

# PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN PERSENTASE RUMAH TANGGA MENURUT KUALITAS FISIK AIR MINUM DENGAN MENGGUNAKAN K-MEANS CLUSTER

**Artanti Indrasetianingsih**

Dosen Program Studi Statistika, FMIPA Universitas PGRI Adi Buana (UNIPA) Surabaya  
Email : [artanti.indra@gmail.com](mailto:artanti.indra@gmail.com)

## ABSTRAK

Kebutuhan akan air bersih dan layak, baik untuk dikonsumsi maupun untuk keperluan lain seperti mandi, masak, dan lain-lain di Indonesia terus meningkat seiring dengan meningkatnya jumlah penduduk. Sedangkan ketersediaan air bersih cenderung berkurang karena adanya kerusakan alam dan pencemaran lingkungan. Peraturan Menteri Kesehatan (Permenkes) No. 492/Menkes/Per/IV/2010 menerangkan bahwa parameter fisik air bersih yang aman bagi kesehatan meliputi bau, warna, total zat padat terlarut (TDS), kekeruhan, rasa dan suhu. Kesadaran masyarakat akan pentingnya menjaga lingkungan dan perilaku masyarakat dalam meningkatkan kualitas lingkungannya agar menjadi lebih baik, khususnya dalam ketersediaan air bersih, masih perlu ditingkatkan.. Tujuan dari penelitian ini adalah mengelompokkan provinsi di Indonesia berdasarkan persentase rumah tangga menurut kualitas fisik air minum dengan menggunakan analisis cluster non hirarki, yaitu *K-Means Cluster*. Data yang digunakan dalam penelitian ini bersumber dari Riset Kesehatan Dasar tahun 2013, yaitu data persentase rumah tangga berdasarkan kualitas fisik air minum menurut provinsi di Indonesia. Kualitas fisik air minum yang dimaksud adalah tidak keruh, tidak berwarna, tidak berasa, tidak berbusa, tidak berbau dan baik. Langkah-langkah analisis yang dilakukan adalah analisis deskriptif dan analisis *K-Means Cluster*. Hasil analisis *K-Means Cluster* diperoleh 2 cluster, yaitu cluster 1 (Persentase rumah tangga yang kualitas fisik air minumnya kurang dari rata-rata persentase kualitas fisik air minum secara nasional) terdiri dari 3 provinsi (Aceh, Nusa Tenggara Timur dan Papua) dan cluster 2 (Persentase rumah tangga yang lebih dari rata-rata persentase kualitas fisik air minum secara nasional) terdiri dari 33 provinsi di luar cluster 1.

**Kata Kunci** : *K-Means Cluster*, Kualitas fisik air minum

## 1. Pendahuluan

Air merupakan salah satu kebutuhan yang sangat penting bagi manusia. Kebutuhan akan air bersih dan layak, baik untuk dikonsumsi maupun untuk keperluan lain seperti mandi, masak, dan lain-lain di Indonesia terus meningkat seiring dengan meningkatnya jumlah penduduk. Sedangkan ketersediaan air bersih cenderung berkurang karena adanya kerusakan alam dan pencemaran lingkungan. Perbandingan dengan tahun 2007 menunjukkan akses air bersih pada tahun 2010 telah mengalami penurunan kira-kira sebesar tujuh persen (Unicef Indonesia, 2012).

Air minum yang aman bagi kesehatan berdasarkan Peraturan Menteri Kesehatan (Permenkes) No. 492/Menkes/Per/IV/2010 pasal 3 ayat 1 adalah apabila memenuhi persyaratan fisika, mikrobiologis, kimiawi dan radioaktif yang dimuat dalam parameter wajib dan parameter tambahan. Parameter wajib merupakan persyaratan kualitas air minum yang wajib diikuti dan ditaati oleh seluruh penyelenggara air minum. Berdasarkan Permenkes tersebut, parameter fisik merupakan salah satu jenis parameter wajib. Parameter fisik tersebut meliputi bau, warna, total zat padat terlarut (TDS), kekeruhan, rasa dan suhu. Kadar maksimum yang diperbolehkan dari masing-masing parameter fisik, untuk bau adalah tidak berbau, untuk warna adalah 15 TCU, untuk TDS adalah 500 mg/l, untuk kekeruhan adalah 5 NTU, untuk rasa adalah tidak berasa, dan untuk suhu adalah suhu udara  $\pm 3$  °C.

