

Pengelompokan Data Rasio Penggunaan Gas Rumah Tangga Berdasarkan Provinsi Di Indonesia Menggunakan Metode K-Means Clustering

Pipit Mutiara Putri¹, Lise Pujiastuti², Iin Parlina¹, Solikhun^{1,*}

¹ AMIK Tunas Bangsa, Kota Pematangsiantar, Indonesia

² STMIK Antar Bangsa, Tangerang, Indonesia

Email: ¹pmp09051999@gmail.com, ²lise.pujiastuti@gmail.com, ³iinparlina@gmail.com, ^{4*}solikhun@amiktunasbangsa.ac.id

Abstrak—Rasio gas rumah tangga merupakan rasio untuk mengetahui kemampuan perusahaan dalam menghasilkan laba selama periode tertentu dan juga memberikan gambaran tentang tingkat efektifitas manajemen perusahaan dalam melaksanakan kegiatan operasinya. Gas rumah tangga saat ini banyak digunakan oleh masyarakat dibanding menggunakan minyak tanah. Banyaknya masyarakat yang beralih dari minyak tanah ke gas rumah tangga disebabkan karena pemakaian atau penggunaannya lebih efisien dan mudah untuk digunakan dan mudah untuk dicari. Dengan pertimbangan pentingnya gas tersebut, pemerintah selalu berupaya untuk meningkatkan kebutuhan industri, pertimbangan tersebut menjadi semakin penting bagi Indonesia karena jumlah penduduknya semakin besar. Tercatat dari sejak tahun 2015 hingga tahun 2018 rasio gas rumah tangga di Indonesia terus mengalami peningkatan. Untuk menstabilkan persediaan dan harga Gas maka, pemerintah harus mengetahui kebutuhan gas di Indonesia. Melihat permasalahan yang cukup kompleks tersebut, tentunya dibutuhkan suatu metode yang dapat lebih efektif dalam melakukan pengelompokan data rasio penggunaan gas rumah tangga berdasarkan provinsi di Indonesia dalam pemenuhan kebutuhan bahan masak di Indonesia. Adapun metode yang akan digunakan dalam pengelompokan data rasio penggunaan gas rumah tangga berdasarkan provinsi di Indonesia penelitian menggunakan metode K-Means Clustering

Kata Kunci: Pengelompokan Data Rasio Gas Rumah Tangga, K-Means Clustering

1. PENDAHULUAN

Seiring dengan program pemerintah dalam upaya mengurangi konsumsi minyak tanah ke gas, masyarakat dituntut beralih bahan bakar dari minyak ke gas. Dengan mengurangi subsidi untuk minyak tanah, sehingga menyebabkan harga minyak tanah semakin tinggi, serta masyarakat beralih menggunakan gas untuk memenuhi kebutuhan rumah tangga. Pemerintah dapat menghemat subsidi hingga Rp. 15–Rp. 20 triliun jika program ini berhasil [1]. Rasio gas rumah tangga merupakan rasio untuk mengetahui kemampuan perusahaan dalam menghasilkan laba selama periode tertentu dan juga memberikan gambaran tentang tingkat efektifitas manajemen perusahaan dalam melaksanakan kegiatan operasinya [2].

Dengan pertimbangan pentingnya gas tersebut, pemerintah selalu berupaya untuk meningkatkan kebutuhan industri. Tercatat dari sejak tahun 2015 hingga tahun 2018 rasio gas rumah tangga di Indonesia terus mengalami peningkatan. Tidak hanya dalam skala nasional tapi juga berdasarkan provinsi. Untuk menstabilkan persediaan dan harga Gas, maka pemerintah harus mengetahui kebutuhan gas di Indonesia. Melihat permasalahan yang cukup kompleks tersebut, tentunya dibutuhkan suatu metode yang dapat lebih efektif dalam melakukan pengelompokan data rasio pengguna gas rumah tangga berdasarkan provinsi di Indonesia dalam pemenuhan kebutuhan bahan masak di Indonesia. Adapun metode yang akan digunakan dalam pengelompokan data rasio penggunaan gas rumah tangga berdasarkan provinsi di Indonesia penelitian menggunakan metode *K-Means Clustering*.

