

# Implementasi K-Means Clustering Dalam Mengelompokkan Rumah Tangga Kumuh(Perkotaan) Menurut Wilayah

Mhd Yuda Rizki<sup>1</sup>, Agus Perdana Windarto<sup>2\*</sup>

<sup>1,2</sup> Prodi Sistem Informasi STIKOM TUNAS BANGSA Pematangsiantar

<sup>1</sup>ryuda1014@gmail.com

<sup>2\*</sup>agus.perdana@amiktunasbangsa.ac.id

**Abstrak** — Kebersihan adalah salah satu hal yang harus paling diutamakan dalam segala hal termasuk pada keadaan rumah dalam sebuah rumah tangga. Penelitian ini membahas tentang “Implementasi K-Means Clustering Dalam Mengelompokkan Rumah Tangga Kumuh(Perkotaan) Menurut Wilayah”. Peneliti memperoleh data bersumber dari sebuah situs website pemerintah yaitu BPS (Badan Pusat Statistik) [www.bps.go.id.Data](http://www.bps.go.id>Data) tersebut mewakili -34 provinsi di Indonesia dan data diambil dari tahun 2015-2018. Terdapat -2 buah cluster dalam penelitian ini yakni cluster tingkat tinggi (C1) dan rendah (C2). Proses clustering berhenti pada iterasi ke -6 dan memperoleh hasil 16 Provinsi menduduki posisi cluster tingkat tinggi dan 18 Provinsi lainnya menduduki posisi cluster tingkat rendah. Diharapkan dengan adanya penelitian ini dapat menjadi masukan kepada pemerintah kota di setiap wilayah agar memberi perhatian lebih kepada wilayah yang tingkat rumah tangga kumuh masih berada di posisi cluster tingkat tinggi (C1) agar meningkatkan kualitas wilayah tersebut.

**Kata Kunci** — Data Mining, K-Means, Clustering, Provinsi, Rumah Tangga, Kumuh.

**Abstract** — Cleanliness is one of the things that must be prioritized in all matters including the state of the house in a household. This research discusses “Implementation of K-Means Clustering in Grouping Slum (Urban) Households by Region”. Researchers obtained data sourced from a government website namely BPS (Statistics Indonesia) [www.bps.go.id.The](http://www.bps.go.id.The) data represents -34 provinces in Indonesia and data is taken from 2015-2018. There were -2 cluster in this study, namely high (C1) and low (C2) cluster. The clustering process stopped at the 6<sup>th</sup> iteration and obtained 16 provinces occupying high-level cluster position and 18 other provinces occupied the low-level cluster positions. With this research can be input to the city government in each region to give more attention to areas where the level of slum households are still in the positions of high-level cluster (C1) in order to improve the quality of the region.

**Keywords** — Data Mining, K-Means, Clustering, Province, Household, Seedy.

## I. PENDAHULUAN

Wilayah yang bersih merupakan keinginan dari setiap wilayah di Indonesia, karena dengan ditandai dengan kebersihannya suatu wilayah maka virus penyakit enggan menyerang masyarakat yang ada pada wilayah tersebut. Jika suatu wilayah dikategorikan bersih maka dapat dilihat masyarakatnya akan jarang terjangkit virus penyakit karena virus penyakit berasal dari tempat-tempat yang kumuh atau kotor. Dalam penelitian ini peneliti membahas tentang “IMPLEMENTASI K-MEANS CLUSTERING DALAM MENGELOMPOKKAN RUMAH TANGGA KUMUH(PERKOTAAN) MENURUT WILAYAH”. Terdapat -2 buah cluster di dalam penelitian ini yakni cluster tingkat tinggi (C1) dan rendah (C2). Proses clustering berhenti pada iterasi ke -6 dan memperoleh hasil 16 Provinsi menduduki posisi cluster tingkat tinggi dan 18 Provinsi lainnya menduduki posisi cluster tingkat rendah. Diharapkan dengan adanya penelitian ini dapat menjadi masukan kepada pemerintah kota di setiap wilayah agar memberi perhatian lebih kepada wilayah yang tingkat rumah tangga kumuh masih berada di posisi cluster tingkat tinggi (C1) agar meningkatkan kualitas wilayah tersebut.