Kesadaran masyarakat akan pentingnya menjaga lingkungan dan perilaku masyarakat dalam meningkatkan kualitas lingkungannya agar menjadi lebih baik perlu ditingkatkan agar upaya-upaya yang telah dan masih dilaksanakan oleh pihak-pihak yang berwenang, khususnya dalam ketersediaan air bersih akan lebih bermanfaat. Hal ini sebaiknya dilakukan dari lingkungan masyarakat yang terkecil yaitu keluarga.

Pada penelitian ini digunakan data persentase rumah tangga berdasarkan kualitas fisik air minum menurut provinsi di Indonesia tahun 2013. Kualitas fisik air minum yang dimaksud adalah tidak keruh, tidak berwarna, tidak berasa, tidak berbusa, tidak berbau dan baik. Tujuan dari penelitian ini adalah mengelompokkan provinsi di Indonesia berdasarkan persentase rumah tangga menurut kualitas fisik air minum dengan menggunakan analisis cluster non hirarki, yaitu *K-Means Cluster*. Pengelompokan yang dilakukan ini diharapkan dapat memberikan informasi yang bermanfaat untuk peningkatan kualitas fisik air minum di masing-masing provinsi sesuai dengan kelompoknya.

Beberapa penelitian sebelumnya yang menggunakan metode K-Means Cluster adalah Rismawan dan Kusumadewi (2008) mengelompokkan nilai Body Mass Index (BMI) dan ukuran rangka mahasiswa dengan menggunakan K-Means Cluster. Hasil yang pengelompokan yang diperoleh adalah 3, yaitu BMI normal dan kerangka besar, BMI obesitas sedang dan kerangka sedang, dan BMI obesitas berat dan kerangka kecil. Febriyana, (2011) melakukan analisis klaster k-means dan k-median pada data indikator kemiskinan (studi kasus data indikator Kemiskinan Kabupaten di Indonesia tahun 2009). Hasil yang diperoleh adalah metode pengklasteran K-Means lebih baik daripada K-Median untuk kasus tersebut, karena ketepatan klasifikasi K-Means sebesar 98,51%, sedangkan K-Median sebesar 97,57%. Alfina,dkk. (2012) meneliti tentang perbandingan metode *Hierarchical Clustering*, K-Means dan Gabungan Keduanya dalam Cluster Data. Hasil yang diperoleh gabungan metode *single linkage* dan K-means memberikan hasil cluster lebih baik dengan parameter uji cluster variance dan metode *silhouette*.

## 2. Tinjauan Pustaka

### 2.1 Analisis Cluster

Analisis cluster merupakan salah satu teknik Multivariate yang masuk dalam klasifikasi metode interdependen, yaitu metode yang mempelajari struktur

hubungan antara variabel, antara kasus/responden dan antara obyek. Analisis cluster bertujuan untuk mengelompokkan obyek atau kasus ke dalam kelompok yang mempunyai sifat yang relatif homogen. Obyek-obyek yang mempunyai kemiripan dikelompokkan dalam cluster yang sama, sedangkan obyek-obyek lain yang tidak mirip dikelompokkan dalam cluster yang lain.

Metode pengelompokan dalam analisis cluster ada dua, yaitu metode hirarki dan metode non hirarki. Metode hirarki merupakan pengelompokan yang dilakukan secara hirarki atau berjenjang dari  $n$ ,  $(n-1)$  sampai 1 kelompok. Metode non hirarki dilakukan dengan menentukan terlebih dahulu jumlah cluster yang diinginkan. Salah satu prosedur non hirarki yang terkenal adalah metode K-Means.