2. METODOLOGI PENELITIAN

2.1 Data Mining

Data mining yang juga dikenal dengan istilah pattern recognition merupakan suatu metode yang digunakan untuk mengolah data sehingga menemukan pola yang tersembunyi dari data yang sudah diolah. Data yang diolah dengan teknik data mining ini kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil dari pengolahan data tersebut dapat digunakan dalam menentukan keputusan di masa depan (Sihombing 2017) [3].

2.2 K-Means Clustering

K-Means Clustering adalah salah satu bentuk pengelompokan yang paling sederhana. Prosedurnya sederhana dan mudah untuk mengklasifikasikan data yang diberikan melalui sejumlah cluster. Penentuan centroid dilakukan dengan cara mengambil data pertama sebagai centroid pertama, data kedua sebagai centroid kedua, dan seterusnya hingga jumlah centroid yang diperlukan. Langkah berikutnya adalah dengan menghitung jarak dari titik yang akan di cluster ke setiap centroid yang ada dan dikelompokkan sesuai dengan jarak terdekat kepada centroid-nya. Bila semua titik sudah masuk kedalam pengelompokan maka langkah pertama selesai [2].

2.3 Algoritma K-Means

Algoritma K-Means merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat klaster (centroid) terdekat dengan data. Tujuan dari K-Means adalah mengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap [4].

3. ANALISA DAN PEMBAHASAN

Pada Metode penelitian ini menggunakan metode *K-means Clustering*, dimana terlebih dahulu yang dihitung adalah rata-rata dari jumlah pengguna gas rumah tangga menurut provinsi di Indonesia periode 2015-2018. Berikut merupakan data jumlah pengguna gas rumah tangga berdasarkan provinsi di Indonesia periode 2015-2018.

Tabel 1. data jumlah pengguna gas rumah tangga

Provinsi	Rasio Penggunaan Gas Rumah Tangga			
	2015	2016	2017	2018
ACEH	74.41	77.94	82.96	85.35
SUMATERA UTARA	72.09	75.43	78.17	80.42
SUMATERA BARAT	41.58	54.73	61.99	66.36
RIAU	77.05	81.18	83.74	86.50
JAMBI	63.69	69.42	75.69	77.29
SUMATERA SELATAN	81.53	84.12	88.29	89.26
BENGKULU	72.28	74.51	80.10	84.80
LAMPUNG	63.83	67.90	76.70	78.04
KEP. BANGKA BELITUNG	73.55	76.48	79.57	81.42
KEP. RIAU	70.72	76.26	75.27	74.66
DKI JAKARTA	89.24	89.38	87.53	91.78
JAWA BARAT	81.45	83.89	85.63	86.17
JAWA TENGAH	70.83	73.30	79.24	80.13
DI YOGYAKARTA	62.91	65.12	71.66	73.16
JAWA TIMUR	70.51	73.77	76.58	78.21
BANTEN	81.47	83.50	84.80	86.57
BALI	67.47	69.53	72.38	74.95
NUSA TENGGARA BARAT	43.20	49.40	54.85	58.87
NUSA TENGGARA TIMUR	0.27	0.57	0.51	1.04
KALIMANTAN BARAT	74.22	79.52	85.97	87.01
KALIMANTAN TENGAH	34.23	51.07	69.27	76.71
KALIMANTAN SELATAN	52.09	64.63	74.17	75.09
KALIMANTAN TIMUR	91.55	93.71	94.90	94.11
KALIMANTAN UTARA	53.91	62.71	68.20	75.33
SULAWESI UTARA	56.68	64.22	68.59	72.61
SULAWESI TENGAH	13.09	29.45	47.88	52.61
SULAWESI SELATAN	78.85	82.55	85.13	86.89
SULAWESI TENGGARA	32.54	43.80	50.55	56.92
GORONTALO	55.96	65.95	77.21	81.22
SULAWESI BARAT	53.08	63.32	69.07	74.47
MALUKU	0.77	0.65	0.44	0.75
MALUKU UTARA	0.51	0.65	0.89	1.70
PAPUA BARAT	3.34	1.59	3.42	4.50
PAPUA	0.61	0.80	0.88	1.13

Data tersebut kemudian terlebih dahulu dicari nilai rata-ratanya, untuk mempermudah pengerjaannya, peneliti dalam melakukan pencarian nilai rata-rata di *Microsoft Excel*. Berikut merupakan data jumlah rata-rata pengguna gas rumah tangga berdasarkan provinsi di Indonesia periode 2015-2018 (Sumber: www.bps.go.id) yang telah diolah ke dalam *Microsoft Excel* untuk mempermudah cara perhitungan.