## II. METODOLOGI PENELITIAN

### A. Data Mining

Data Mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis [1].

### B. K-Means

Merupakan algoritma clustering yang berulang-ulang. Algoritma K-Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil) [2].

## III. HASIL DAN PEMBAHASAN

Menurut [3] Dalam penelitian ini metode yang digunakan yaitu metode Data Mining sebagai berikut :

1. Tahap pengumpulan data,
2. Tahap pengolahan data,
3. Tahap Clustering dan
4. Tahap Analisis Format Penulisan

TABEL I  
DATA RUMAH TANGGA KUMUH MENURUT WILAYAH

Provinsi	Persentase Rumah Tangga Kumuh Perkotaan (40% Ke Bawah), Menurut Provinsi (Persen)			
ACEH	9,33	8,45	8,1	9,12
SUMATERA UTARA	9,82	9,36	8,25	8,74
SUMATERA BARAT	12,58	13,02	9,97	9,36
RIAU	8,89	5,63	5,8	3,36
JAMBI	6,48	8,73	5,99	5,21
SUMATERA SELATAN	15,01	16,13	11,22	12,68
BENGKULU	11,41	12,44	14,85	9,13
LAMPUNG	6,88	4,82	4,44	4,33
KEP. BANGKA BELITUNG	5,85	5,42	3,41	5,13
KEP. RIAU	4,23	3,83	3,43	5,7
DKI JAKARTA	13,25	11,02	14,55	14,36
JAWA BARAT	14,09	11,81	11,86	11,16
JAWA TENGAH	4,49	3,7	3,53	3,21
DI YOGYAKARTA	4,18	3,32	3,95	2,77
JAWA TIMUR	4,78	4,57	3,4	4,06
BANTEN	8,73	7,16	5,7	4,7
BALI	4,73	2,97	2,74	2,75
NUSA TENGGARA BARAT	12,5	11,35	9,97	10,9
NUSA TENGGARA TIMUR	28,13	27,09	23,21	16,15
KALIMANTAN BARAT	7,89	6,07	7,93	4,01
KALIMANTAN TENGAH	12,47	9,57	9,78	8,88
KALIMANTAN SELATAN	11,34	9,29	7,81	10,07
KALIMANTAN TIMUR	7,4	6,69	7,37	5,93
KALIMANTAN UTARA	16,15	12,61	13,12	9,52
SULAWESI UTARA	19,15	11,27	8,41	12,92
SULAWESI TENGAH	14,52	14,69	9,37	14,63
SULAWESI SELATAN	7,49	5,56	5,82	7,74
SULAWESI TENGGARA	11,26	13,14	13,49	11,16
GORONTALO	11,54	19,55	15,13	12,83
SULAWESI BARAT	12,44	11,05	22,66	15,22
MALUKU	19,22	16,72	19,09	14,72
MALUKU UTARA	5,22	7,56	4,25	6,93
PAPUA BARAT	19,31	25,7	20,65	14,55
PAPUA	20,85	14,44	12,48	17

Setelah data diperoleh selanjutnya adalah menentukan centroid data awal dari data yang telah diperoleh. Adapun centroid data awal yang diperoleh dapat dilihat pada table II berikut:

TABEL II  
CENTROID DATA AWAL

	a	b	c	d
c1	28,13	27,09	23,21	17
c2	4,18	2,97	2,74	2,75

Keterangan :

A = 2015  
B = 2016  
C = 2016  
D = 2018

Selanjutnya adalah melakukan proses pencarian clushter tingkat tinggi (C1) dan clushter tingkat rendah(C2) dapat dilihat pada table III.