Ada dua asumsi yang harus dipenuhi dalam analisis cluster, yaitu sampel yang sudah representatif (mewakili populasi) dan tidak ada kasus multikolinieritas antar variabel (Hair, dkk., 2006). Sampel yang representatif dapat dilihat dari nilai Kaiser-Meyer-Olkin (KMO) yang lebih besar dari 0.5 (Santoso, 2012). Ada tidaknya multikolinieritas antar variabel dapat diketahui dari nilai *Variance Inflation Factor* (VIF) yang lebih besar dari 10 (Gujarati, 2004).

## 2.2 K-Means Cluster

Mc Queen dalam Johnson dan Wichern (1998) menyarankan bahwa algoritma pada metode K-Means Cluster menetapkan suatu obyek ke dalam cluster yang mempunyai centroid (mean) terdekat. Proses ini dibagi dalam tiga tahap :

1. Mempartisi obyek-obyek ke dalam K cluster awal.
2. Menentukan obyek yang mempunyai jarak terdekat dengan centroid (mean). (Ukuran jarak yang biasa digunakan adalah jarak Euclidean, yang digunakan baik untuk pengamatan yang distandarkan maupun yang tidak distandarkan). Menghitung ulang centroid yang digunakan cluster untuk mendapatkan obyek yang baru dan untuk cluster yang kehilangan obyek.
3. Mengulangi langkah ke-2 sampai tidak ada lagi obyek yang dipindahkan.

Jarak Euclidean antara  $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$  dan  $\mathbf{y} = [y_1, y_2, \dots, y_p]^T$  adalah sebagai berikut (Johnson dan Wichern, 1998) :

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

## 3. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang bersumber dari Riskesdas 2013, yaitu data persentase rumah tangga menurut kualitas fisik air minum dan provinsi di Indonesia (Kementerian Kesehatan Republik Indonesia, 2014).

Variabel-variabel yang digunakan pada penelitian ini meliputi variabel provinsi dan variabel kualitas fisik air minum, yaitu :

1. Variabel Provinsi, terdapat 33 provinsi di Indonesia
2. Variabel Tidak Keruh, merupakan persentase rumah tangga di masing-masing provinsi yang kualitas fisik air minumnya tidak keruh.
3. Variabel Tidak Berwarna, merupakan persentase rumah tangga di masing-masing provinsi yang kualitas fisik air minumnya tidak berwarna.

4. Variabel Tidak Berasa, merupakan persentase rumah tangga di masing-masing provinsi yang kualitas fisik air minumnya tidak berasa.
5. Variabel Tidak Berbusa, merupakan persentase rumah tangga di masing-masing provinsi yang kualitas fisik air minumnya tidak berbusa.
6. Variabel Tidak Berbau, merupakan persentase rumah tangga di masing-masing provinsi yang kualitas fisik air minumnya tidak berbau.
7. Variabel Baik, merupakan persentase rumah tangga di masing-masing provinsi yang kualitas fisik air minumnya baik (tidak keruh, tidak berwarna, tidak berasa, tidak berbusa dan tidak berbau).

Langkah-langkah analisis yang dilakukan pada penelitian ini dilakukan dalam dua tahap, yaitu sebagai berikut :

1. Analisis deskriptif, yaitu dengan membuat tabel dan grafik dari variabel-variabel penelitain.
2. Analisis untuk memperoleh kelompok dari persentase rumah tangga menurut kualitas fisik air minum berdasarkan provinsi dengan menggunakan metode K-Means Cluster.