Tabel 2. Jumlah Rata-rata Penggunaan Gas Rumah Tangga berdasarkan Provinsi

Provinsi	Rata-rata
ACEH	74.41
SUMATERA UTARA	72.09
SUMATERA BARAT	48.16
RIAU	80.66
JAMBI	69.42
SUMATERA SELATAN	86.21
BENGKULU	73.40
LAMPUNG	67.90
KEP. BANGKA BELITUNG	76.48

KEP. RIAU	75.77
DKI JAKARTA	89.66
JAWA BARAT	86.17
JAWA TENGAH	79.24
DI YOGYAKARTA	73.16
JAWA TIMUR	76.58
BANTEN	84.80
BALI	74.95
NUSA TENGGARA BARAT	54.85
NUSA TENGGARA TIMUR	1.04
KALIMANTAN BARAT	87.01
KALIMANTAN TENGAH	76.71
KALIMANTAN SELATAN	75.09
KALIMANTAN TIMUR	94.11
KALIMANTAN UTARA	75.33
SULAWESI UTARA	60.45
SULAWESI TENGAH	52.61
SULAWESI SELATAN	86.89
SULAWESI TENGGARA	47.18
GORONTALO	81.22
SULAWESI BARAT	69.07
MALUKU	0.75
MALUKU UTARA	1.70
PAPUA BARAT	3.42
PAPUA	1.13

3.1 Centroid Data

Centroid adalah titik tengah suatu *cluster*. *Centroid* berupa nilai. *Centroid* digunakan untuk menghitung jarak suatu objek data terhadap *centroid*. Suatu objek data termasuk dalam *cluster* jika memiliki jarak terpendek terhadap *centroid cluster* tersebut[5]. Menentukan *centroid* awal dapat di peroleh secara acak. *Centroid* awal merupakan titik pusat *cluster* pertama. Penentuan *cluster* ada tiga bagian yaitu *cluster* tingkat Tinggi (C1), *cluster* tingkat Sedang (C2), dan *cluster* tingkat Rendah (C3). Untuk *cluster* tingkat Tinggi (Max) untuk *cluster* tingkat sedang (Rata-rata), dan untuk *cluster* tingkat rendah (Min). Berikut adalah *centroid* data awal dari penelitian ini:

Tabel 3. Centroid data C1

MAX (C1)	8031411
AVERAGE(C2)	675625
MIN(C3)	1608

Tabel 4. Centroid data C3

MAX (C1)	7912933
AVERAGE(C2)	6299556
MIN(C3)	1608

3.2 Clustering Data

Proses *cluster* atau pengelompokan data bisa dilakukan dengan dua metode, yaitu Metode *Hirarki* dan Metode *Non-Hirarki*. Metode yang dilakukan dalam penelitian ini adalah metode *Non-Hirarki*. Metode ini digunakan sebagai alternatif metode *cluster* untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode *hirarki*. [6]. Metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan (dua *cluster*, tiga *cluster*, atau lain sebagainya). Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti proses *hierarki*. Berikut adalah data *clustering* yang telah dilakukan:

