai tertinggi data lain ada pada 60 hari serta masa aktif kartu berada pada kisaran 60 hari jika tidak diisi ulang.

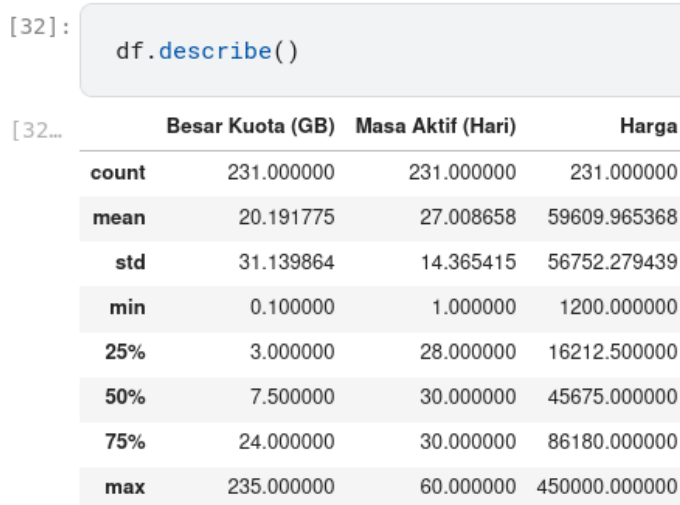
49	Tri 30GB 24 Jam	30	30	65,000.00
50	Tri 65GB 24 Jam	65	30	100,000.00
51	Tri Rp 25.000			
52	Rp 10.000			
53	Tri 110GB 24 Jam	110	30	130,000.00
54	Tri (NEW) 40GB 30 Hari	40	30	70,000.00
55	Xtra Combo Mini 1.5GB	1.5	7	9,600.00
56	Xtra Combo Mini 2.5GB	2.5	7	13,650.00

Gambar 2. Dataset Tidak Lengkap

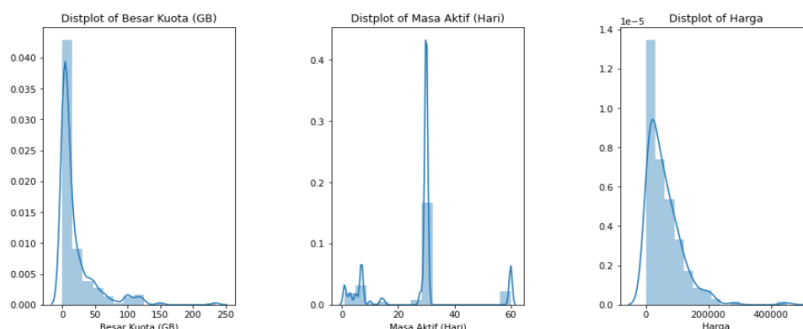
31	AON Full 40GB	40	mengikuti masa ak	99,000.00
32	AON Unlimited 26GB	26	30	73,000.00
33	AON Full 12GB	12	mengikuti masa ak	55,000.00
34	AON Full 9GB	9	mengikuti masa ak	45,000.00
35	AON Full 6GB	6	mengikuti masa ak	31,000.00
36	AON Full 3.5GB	3.5	mengikuti masa ak	21,000.00
37	AON Full 2.5GB	2.5	mengikuti masa ak	15,500.00
38	Home 150GB	150	30	150,000.00

Gambar 3. Dataset pada Kolom Masa Aktif Tidak Relevan

Selain itu dilakukan pengecekan terhadap outlier, dimana bisa dilihat dengan cara mencari mean, median, dan modus, serta melihat hasil visualisasinya.