#### 4. Hasil Penelitian Dan Pembahasan

##### 4.1 Statistik Deskriptif

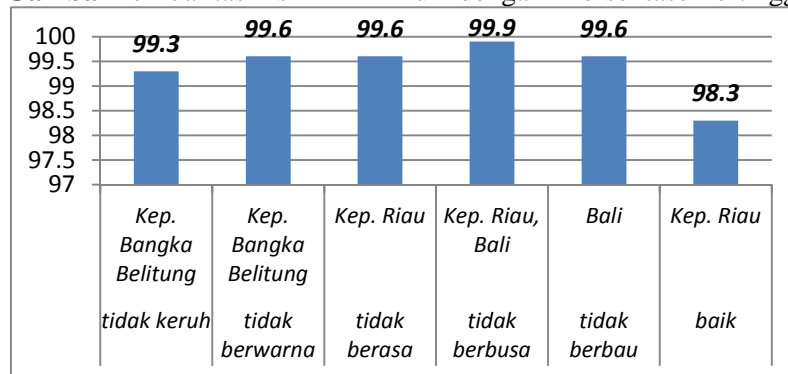
Tabel 1 menerangkan variabel-variabel persentase kualitas fisik air minum secara deskriptif. Berdasarkan Gambar 1 dapat diketahui bahwa persentase rumah tangga dengan kualitas fisik air minum tidak berbusa, mempunyai persentase tertinggi, yaitu sebesar 99,90%, adalah rumah tangga yang berasal dari propinsi Kepulauan Riau dan Bali. Sedangkan rumah tangga dengan persentase terendah sebesar 78,60% untuk kualitas fisik air minum baik, yaitu rumah tangga yang berasal dari propinsi papua (dapat dilihat pada Gambar 2).

**Tabel 1.** Deskriptif Variabel Proporsi Kualitas Fisik Air Minum

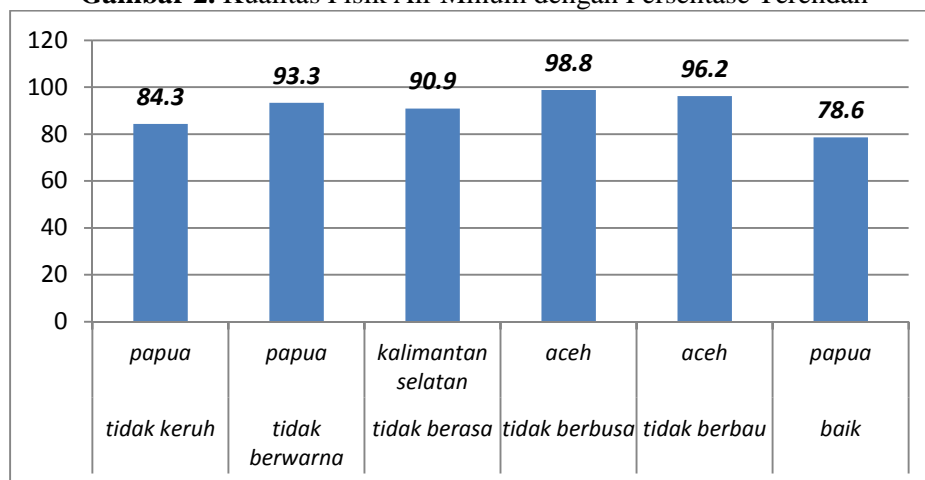
Variabel	Minimum	Maksimum	Mean	Standar Deviasi
Tidak Keruh	84,30	99,30	95,95	8,07
Tidak Berwarna	93,40	99,60	97,94	2,08
Tidak Berasa	90,90	99,60	96,99	4,17
Tidak Berbusa	98,80	99,90	99,43	0,06
Tidak Berbau	96,20	99,60	98,52	0,58
Baik	78,60	98,30	93,10	15,5

5) *Pengelompokan Provinsi di Indonesia Berdasarkan Persentase Rumah Tangga Menurut Kualitas Fisik Air Minum Dengan Menggunakan K-Means Cluster*