Iterasi 2			Jarak Terpendek	C1	C2	C3
C1	C2	C3				
5.904111111	6.8475	72.802	5.904111111	1		
8.224111111	4.5275	70.482	4.5275		1	
32.15911111	19.4075	46.547	19.4075	1		
0.342555556	13.09416667	79.04866667	0.342555556	1		
10.89411111	1.8575	67.812	1.8575		1	
5.890888889	18.6425	84.597	5.890888889	1		
6.919111111	5.8325	71.787	5.8325		1	
12.14111111	0.3375	66.292	0.3375		1	
3.834111111	8.9175	74.872	3.834111111	1		
4.549111111	8.2025	74.157	4.549111111	1		
9.340888889	22.0925	88.047	9.340888889	1		
5.855888889	18.6075	84.562	5.855888889	1		
1.074111111	11.6775	77.632	1.074111111	1		
7.154111111	5.5975	71.552	5.5975	1		
3.734111111	9.0175	74.972	3.734111111	1		
4.485888889	17.2375	83.192	4.485888889	1		
5.364111111	7.3875	73.342	5.364111111	1		
25.46411111	12.7125	53.242	12.7125		1	
79.27411111	66.5225	0.568	0.568			1
6.695888889	19.4475	85.402	6.695888889	1		
3.604111111	9.1475	75.102	3.604111111	1		
5.224111111	7.5275	73.482	5.224111111	1		
13.795888889	26.5475	92.502	13.795888889	1		
4.984111111	7.7675	73.722	4.984111111	1		
19.86411111	7.1125	58.942	7.1125		1	
27.70411111	14.9525	51.002	14.9525		1	
6.575888889	19.3275	85.282	6.575888889	1		
33.13911111	20.3875	45.567	20.3875		1	
0.905888889	13.6575	79.612	0.905888889	1		
11.24411111	1.5075	67.462	1.5075		1	
79.56411111	66.8125	0.858	0.858			1
78.61411111	65.8625	0.092	0.092			1
76.89411111	64.1425	1.812	1.812			1
79.18411111	66.4325	0.478	0.478			1
Jumlah				20	9	5
				79.12933	62.99556	1.608

Iterasi 1			Jarak Terpendek	C1	C2	C3
C1	C2	C3				
19.70	10.9515686	73.66	10.95	1		
22.02	8.63156863	71.34	8.631568627	1		
45.96	15.3034314	47.405	15.30	1		
13.45333333	17.1982353	79.9066667	13.45333333	1		
24.69	5.96156863	68.67	5.96		1	
7.905	22.7465686	85.455	7.905	1		
20.72	9.93656863	72.645	9.94		1	
26.21	4.44156863	67.15	4.441568627	1		
17.63	13.0215686	75.73	13.02	1		
18.345	12.3065686	75.015	12.30656863		1	
4.46	26.1965686	88.905	4.46	1		
7.94	22.7115686	85.42	7.94	1		
14.87	15.7815686	78.49	14.87	1		
20.95	9.70156863	72.41	9.701568627		1	
17.53	13.1215686	75.83	13.12	1		
9.31	21.3415686	84.05	9.31	1		
19.16	11.4915686	74.2	11.49		1	
39.26	8.60843137	54.1	8.608431373		1	
93.07	62.4184314	0.29	0.29			1
7.1	23.5515686	86.26	7.1	1		
17.40	13.2515686	75.96	13.25		1	
19.02	11.6315686	74.34	11.63156863		1	
0.00	30.6515686	93.36	0.00	1		
18.78	11.8715686	74.58	11.87156863		1	
33.66	3.00843137	59.7	3.01		1	
41.5	10.8484314	51.86	10.84843137		1	
7.22	23.4315686	86.14	7.22	1		
46.935	16.2834314	46.425	16.28343137		1	
12.89	17.7615686	80.47	12.89	1		
25.04	5.61156863	68.32	5.611568627		1	
93.36	62.7084314	0	0.00			1
92.41	61.7584314	0.95	0.95			1
90.69	60.0384314	2.67	2.67			1
92.98	62.3284314	0.38	0.38			1
Jumlah				15	14	5
				80.31411	67.5625	1.608

Gambar 1. Hasil Perhitungan dan pengelompokan data dari Centroid Awal (Iterasi 1), Iterasi 2, dan Iterasi 3