TABEL III  
PROSES ITERASI -I

Provinsi	Persentase Rumah Tangga Kumuh Perkotaan (40% Ke Bawah), Menurut Provinsi (Persen)				Jp(jarak terpenddek)		
	c1	c2	c1	c2	c1	c2	Jp(jarak terpenddek)
ACEH	9,33	8,45	8,1	9,12	31,48486	11,21871	11,21871
SUMATERA UTARA	9,82	9,36	8,25	8,74	30,68612	11,78482	11,78482
SUMATERA BARAT	12,58	13,02	9,97	9,36	25,95062	16,35627	16,35627
RIAU	8,89	5,63	5,8	3,36	36,32997	6,24463	6,24463
JAMBI	6,48	8,73	5,99	5,21	35,23272	7,421705	7,421705
SUMATERA SELATAN	15,01	16,13	11,22	12,68	21,32319	21,47067	21,32319
BENKULU	11,41	12,44	14,85	9,13	25,02014	18,14691	18,14691
LAMPUNG	6,88	4,82	4,44	4,33	38,21462	4,012343	4,012343
KEP. BANGKA BELITUNG	5,85	5,42	3,41	5,13	38,71594	3,860661	3,860661
KEP. RIAU	4,23	3,83	3,43	5,7	40,38782	3,149714	3,149714
DKI JAKARTA	13,25	11,02	14,55	14,36	23,69862	20,52646	20,52646
JAWA BARAT	14,09	11,81	11,86	11,16	24,36243	18,17295	18,17295
JAWA TENGAH	4,49	3,7	3,53	3,21	41,02936	1,210248	1,210248
DI YOGYAKARTA	4,18	3,32	3,95	2,77	41,377	1,259762	1,259762
JAWA TIMUR	4,78	4,57	3,4	4,06	40,15287	2,252044	2,252044
BANTEN	8,73	7,16	5,7	4,7	35,09209	7,129004	7,129004
BALI	4,73	2,97	2,74	2,75	41,84994	0,55	0,55
NUSA TENGGARA BARAT	12,5	11,35	9,97	10,9	26,5434	16,0668	16,0668
NUSA TENGGARA TIMUR	28,13	27,09	23,21	16,15	0,85	41,88028	0,85
KALIMANTAN BARAT	7,89	6,07	7,93	4,01	35,40786	7,204013	7,204013
KALIMANTAN TENGAH	12,47	9,57	9,78	8,88	28,25748	14,12171	14,12171
KALIMANTAN SELATAN	11,34	9,29	7,81	10,07	29,73094	13,05738	13,05738
KALIMANTAN TIMUR	7,4	6,69	7,37	5,93	34,9191	7,467001	7,467001
KALIMANTAN UTARA	16,15	12,61	13,12	9,52	22,60419	19,74304	19,74304
SULAWESI UTARA	19,15	11,27	8,41	12,92	23,80334	20,7019	20,7019
SULAWESI TENGAH	14,52	14,69	9,37	14,63	23,15501	20,72113	20,72113
SULAWESI SELATAN	7,49	5,56	5,82	7,74	35,74507	7,214617	7,214617
SULAWESI TENGGARA	11,26	13,14	13,49	11,16	24,65326	18,43491	18,43491
GORONTALO	11,54	19,55	15,13	12,83	20,36553	24,16991	20,36553
SULAWESI BARAT	12,44	11,05	22,66	15,22	22,51507	26,18819	22,51507
MALUKU	19,22	16,72	19,09	14,72	14,46021	28,73791	14,46021
MALUKU UTARA	5,22	7,56	4,25	6,93	36,97534	6,473191	6,473191
PAPUA BARAT	19,31	25,7	20,65	14,55	9,606279	34,72143	9,606279
PAPUA	20,85	14,44	12,48	17	18,11502	26,59661	18,11502

Setelah diperoleh nilai C1 dan C2 serta menentukan jarak terpendaknya. Selanjutnya hasil akan di kelompokkan dalam sebuah table. Hasil iterasi -1 dapat dilihat pada table IV berikut ini :

TABEL IV  
HASIL PROSES ITERASI -1

Provinsi		
	c1	c2
ACEH		1
SUMATERA UTARA		1
SUMATERA BARAT		1
RIAU		1
JAMBI		1
SUMATERA SELATAN	1	
BENGKULU		1
LAMPUNG		1
KEP. BANGKA BELITUNG		1
KEP. RIAU		1
DKI JAKARTA		1
JAWA BARAT		1
JAWA TENGAH		1
DI YOGYAKARTA		1
JAWA TIMUR		1
BANTEN		1
BALI		1