**Gambar 1.** Kualitas Fisik Air Minum dengan Persentase Tertinggi



**Gambar 2.** Kualitas Fisik Air Minum dengan Persentase Terendah



#### 4.2 K-Means Cluster

Sebelum dilakukan analisis K-Means cluster, maka perlu dilakukan pengujian asumsi dalam analisis cluster, yaitu sampel yang diambil sudah representatif mewakili populasi dan tidak terjadi multikolinieritas. Nilai KMO yang diperoleh adalah 0,634, karena nilai KMO sudah lebih besar dari 0.5, maka asumsi pertama sudah terpenuhi. Multikolinieritas dapat diketahui dari nilai VIF, karena nilai VIF untuk variabel tidak keruh, tidak berwarna, tidak berasa dan baik mempunyai nilai lebih dari 10, maka hal ini menunjukkan adanya kasus multikolinier. Jika dilihat dari nilai korelasi antara variabel-variabel penelitian, maka nilai korelasi yang tertinggi adalah antara variabel baik dengan variabel tidak keruh, yaitu sebesar 0,91. Variabel baik dan variabel tidak berwarna mempunyai nilai korelasi yang tinggi dengan variabel-variabel yang lain, yaitu nilainya lebih besar dari 0,5. Untuk mengatasi kasus multikolinieritas, maka variabel baik dan variabel tidak berwarna dikeluarkan dari analisis cluster, karena mempunyai nilai korelasi yang tinggi dengan variabel-variabel yang lain. Setelah variabel baik dan variabel tidak berwarna dikeluarkan, nilai KMO yang diperoleh sebesar 0.683, sehingga asumsi sampel yang representative terpenuhi. Sedangkan nilai VIF untuk variabel tidak keruh sebesar 2,1, nilai VIF untuk variabel tidak berasa sebesar 1,8, nilai VIF untuk variabel tidak berasa sebesar 2,3, dan nilai VIF

variabel tidak berbau sebesar 2,1. Karena semua nilai VIF  $< 10$ , maka tidak ada kasus multikolinieritas. Dengan demikian kedua asumsi dalam analisis cluster sudah terpenuhi.

Pada analisis K-Means Cluster, jumlah cluster perlu ditentukan terlebih dahulu. Jumlah cluster yang ditentukan pada penelitian ini adalah 2. Standarisasi tidak dilakukan sebelum melakukan analisis cluster, karena semua variabel mempunyai satuan yang sama, yaitu dalam persen.

Setelah dilakukan K-Means Cluster diperoleh hasil pengelompokan 33 provinsi di Indonesia berdasarkan persentase rumah tangga menurut kualitas fisik air minum (tidak keruh, tidak berasa, tidak berbusa dan tidak berbau) dapat dilihat pada Tabel 2. Tabel 2 menunjukkan bahwa anggota cluster 1 terdiri dari 3 provinsi, yaitu Aceh, Nusa Tenggara Timur dan Papua, sedangkan cluster 2 terdiri dari 30 provinsi, yaitu semua provinsi di luar provinsi Aceh, Nusa Tenggara Timur dan Papua.

Tabel 3 menerangkan tentang *final cluster centers*. Berdasarkan Tabel 3, maka dapat diketahui rata-rata persentase rumah tangga menurut kualitas fisik air minum di setiap cluster. Jika dilihat dari rata-rata tersebut, maka :

- Cluster 1 merupakan persentase rumah tangga dengan kualitas fisik air minum yang kurang dari rata-rata persentase kualitas fisik air minum secara nasional.
- Cluster 2 merupakan persentase rumah tangga dengan kualitas fisik air minum yang lebih besar dari rata-rata persentase kualitas fisik air minum secara nasional.

**Tabel 3.** Rata-rata Persentase Rumah Tangga menurut Kualitas Fisik Air Minum di setiap Cluster

Kualitas fisik air minum	Cluster 1	Cluster 2
Tidak Keruh	88,90	96,65
Tidak Berasa	93,60	97,33
Tidak Berbusa	99,07	99,47
Tidak Berbau	97,53	98,62

**Tabel 4.** Anggota-anggota Cluster

cluster 1	cluster 2
Aceh	Sumatera Barat
Nusa Tenggara Timur	Sumatera Utara
Papua	Jambi
	Bengkulu
	Riau
	Kep. Bangka Belitung
	Kep. Riau
	Sumatera Selatan
	Lampung
	DKI Jakarta
	Banten
	Jawa Barat

7) *Pengelompokan Provinsi di Indonesia Berdasarkan Persentase Rumah Tangga Menurut Kualitas Fisik Air Minum Dengan Menggunakan K-Means Cluster*