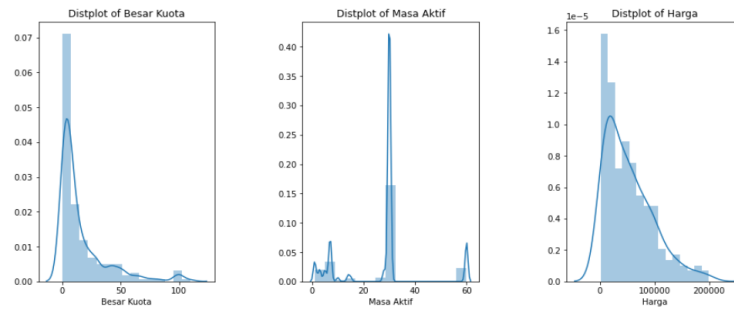


Gambar 4. Deskripsi Dataset



Gambar 5. Visualisasi Data

Dari kedua gambar di atas, kolom besar kuota dan harga memiliki outlier. Pada gambar 5 ditunjukkan oleh nilai *means/75%* sangat jauh dengan *max*. Pada gambar 6 ditunjukkan oleh adanya persebaran data yang tidak merata. Oleh karena itu, dilakukan penghapusan data yang memiliki outliers. Untuk mendeteksi outliers menggunakan pendekatan standar deviasi, yaitu hapus nilai-nilai yang merupakan sejumlah standar deviasi dari rata-rata, jika data memiliki distribusi Gaussian. Praktik industri yang umum adalah menggunakan 3 standar deviasi dari mean untuk membedakan outlier dari non-outlier. Dengan menggunakan 3 standar deviasi dapat menghapus data ekstrim sebanyak 0,3%. Hasilnya, jumlah baris data berubah menjadi 221 baris dengan nilai max yang berubah juga.



Gambar 6. Visualisasi Dataset Setelah Penghapusan Outliers

	Besar Kuota	Masa Aktif	Harga
count	221.000000	221.000000	221.000000
mean	16.164253	26.882353	52875.823529
std	21.925275	14.675109	43723.693124
min	0.100000	1.000000	1200.000000
25%	3.000000	28.000000	15500.000000
50%	7.000000	30.000000	44505.000000
75%	20.000000	30.000000	78410.000000
max	110.000000	60.000000	199360.000000

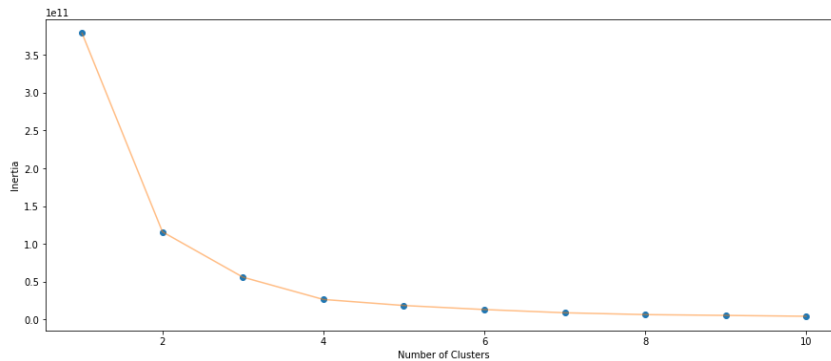
Gambar 7. Deskripsi Dataset Setelah Penghapusan Outliers

Tabel 2. Hasil *Preprocessing* Data

Nama Paket Data	Besar Kuota (GB)	Masa Aktif (Hari)	Harga (Rp.)
XL Data Pure 4GB	4	30	34050
Telkomsel 12 GB/30 Hari	12	30	103105
Bronet 8 GB 30 Hari	8	30	43850
Happy 2GB 1 Hari	2	1	4000
Freedom Internet 6,5 GB 30 Hari	6.5	30	36125
Unlimited FUP 500 MB/Hari 28 Hari	14	28	50845
...

3.2 Tentukan jumlah cluster

Penentuan jumlah cluster dilakukan menggunakan metode *elbow*. Konfigurasi yang digunakan adalah maksimal iterasi per kluster adalah 300. Hasilnya bisa dilihat pada gambar 8. Dari gambar, terlihat nilai yang bisa digunakan (perbedaan sudut garis) adalah 4 cluster.



