

# Pengelompokan Rumah Tangga di Pulau Madura Berdasarkan Faktor-Faktor yang Mempengaruhi Konsumsi Air Bersih Menggunakan *Similarity Weight and Filter Method*

Astarani Wili Martha dan Ismaini Zain

Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data,

Institut Teknologi Sepuluh Nopember (ITS)

*e-mail:* ismaini\_z@statistika.its.ac.id

**Abstrak**—Clean Water and Sanitation adalah salah satu indikator pada SDGs yang berhubungan dengan kebutuhan air bersih. Tiga dari empat kabupaten di pulau Madura mengalami kekeringan, hal ini tentu menjadi sorotan untuk dilakukannya penelitian supaya terpenuhinya kebutuhan air bersih. Pulau Madura terdapat 3097 rumah tangga yang perlu mengonsumsi air bersih. Namun, belum semua rumah tangga mengakses air bersih. Penelitian ini bertujuan mengelompokan rumah tangga di Pulau Madura berdasarkan faktor yang mempengaruhi konsumsi air bersih dengan analisis kelompok. Pengelompokan berdasarkan data numerik dan data kategorik. Oleh karena itu digunakan pengelompokan dengan metode *Similarity Weight and Filter Method*. SWFM adalah salah satu metode pengelompokan campuran, pengelompokan data numerik menggunakan metode hirarki ward dan data kategorik menggunakan k-modes. Untuk analisis kelompok data numerik, terdapat 3 variabel dan diperoleh dua jumlah kelompok optimum menggunakan metode ward dengan nilai pseudo-F sebesar 1001,172. Analisis kelompok berskala kategorik terdapat 6 variabel dengan menggunakan k-modes dihasilkan tiga kelompok. Metode ensemble SWFM menghasilkan lima kelompok. Pengambilan lima kelompok karena menghasilkan nilai rasio terkecil pada jumlah kelompok sebesar 0,006627.

**Kata Kunci**—Air Bersih, Analisis Kelompok, K-modes, Rumah Tangga, SWFM.

## I. PENDAHULUAN

**S**USTAINABLE *Development Goals* (SDGs) merupakan program pembangunan berkelanjutan yang memiliki 17 tujuan dan 169 target pembangunan. Tujuan program SDGs diharapkan dapat lebih tanggap dalam menjawab masalah ketertinggalan pembangunan negara di seluruh dunia baik negara maju maupun negara berkembang seperti berkaitan dengan perubahan situasi dunia mengenai isu seperti kerusakan lingkungan, perubahan iklim, perlindungan sosial, deflasi sumber daya alam dan pembangunan lainnya yang berkaitan pada kemiskinan. Dari 17 tujuan, terdapat salah satu tujuan yang berhubungan dengan kebutuhan air bersih yaitu *Clean Water and Sanitation*. Permasalahan kebutuhan air bersih harus teratasi agar terpenuhinya tujuan SDGs hingga tahun 2030.

Pembangunan ekonomi tidak lepas dari studi akses penyediaan air bersih. Pembangunan adalah kemampuan

pemerintah untuk memenuhi kebutuhan dasar (*basic needs*), antara lain yaitu pangan, papan, kesehatan, dan perlindungan (termasuk akses terhadap air bersih) [1]. Jika salah satu diantaranya tidak ada, maka masyarakat akan mengalami kondisi kemiskinan. Menurut UNDP (2002), kemiskinan adalah suatu situasi atau kondisi seseorang (individu) tidak memiliki pendapatan untuk memenuhi kebutuhan dasar terdiri dari hak dapat menikmati kehidupan yang bermartabat dan hak yang diakui dalam peraturan perundang-undangan [2]. Salah satu hak dasar yang diakui adalah akses terhadap air bersih. Menurut Suparman, bahwa penduduk atau rumah tangga yang mempunyai status rumah tangga miskin justru jauh lebih sulit mengakses air bersih atau minum, ironisnya justru penduduk atau rumah tangga miskin membayar kebutuhan air minum lebih mahal daripada rumah tangga non miskin [3].

Menurut Badan Penanggulangan Bencana Daerah (BPBD) menyatakan bahwa 15 Kabupaten di Jawa Timur mengalami kekeringan diantaranya terdapat 3 Kabupaten di Pulau Madura. Hal ini menjadi salah satu permasalahan pada ketidakmerataan distribusi air bersih pada setiap daerah terutama daerah di Pulau Madura. Berdasarkan hasil kinerja PDAM tahun 2017, 3 dari 4 Kabupaten di Pulau Madura dikatakan kurang sehat [4]. Hal ini tentunya menjadi tantangan bagi pemerintah dan PDAM untuk meningkatkan pelayanan akses air bersih, mengingat dalam RPJMN tahun 2019 pemerintah mempunyai tujuan pelayanan 100% terhadap akses air minum layak berkelanjutan terhadap seluruh masyarakat.