---

	Jawa Tengah
	DIY
	Jawa Timur
	Kalimantan Barat
	Kalimantan Selatan
	Kalimantan Tengah
	Kalimantan Timur
	Gorontalo
	Bali
	Nusa Tenggara Barat
	Sulawesi Tengah
	Sulawesi Tenggara
	Sulawesi Barat
	Sulawesi Selatan
	Sulawesi Utara
	Maluku Utara
	Maluku
	Papua Barat

## 5. Kesimpulan

Berdasarkan hasil dan pembahasan di atas, maka dapat disimpulkan:

1. Provinsi Kepulauan Riau dan Bali merupakan provinsi dengan persentase rumah tangga yang kualitas fisik air minumnya tertinggi untuk kategori tidak berbasa, yaitu sebesar 99,90%. Sedangkan provinsi Papua merupakan provinsi dengan persentase rumah tangga yang kualitas fisik air minumnya terendah untuk kategori baik, yaitu sebesar 78,6%.
2. Pengelompokan 33 provinsi di Indonesia berdasarkan persentase rumah tangga menurut kualitas fisik air minum kategori tidak keruh, tidak berasa, tidak berbasa dan tidak berbau dengan menggunakan metode K-Means Cluster diperoleh hasil cluster 1 terdiri dari 3 provinsi, yaitu Aceh, Nusa Tenggara Timur dan Papua (Persentase rumah tangga yang kualitas fisik air minumnya kurang dari rata-rata persentase kualitas fisik air minum secara nasional) dan cluster 2 terdiri dari 33 provinsi (Persentase rumah tangga yang lebih dari rata-rata persentase kualitas fisik air minum secara nasional)

## Daftar Pustaka

- Alfina,dkk., (2012), Analisis Perbandingan Metode *Hierarchical Clustering*, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS, Jurnal Teknik ITS, Vol. 1., ISSN:2301-9271, <http://ejurnal2.its.ac.id/index.php/teknik/article/viewFile/1794/600>, Tanggal akses : 2 Juli 2015.
- Alfina,dkk., (2012), Analisis Perbandingan Metode *Hierarchical Clustering*, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem

- Kerja Praktek Jurusan Teknik Industri ITS, Jurnal Teknik ITS, Vol. 1., ISSN:2301-9271,
- Febriyana, (2011), Analisis Klaster K-Means dan K-Median pada Data Indikator Kemiskinan (studi kasus data indikator Kemiskinan Kabupaten di Indonesia tahun 2009), Skripsi, repository.uinjkt.ac.id., Tanggal akses : 2 Juli 2015
- Gujarati, D. N., (2004), *Basic Econometrics*, Fourth Edition, McGraw-Hill, New York.
- Hair, J. F, Jr. dkk, (2006), *Multivariate Data Analysis*, Sixth Edition, Pearson Prentice Hall, New Jersey.
- Johnson, R.A. dan Wichern, D. W., (1998), *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc., New Jersey.
- Kementrian Kesehatan Republik Indonesia, (2014), Profil Data Kesehatan Indonesia Tahun 2013, [www.depkes.go.id](http://www.depkes.go.id), Tanggal akses : 10 Juni 2015.
- Peraturan Menteri Kesehatan Republik Indonesia No. 492/Menkes/Per/IV/2010, [www.litbang.depkes.go.id](http://www.litbang.depkes.go.id)., Tanggal akses : 10 Juni 2015.
- Rismawan, T. dan Kusumadewi, S., (2008), Aplikasi K-Means untuk Pengelompokkan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) dan Ukuran Kerangka, Prosiding Seminar Nasional Aplikasi Teknologi Informasi, ISSN:1907-5022., [Journal.uii.ac.id/index.snati/article/](http://Journal.uii.ac.id/index.snati/article/), Tanggal akses : 2 Juli 2015.
- Santoso, S., (2012), *Aplikasi SPSS pada Statistik Multivariate*, Elex Media Komputindo, Jakarta.
- Unicef Indonesia, (2012), Air Bersih, Sanitasi dan Kebersihan, Ringkasan Kajian, Oktober 2012, [www.unicef.or.id](http://www.unicef.or.id), Tanggal akses : 14 Juni 2015.