Gambar 8. Grafik Metode Elbow

3.3 Hasil Cluster dengan K-means

Dengan menggunakan algoritma k-means didapatkan hasil empat cluster. Kluster pertama dengan karakteristik harga murah hingga sedang (0 - Rp. 32.000), kuota sedikit (0 - 11 GB), masa aktif sebentar hingga lama (1 - 60 hari). Kluster kedua dengan karakteristik harga mahal (Rp. 79.000 - Rp. 110.000), kuota murah hingga mahal (6 - 100 GB), masa aktif sedang hingga lama (7 - 60 hari). Kluster ketiga dengan karakteristik harga sangat mahal (Rp. 125.000 - Rp. 180.000), masa aktif lama (28 - 60 hari), kuota sedang hingga banyak (12 - 110 GB). Kluster keempat dengan karakteristik harga sedang (Rp. 34.000 - Rp. 70.000), masa aktif lama (14 - 60 hari), kuota sedikit hingga sedang (1,5 - 40 GB). Total data pada kluster 1 adalah 94 data, kluster 2 sebanyak 42 data, kluster 3 sebanyak 16 data, kluster 4 sebanyak 59 data. Hasil klusterisasi ditampilkan pada gambar 9, gambar 10, gambar 11, dan gambar 12.

Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga
1 Happy 2GB 1 Hari	2	1	4000	XL Data Pure 0.8GB	0.8	30	7025	QWSEM 8 GB	4	30	28600
2 Happy 7GB 3 Hari	7	1	10000	XL Data Pure 1GB	1	30	8550	Freedom Internet 2 GB/1 Hari	2	1	7925
3 Happy 7GB 3 Hari	7	3	15000	XL Data Pure 2GB	2	30	17000	Freedom Internet 2.5 GB/5 Hari	2.5	5	11475
4 Happy 11GB 30 Hari	11	30	30000	XL Data Pure 3GB	3	30	24450	Freedom Internet 5 GB/3 Hari	5	3	14925
5 Happy 11GB 30 Hari	11	30	30000	XL HotRod 0.5GB	0.5	7	5430	Freedom Internet 2 GB/30 Hari	2	30	16500
6 Happy 7GB 30 Hari	7	30	15000	XL HotRod 15GB	1	7	8130	Freedom Internet 3 GB/30 Hari	3	30	17700
7 Happy 3GB 10 Hari	3	10	31000	XL HotRod 15GB 3Hari	1.5	3	9320	Freedom Internet 7 GB/5 Hari	7	5	20895
8 Happy 9GB 10 Hari	9	10	20000	XL HotRod 1GB	1	7	9345	Freedom Internet 4 GB/30 Hari	4	30	27300
9 Happy 5GB 7 Hari	5	7	9000	XL HotRod 15GB	1	7	9450	Freedom Internet 5.5 GB/30 Hari	5.5	30	27855
10 Happy 15GB 7 Hari	1.5	7	21000	XL HotRod 2.5GB	2.5	7	11475	Indosat Data Pure 100 MB 30 Hari	0.1	30	1200
11 Happy 6GB 5 Hari	6	5	16000	XL HotRod 4GB	4	7	14660	Indosat Data Pure 200 MB 30 Hari	0.2	30	2025
12 Happy 2.5GB 5 Hari	3.5	5	15000	XL HotRod 2GB	2	7	14780	Indosat Data Pure 250 MB 30 Hari	0.25	30	2450
13 Happy 5GB 3 Hari	5	3	11000	XL HotRod 3GB	3	7	18205	Indosat Data Pure 300 MB 30 Hari	0.3	30	2950
14 Happy 5GB 3 Hari	3	3	10000	XL HotRod 5GB	6	7	18300	Indosat Data Pure 500 MB 30 Hari	0.5	30	4500
15 Happy 2.5GB 1 Hari	2.5	1	5500	XL HotRod 0.8GB	0.8	30	24605	Indosat Data Pure 750 MB 30 Hari	0.75	30	6355
16 Happy 15GB 1 Hari	1.5	1	4000	XL Xtra Combo Flex 4GB	4	30	13975	Indosat Data Pure 1 GB 30 Hari	1	30	8525
17 AON Full 6GB	6	60	31000	XL Xtra Combo Flex 7.5G	7.5	30	25005	Indosat Data Pure 1.5 GB 30 Hari	1.5	30	12500
18 AON Full 3.5GB	3.5	60	21000	XL Xtra Combo Flex 13G	13	30	32500	Indosat Data Pure 2 GB 30 Hari	2	30	15850
19 AON Full 2.5GB	2.5	60	15500	Telkomsel 500 MB/15 Hari	0.5	15	8650	Indosat Data Pure 3 GB 30 Hari	3	30	20100
20 Trj (NEW) 2.5GB 1 Hari	2.5	1	4000	Telkomsel 1 GB/30 Hari	1	30	16425	Indosat Data Pure 5 GB 30 Hari	5	30	31025
21 Trj (NEW) 4GB 3 Hari	4	3	10000	Telkomsel 2 GB/30 Hari	2	30	32350	Freedom Combo 6 GB	4	30	30175
22 Trj (NEW) 10GB 5 Hari	10	5	20000	Bronet 1 GB 30 Hari	1	30	9830	Yellow 1 GB 1 Hari	1	1	4640
23 Xtra Combo Mini 1.5GB	1.5	7	9600	Bronet 1.5 GB 30 Hari	1.5	30	13360	Yellow 1 GB 2 Hari	1	2	5300
24 Xtra Combo Mini 2.5GB	2.5	7	13650	Bronet 2 GB 30 Hari	2	30	17355	Yellow 1 GB 7 Hari	1	7	9700
25 Xtra Combo Mini 4GB	4	7	17500	Bronet 3 GB 30 Hari	3	30	23505	Yellow 1 GB 15 Hari	1	15	11700
26 Xtra Combo Mini 6GB	6	7	22250	Bronet 5 GB 30 Hari	5	30	29873	Freedom Kuota Harian 7 GB 7 Hari	7	7	26655
27 Xtra On 1GB	1	60	13450	Bronet 2 GB 60 Hari	2	60	25925	Smartfren 1 GB 7 Hari	1	7	7500
28 Xtra On 2GB	2	60	14855	Bronet 3 GB 60 Hari	3	60	32485	Smartfren 3 GB/5 Hari	3	5	9355
29 Xtra On 2GB	2	60	22500	Mini 7 Hari 1 GB/7 Hari	1	7	9368	Smartfren 15 GB 7 Hari	1.5	7	9905
30 XL Data Pure 0.5GB	0.5	30	4500	Mini 7 Hari 1.5 GB/7 Hari	1.5	7	14475	Unlimited FUP 1 GB/Hari 7 Hari	7	7	22247
				Mini 2 GB/1 Hari	2	1	7600	Unlimited Nonstop 2 GB 10 Hari	2	10	10192
				Mini 3 GB/1 Hari	3	1	8500				
				QWSEM 4 GB	2	30	16750				