Iterasi 3			Jarak Terpendek	C1	C2	C3
C1	C2	C3				
4.719333	11.41444	72.802	4.719333333	1		
7.039333	9.094444	70.482	7.039333333	1		
30.97433	14.84056	46.547	14.84055556		1	
1.527333	17.66111	79.04867	1.527333333	1		
9.709333	6.424444	67.812	6.424444444		1	
7.075667	23.20944	84.597	7.075666667	1		
5.734333	10.39944	71.787	5.734333333	1		
11.22933	4.904444	66.292	4.904444444		1	
2.649333	13.48444	74.872	2.649333333	1		
3.364333	12.76944	74.157	3.364333333	1		
10.52567	26.65944	88.047	10.52566667	1		
7.040667	23.17444	84.562	7.040666667	1		
0.110667	16.24444	77.632	0.110666667	1		
5.969333	10.16444	71.552	5.969333333	1		
2.549333	13.58444	74.972	2.549333333	1		
5.670667	21.80444	83.192	5.670666667	1		
4.179333	11.95444	73.342	4.179333333	1		
24.27933	8.145556	53.242	8.145555556		1	
78.08933	61.95556	0.568	0.568			1
7.880667	24.01444	85.402	7.880666667	1		
2.419333	13.71444	75.102	2.419333333	1		
4.039333	12.09444	73.482	4.039333333	1		
14.98067	31.11444	92.502	14.98066667	1		
3.799333	12.33444	73.722	3.799333333	1		
18.67933	2.545556	58.842	2.545555556		1	
26.51933	10.38556	51.002	10.38555556		1	
7.760667	23.89444	85.282	7.760666667	1		
31.95433	15.82056	45.567	15.82055556		1	
2.090667	18.22444	79.612	2.090666667	1		
10.05933	6.074444	67.462	6.074444444		1	
78.37933	62.24556	0.858	0.858			1
77.42933	61.29556	0.092	0.092			1
75.70933	59.57556	1.812	1.812			1
77.99933	61.86556	0.478	0.478			1
Jumlah				21	8	5
				79.99603	58.70375	1.608

Gambar 2. Hasil Perhitungan dari Centroid Awal (Iterasi 1), Iterasi 2, dan Iterasi 3

Proses *K-Means* akan terus beriterasi terus sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya atau sama dengan proses akan terus melakukan iterasi sampai data pada iterasi terakhir sama dengan iterasi sebelumnya.

4. KESIMPULAN

Dari hasil penelitian, dapat disimpulkan bahwa jumlah data pengguna gas rumah tangga menurut provinsi 2015-2018 menggunakan metode *Clustering* dengan algoritma *K-Means* dapat digunakan untuk mengelompokkan data jumlah pengguna. Data tersebut diolah menggunakan Ms. Excel untuk menentukan nilai *centroid* dalam 3 *cluster* yaitu *cluster* tingkat tinggi, *cluster* tingkat sedang, *cluster* tingkat rendah. *Centroid* data untuk tingkat tinggi sebanyak 45.96 yaitu untuk provinsi Sumatera Barat, untuk tingkat sedang sebanyak 16.2834314 yaitu untuk provinsi Sulawesi Tenggara, dan provinsi lainnya untuk tingkat rendah sebanyak 62.7084314

REFERENCES

- [1] A. Fajar and D. Y. Restivia, "Pengaruh Pemberitaan Surat Kabar Kompas , Seputar Indonesia dan Media Indonesia Terhadap Persepsi Masyarakat Pengguna Tabung Gas," vol. 1, 2011.
- [2] N. H. Kristanto, A. C. L. A, and H. B. S, "Implementasi K-Means Clustering untuk Pengelompokan Analisis Rasio Profitabilitas dalam Working Capital," vol. 02, no. 01, 2016.
- [3] K. Riset *et al.*, "PROPOSAL PENELITIAN 2019," no. 8, 2020.
- [4] I. Method, K. C. Based, S. Value, W. Interface, C. Study, and I. U. M. M. Magelang, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," vol. 18, no. 1, pp. 76–82, 2015.
- [5] A. P. WINDARTO, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering, Program Studi Sistem Informasi, Data Mining," vol. 16, no. 4, pp. 348–357, 2017.
- [6] E. RIANTI, "Data Mining Dalam Menentukan Kacamata Pada Optik Zal Laris Dan Tidak Laris Menggunakan Metode Clustering, Universitas Indonesia," vol. 4, no. 2, pp. 267–283, 2017.