NUSA TENGGARA BARAT		1
NUSA TENGGARA TIMUR	1	
KALIMANTAN BARAT		1
KALIMANTAN TENGAH		1
KALIMANTAN SELATAN		1
KALIMANTAN TIMUR		1
KALIMANTAN UTARA		1
SULAWESI UTARA		1
SULAWESI TENGAH		1
SULAWESI SELATAN		1
SULAWESI TENGGARA		1
GORONTALO	1	
SULAWESI BARAT	1	
MALUKU	1	
MALUKU UTARA		1
PAPUA BARAT	1	
PAPUA	1	

Proses akan terus berlanjut sampai hasil iterasi sama dan cocok dengan hasil iterasi sebelumnya. Pada penelitian ini proses berhenti pada iterasi ke -6. Proses iterasi -6 dapat dilihat pada table V, VI dan VII berikut ini :

TABEL V  
CENTROID DATA ITERASI -6

	a	b	C	d
c1	15,71313	15,12688	14,37688	12,89313
c2	7,233333	6,261111	5,65	5,702222

TABEL VI  
PROSES ITERASI -6

Provinsi	Persentase Rumah Tangga Kumuh Perkotaan (40% Ke Bawah), Menurut Provinsi (Persen)				Jp(jarak terpendek)		
	c1	c2	c3	c4	c1	c2	c3
ACEH	9,33	8,45	8,1	9,12	11,78815	5,18372	5,18372
SUMATERA UTARA	9,82	9,36	8,25	8,74	11,08029	5,68173	5,68173
SUMATERA BARAT	12,58	13,02	9,97	9,36	6,794035	10,31073	6,794035
RIAU	8,89	5,63	5,8	3,36	17,3548	2,941318	2,941318
JAMBI	6,48	8,73	5,99	5,21	15,98564	2,64968	2,64968
SUMATERA SELATAN	15,01	16,13	11,22	12,68	3,392923	15,41382	3,392923
BENGKULU	11,41	12,44	14,85	9,13	6,334125	12,32935	6,334125
LAMPUNG	6,88	4,82	4,44	4,33	18,87656	2,355576	2,355576
KEP. BANGKA BELITUNG	5,85	5,42	3,41	5,13	19,28842	2,822431	2,822431
KEP. RIAU	4,23	3,83	3,43	5,7	20,7619	4,456312	4,456312
DKI JAKARTA	13,25	11,02	14,55	14,36	5,011497	14,59501	5,011497

JAWA BARAT	14,09	11,81	11,86	11,16	4,79318	12,08948	4,79318
JAWA TENGAH	4,49	3,7	3,53	3,21	21,63214	4,97903	4,97903
DI YOGYAKARTA	4,18	3,32	3,95	2,77	21,99119	5,42779	5,42779
JAWA TIMUR	4,78	4,57	3,4	4,06	20,7243	4,078982	4,078982
BANTEN	8,73	7,16	5,7	4,7	15,95778	2,013694	2,013694
BALI	4,73	2,97	2,74	2,75	22,51041	5,855067	5,855067
NUSA TENGGARA BARAT	12,5	11,35	9,97	10,9	6,926908	9,965634	6,926908
NUSA TENGGARA TIMUR	28,13	27,09	23,21	16,15	19,64501	35,88904	19,64501
KALIMANTAN BARAT	7,89	6,07	7,93	4,01	16,23885	2,920574	2,920574
KALIMANTAN TENGAH	12,47	9,57	9,78	8,88	8,867534	8,09485	8,09485
KALIMANTAN SELATAN	11,34	9,29	7,81	10,07	10,21211	7,055633	7,055633
KALIMANTAN TIMUR	7,4	6,69	7,37	5,93	15,42304	1,794995	1,794995
KALIMANTAN UTARA	16,15	12,61	13,12	9,52	4,413981	13,791	4,413981
SULAWESI UTARA	19,15	11,27	8,41	12,92	7,892523	15,06021	7,892523
SULAWESI TENGAH	14,52	14,69	9,37	14,63	5,449765	14,75416	5,449765
SULAWESI SELATAN	7,49	5,56	5,82	7,74	16,09098	2,176895	2,176895
SULAWESI TENGGARA	11,26	13,14	13,49	11,16	5,250549	12,44131	5,250549
GORONTALO	11,54	19,55	15,13	12,83	6,127821	18,32533	6,127821
SULAWESI BARAT	12,44	11,05	22,66	15,22	10,06771	20,73574	10,06771
MALUKU	19,22	16,72	19,09	14,72	6,355096	22,69411	6,355096
MALUKU UTARA	5,22	7,56	4,25	6,93	17,47786	3,034479	3,034479
PAPUA BARAT	19,31	25,7	20,65	14,55	12,91611	28,7576	12,91611
PAPUA	20,85	14,44	12,48	17	6,879232	20,65421	6,879232