Gambar 9. Kluster 1

Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga
1 AON 50GB 99ribu	50	60	99000	XL Xtra Combo 45GB	15	30	114610	QWSEM 80 GB - Unlimited GAMP	10	30	95592
2 Happy 5GB 30 Hari	55	30	110000	XL HotRod 6GB	6	30	89555	Freedom Internet 25 GB/30 Hari	25	30	79800
3 Happy 2GB 30 Hari	42	30	85000	XL HotRod 8GB	8	30	113010	Freedom Internet 28 GB/30 Hari	28	30	83405
4 Happy 5GB 14 30 Hari	52	30	77000	XL Xtra Combo Flex 45G	45	30	74200	Freedom Internet 32 GB 30 Hari	32	30	92530
5 Xtra On 4GB	40	60	99000	XL Xtra Combo Flex 72G	72	30	104625	Freedom Internet 42 GB/30 Hari	42	30	101445
6 AON Unlimited 26GB	26	30	73000	Telkomsel 10 GB/30 Hari	10	30	89255	Freedom Internet 50/GB 30 Hari	50	30	105450
7 Home 100GB	100	7	89000	Telkomsel 12 GB/30 Hari	12	30	103105	Freedom Combo 30 GB	23	30	82140
8 Home 32GB	32	30	75000	Kartu As - 12 GB Zona B	10	30	97920	Freedom Combo 40 GB	30	30	104180
9 Trj (NEW) 65GB 30 Hari	66	30	90000	Kartu LOOP - 12 GB ZonP	10	30	87360	Freedom Kuota Harian 28 GB 28 Hari	28	28	84515
10 Trj 10GB 24 Jam	65	30	100000	Kartu LOOP - 20 GB ZonP	18	30	115140	Unlimited FUP 2 GB/Hari 28 Hari	56	28	78105
11 Xtra On 17GB PB	17	60	88946	Kartu simPATI - 8 GB ZP	6	30	78120	Unlimited FUP 3 GB/Hari 28 Hari	84	28	97869
12 Xtra On 20GB PB	20	60	93862	Kartu simPATI - 17 GB ZP	15	30	108600	Bronet 10 GB 60 Hari	10	60	78410
13 Akrab Reguler 42 GB	29	30	75000	Bronet 24 GB 30 Hari	24	30	79742	Bronet 25 GB 60 Hari	25	60	87850
14 Akrab Reguler 65 GB	45	30	110000	Bronet 25 GB 30 Hari	25	30	87650	XL Xtra Combo 30GB	10	30	83000