Dalam representasi akses air bersih di Pulau Madura, diperlukan pengelompokan dalam pemenuhan kebutuhan air bersih berdasarkan faktor-faktor yang mempengaruhi konsumsi air bersih. Penelitian ini menggunakan variabel penelitian berskala numerik dan kategorik. *Similarity Weight and Filter Method* merupakan metode pengelompokan data campuran berskala numerik dan kategorik.

Berdasarkan hasil uraian tersebut, maka akan dilakukan penelitian tentang pengelompokan faktor-faktor yang mempengaruhi konsumsi air bersih rumah tangga di Pulau Madura menggunakan *Similarity Weight and Filter Method* (SWFM). Penelitian ini diharapkan bisa menjadi masukan Pemerintah dalam penanganan kebutuhan air bersih berdasarkan faktor yang terkait konsumsi air bersih pada rumah tangga di Pulau Madura.

II. TINJAUAN PUSTAKA

A. Statistika Deskriptif

Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Statistika deskriptif hanya memberikan informasi mengenai data yang tersedia dan tidak dapat digunakan untuk generalisasi gugus data induknya yang lebih besar. Statistika deskriptif juga dapat disajikan dalam susunan tabel, diagram, grafik, dan besaran-besaran lain [5]. Pada metode ini data dibedakan menjadi dua yaitu data kategorik dan numerik.

B. Analisis Kelompok (Cluster Analysis)

Analisis Kelompok (Cluster Analysis) merupakan salah satu teknik analisis multivariat yang bertujuan menempatkan sekumpulan objek ke dalam dua atau lebih cluster berdasarkan kesamaan karakteristiknya [6]. Analisis ini mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. Analisis kelompok mengelompokkan objek-objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam kelompok yang sama, serta mempunyai kemiripan satu dengan yang lain [7]. Hasil analisis kelompok dipengaruhi oleh beberapa hal, antara lain yaitu objek yang dikelompokkan, variabel yang diamati, ukuran kemiripan dan ketakmiripan, serta metode pengelompokan yang digunakan.

Ukuran kemiripan dan ketidakmiripan pada umumnya diukur berdasarkan jarak. Salah satu faktor yang sangat berpengaruh terhadap hasil dari kelompok yang dibentuk adalah jarak antara objek pengamatan [8]. Berikut ini merupakan metode-metode pengukuran jarak antara objek ke- $i$  ( $x_i$ ) dengan objek ke- $j$  ( $x_j$ ) berdasarkan karakteristik variabel yang dikelompokkan.

a. Jarak Euclidean

Metode pengelompokan data numerik didasarkan pada ukuran ketidakmiripan atau jarak. Ukuran ketidakmiripan yang biasa digunakan adalah jarak Euclidean. Penggunaan jarak Euclidean relatif mudah dimengerti dan dapat digunakan pada data yang memiliki jumlah variabel lebih dari dua yang ditunjukkan pada persamaan berikut.

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2} \quad (1)$$

Dimana

$$i = (1, 2, \dots, n) \text{ dan } j = (1, 2, \dots, n)$$

C. Pengelompokan Data Numerik

Metode hirarki digunakan bila banyaknya kelompok yang akan dibentuk tidak diketahui sebelumnya dan banyaknya amatan tidak terlalu besar. Terdapat dua teknik pengelompokan dalam analisis kelompok hirarki yaitu teknik pembagian (divisive) dan teknik penggabungan (agglomerative). Beberapa teknik pengelompokan antar kelompok dinyatakan sebagai berikut [7].

Metode ward menggunakan dasar pertimbangan untuk meminimalkan informasi yang hilang dari penggabungan dua kelompok. Jumlah kuadrat antara dua kelompok untuk seluruh variabel merupakan jarak antara dua kelompok, dimana metode ini meminimalkan variansi dalam kelompok. Nilai ESS dapat ditunjukkan pada persamaan (1).

$$ESS = \sum_{j=1}^N (x_j - \bar{x})' (x_j - \bar{x}) \quad (2)$$

Menentukan jumlah kelompok optimum merupakan tahapan penting setelah proses pengelompokan. Tahapan ini disebut sebagai validasi pengelompokan [8]. Indeks R-squared merupakan salah satu indeks yang dapat digunakan untuk menentukan jumlah kelompok optimum pada pengelompokan hirarki [9]. Indeks tersebut melibatkan perhitungan keragaman data baik keragaman total, keragaman dalam kelompok, maupun keragaman antar kelompok. Indeks validitas untuk menentukan jumlah kelompok optimum pada pengelompokan hirarki dapat dituliskan sebagai berikut.