TABEL VII  
HASIL PROSES ITERASI -6

Provinsi				
	c1	c2		
ACEH		1	NUSA TENGGARA TIMUR	1
SUMATERA UTARA		1	KALIMANTAN BARAT	1
SUMATERA BARAT	1		KALIMANTAN TENGAH	1
RIAU		1	KALIMANTAN SELATAN	1
JAMBI		1	KALIMANTAN TIMUR	1
SUMATERA SELATAN	1		KALIMANTAN UTARA	1
BENGGKULU	1		SULAWESI UTARA	1
LAMPUNG		1	SULAWESI TENGAH	1
KEP. BANGKA BELITUNG		1	SULAWESI SELATAN	1
KEP. RIAU		1	SULAWESI TENGGARA	1
DKI JAKARTA	1		GORONTALO	1
JAWA BARAT	1		SULAWESI BARAT	1
JAWA TENGAH		1	MALUKU	1
DI YOGYAKARTA		1	MALUKU UTARA	1
JAWA TIMUR		1	PAPUA BARAT	1
BANTEN		1	PAPUA	1
BALI		1		
NUSA TENGGARA BARAT	1			

Selanjutnya peneliti memanfaatkan software RapidMiner untuk menguji keakuratan pada penelitian yang telah dilakukan dan hasilnya dapat dilihat pada gambar 1 dan 2 berikut :

Attribute	cluster_0	cluster_1
2015.0	15.713	7.233
2016.0	15.127	6.261
2017.0	14.377	5.650
2018.0	12.893	5.702

Gambar 1. Hasil centroid iterasi -6 menggunakan RapidMiner

Cluster Model
Cluster 0: 16 items
Cluster 1: 18 items
Total number of items: 34

Gambar 2. Hasil clushter menggunakan RapidMiner

#### IV. KESIMPULAN

Algoritma K-Means clushtering dapat di terapkan dalam menentukan populasi rumah tangga kumuh menurut wilayah.

Dan Dengan memanfaatkan software RapidMiner dapat menjadi pilihan terbaik dalam menguji algoritma K-Means clushtering. Penelitian ini diharapkan menjadi masukan kepada pemerintah kota disetiap wilayah agar memberi perhatian lebih kepada wilayah yang berada di posisi clushter tingkat tinggi(C1).

#### REFERENSI

- [1] M. Gading Sadewo, A. Perdana Windarto, and S. Retno Andani, "KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) PEMANFAATAN ALGORITMA CLUSHTERING DALAM MENGELOMPOKKAN JUMLAH DESA/KELURAHAN YANG MEMILIKI SARANA KESEHATAN MENURUT PROVINSI DENGAN K-MEANS."
- [2] T. Rismawan and D. S. Kusumadewi, "Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (Bmi) & Ukuran Kerangka," *Semin. Nas. Apl. Teknol. Inf.*, vol. 21, no. 01, pp. 1907-5022, 2008.
- [3] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216-223, 2018, doi: 10.30865/komik.v2i1.929.