Gambar 10. Kluster 2

Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga
1 Happy 100GB 30 Hari	100	30	135000	XL Xtra Combo Flex 100GB	100	30	123525	Bronet 50 GB 60 Hari	50	60	135300
2 Trj 10GB 24 Jam	110	30	130000	Telkomsel 50 GB/30 Hari	50	30	190360	Bronet 65 GB 60 Hari	65	60	135700
3 Akrab Reguler 147GB	78	30	180000	Kartu As - 20 GB Zona B (1-11)	18	30	124820	Freedom Internet 100 GB	100	30	135500
4 XL Xtra Combo 60GB	20	30	161555	Kartu As - 42 GB Zona B (1-11)	40	30	185180	Freedom Combo 50 GB	40	30	150900
5 XL HotRod 12GB	12	30	164875	Kartu simPATI - 32 GB Zona B	30	30	173340	Unlimited FUP 3.5 GB/Hari	98	28	151535
6				Bronet 50 GB 30 Hari	50	30	135900				
7											
8											

Gambar 11. Klaster 3

No	Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga	Nama Paket	Besar Kuota	Masa Aktif	Harga
1	Happy 17GB 30 Hari	17	30	45000	XL Xtra Combo 15GB	5	30	55575	QWSEM 16 GB	8	30	45675
2	Happy 40GB 30 Hari	40	30	70000	XL HotRod 15GB	1.5	30	44505	QWSEM 32 GB + Unlimited	4	30	48505
3	AON 20GB 55ribu	20	60	55000	XL HotRod 3GB	3	30	53755	QWSEM 24 GB	12	30	60150
4	Happy 30GB 30 Hari	30	30	70000	XL Xtra Combo Flex 10GB	10	30	40875	QWSEM 48 GB + Unlimited	6	30	60325
5	Happy 18GB 30 Hari	18	30	60000	XL Xtra Combo Flex 25GB	25	30	49350	Freedom Internet 6,5 GB*	6.5	30	36125
6	Happy 14GB 30 Hari	14	30	55000	Telkomsel 3 GB/30 Hari	3	30	40650	Freedom Internet 9 GB/3*	9	30	38740
7	Happy 16GB 30 Hari	11	30	45000	Telkomsel 4 GB/30 Hari	4	30	50300	Freedom Internet 10 GB*	10	30	46255
8	AON Full 12GB	12	60	55000	Telkomsel 5 GB/30 Hari	5	30	56550	Freedom Internet 16 GB*	16	30	59625
9	AON Full 9GB	9	60	45000	Telkomsel 6 GB/30 Hari	6	30	65500	Freedom Internet 18 GB*	18	30	65725
10	Tri (NEW) 20GB 30 Hari	20	30	50000	Telkomsel 8 GB/30 Hari	8	30	68900	Indosat Data Pure 6 GB 3*	6	30	38800
11	Tri (NEW) 24GB 30 Hari	24	30	55000	Kartu As 5 GB Zona B (1-11)	4	30	46820	Indosat Data Pure 7 GB 3*	7	30	43400
12	Tri 30GB 24 Jam	30	30	65000	Kartu LOOP - 5 GB Zona B (1-7)	4	30	51220	Indosat Data Pure 8 GB 3*	8	30	44600
13	Tri (NEW) 40GB 30 Hari	40	30	70000	Kartu simpATI - 5 GB Zona B (*)	4	30	47320	Indosat Data Pure 9 GB 3*	9	30	49100
14	Akrab Reguler 30 GB	20	30	56000	Bronet 8 GB 30 Hari	8	30	43850	Indosat Data Pure 10 GB*	10	30	53600
15	XL Data Pure 4GB	4	30	34050	Bronet 12 GB 30 Hari	12	30	57295	Freedom Combo 10 GB	7	30	40620
16	XL Data Pure 5GB	5	30	42550	Bronet 15 GB 30 Hari	15	30	64650	Freedom Combo 20 GB	15	30	61665
17	XL Data Pure 6GB	6	30	54705	Bronet 19,5 GB 30 Hari	19.5	30	68338	Freedom Kuota Harian 1*	14	14	48995
18	XL Data Pure 8GB	8	30	66655	Bronet 5 GB 60 Hari	5	60	48465	Unlimited FUP 1 GB/Hari*	14	14	39559
19	QWSEM 24 GB + Unlimi	3	30	41825	Bronet 8 GB 60 Hari	8	60	66702	Unlimited FUP 500 MB/H*	14	28	50845
20					QWSEM 16 GB + Unlimited GA*	2	30	40420	Unlimited FUP 700 MB/H*	19.6	28	58836

Gambar 12. Klaster 4

4. Kesimpulan

Dari hasil penelitian yang telah dilakukan, dapat diambil beberapa kesimpulan.

1. Klasterisasi paket data internet berbagai operator Indonesia dapat dikelompokkan menjadi empat kategori dengan karakteristik berupa harga, besar kuota, dan masa aktif. Klaster pertama cocok bagi pelanggan yang tidak terlalu sering menggunakan internet dan kebutuhan biaya sedikit. Klaster kedua cocok bagi pelanggan yang sering menggunakan internet dan mampu membeli paket data dengan harga mahal. Klaster keempat cocok bagi pelanggan yang menggunakan internet dalam taraf standar.
2. Klasterisasi di atas dibentuk menjadi empat kategori, tetapi masih ada data yang berbeda dengan yang lainnya, contohnya pada klaster 1, terdapat satu paket data dengan masa aktif cuma 7 hari sedangkan data yang lain masa aktifnya di atas 28 hari. Disarankan menggunakan jumlah yang lebih dari empat.
3. Outliers pada data paket data sangat mungkin terjadi karena karakteristik data paket data internet sangat tidak menyebar. Hal ini dibuktikan saat penggambaran diagram visualisasi. Hal ini dapat diatasi dengan menghilangkan outliers.

Referensi

- [1] -, "APJII", 2019. [Online]. Available: <https://apji.or.id/survei2019x> [1 Oktober 2022]
- [2] -, "We Are Social", 26 January 2022. [Online]. Available: <https://wearesocial.com/uk/blog/2022/01/digital-2022-another-year-of-bumper-growth-2/>. [1 October 2022]
- [3] Suyanto, Data Mining untuk Klasifikasi dan Klasterisasi Data, First ed., Bandung: Informatika, 2019, pp. 4.
- [4] D. Triyansyah and D. Fitriah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing" IncomTech Jurnal Telekomunikasi dan Komputer, vol. 8, no. 3, p. 164, 2018.
- [5] Suyanto, Machine Learning Tingkat Dasar dan Lanjut, First ed., Bandung: Informatika, 2018, pp. 205.
- [6] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means dalam Pengelompokan Penjualan Produk" Jurnal Media Infotama, vol. 12, no. 2, p. 157, 2016.
- [7] R. Muliono and Z. Sembiring, "DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASTERISASI TINGKAT TRIDARMA PENGAJARAN DOSEN" CESS (Journal of Computer Engineering System and Science), vol. 4, no. 2, p. 274, 2019.
- [8] F. Nur, M. Zarlis, and B. Benyamin, "PENERAPAN ALGORITMA K-MEANS PADA SISWA BARU SEKOLAH MENENGAH KEJURUAN UNTUK CLUSTERING JURUSAN" Jurnal Nasional Informatika dan Teknologi Jaringan, vol. 1, no. 2, p. 101, 2017.