$$SST = \sum_{l=1}^{m_{numerik}} \sum_{i=1}^n (x_{il} - \bar{x}_l)^2 \quad (3)$$

$$SSW = \sum_{c=1}^C \sum_{l=1}^{m_{numerik}} \sum_{i=1}^{n_c} (x_{ilc} - \bar{x}_{lc})^2 \quad (4)$$

$$SSB = SST - SSW \quad (5)$$

dengan,

$m_{numerik}$  : Jumlah variabel numerik dalam pengamatan,

$C$  : Jumlah kelompok yang dibentuk dalam pengamatan,

$n$  : Total jumlah objek pengamatan,

$n_c$  : Jumlah anggota pada kelompok ke- $c$  untuk  $c=1, 2, \dots, C$

$\bar{x}_l$  : Rata-rata keseluruhan objek pada variabel ke- $l$  untuk  $l=1, 2, \dots, m_{numerik}$ ,

$\bar{x}_{lc}$  : Rata-rata variabel ke- $l$  pada kelompok ke- $c$  untuk  $c=1, 2, \dots, C$

Nilai  $R^2 = 0$  menunjukkan bahwa tidak terdapat perbedaan antara kelompok, sedangkan  $R^2 = 1$  menunjukkan adanya perbedaan yang signifikan antar kelompok yang terbentuk.

$$R^2 = \frac{SSB}{SST} = \frac{[SST - SSW]}{SST} \quad (6)$$

Penentuan jumlah kelompok yang terbentuk dapat dilihat berdasarkan nilai maksimum dari Pseudo-F. Rumus yang digunakan untuk menghitung nilai pseudo-F statistics yaitu.

$$Pseudo-F = \frac{\left(\frac{R^2}{c-1}\right)}{\left(\frac{1-R^2}{n-c}\right)} \quad (7)$$

D. Pengelompokan Data Kategorik

Pengelompokan data bersifat kategorik dapat dilakukan dengan ukuran kemiripan data dengan menggunakan metode non-hirarki k-modes. Berdasarkan Kaufman dan Rousseeuw (1990), andaikan X dan Y adalah dua data dengan fitur bertipe kategorikal [10]. Ukuran ketidakmiripan di antara X dan Y

dapat diukur dengan jumlah ketidakcocokan nilai dari fitur yang berkorespondensi dari dua data. Semakin kecil nilai ketidakcocokan, maka semakin mirip dua data tersebut. Berikut merupakan rumus yang digunakan sebagaimana pada persamaan (8).

$$d(X, Y) = \sum_{j=1}^r \delta(x_j, y_j) \quad (8)$$

**E. Pengelompokan Data Campuran**

Pengelompokan SWFM memiliki konsep analisis yang sama dengan pengelompokan lainnya, yang membedakan adalah pada tahap pembentukan klaster akhir. tahapan metode *similarity weight* digunakan ukuran kemiripan yang memasukkan faktor bobot pada rumus ukuran kemiripan. Bobot yang diberikan tergantung pada jumlah anggota pengamatan ( $n_i$  atau  $n_j$ ). Rumus yang digunakan untuk menghitung ukuran kemiripan Antara pasangan objek ke- $i$  dan objek ke- $j$  adalah sebagaimana pada persamaan (9) [11].

$$sim(x_i, x_j) = \sum_{i \leq n_i, j \leq n_j} \frac{S_{ij}}{\max(n_i, n_j)}, i \neq j \quad (9)$$

$$S_{ij} = \left| \frac{X_i \cap X_j}{X_i \cup X_j} \right|$$

$X_i$  : Himpunan pengamatan ke- $i$   
dengan  $X_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{ki}\}$

$X_j$  : Himpunan pengamatan ke- $j$   
dengan  $X_j = \{x_{j1}, x_{j2}, x_{j3}, \dots, x_{kj}\}$

$|X|$  : Bilangan cardinal atau jumlah anggota dari himpunan  $X$

$m_k$  : Jumlah variabel kategorik dalam pengamatan

$n$  : Total jumlah objek pengamatan

$n_i$  : Jumlah anggota dalam kelompok ke- $i$

$n_j$  : Jumlah anggota dalam kelompok ke- $j$ .

Selanjutnya hasil pengelompokan data numerik dan kategorik digabungkan untuk memperoleh pengelompokan akhir menggunakan metode *filter* sebagaimana pada persamaan 10 [11].

$$F(X_i, X_j) = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w_{ij} d(X_i, X_j) \quad (10)$$

**F. Kinerja Hasil Pengelompokan**

Kinerja hasil pengelompokan untuk variabel dengan skala numerik dapat diketahui dari rasio nilai  $S_w$  dan  $S_B$  [12]. Dengan menggunakan nilai rata-rata variabel, simpangan baku di dalam kelompok ( $S_w$ ) dan simpangan baku antar kelompok ( $S_B$ ) dapat dirumuskan dalam persamaan (11) dan (12).

$$S_w = [MSW]^{1/2} \quad (11)$$

$$S_B = [MSB]^{1/2} \quad (12)$$

Seperti halnya data numerik kinerja suatu pengelompokan dengan data kategorik juga didasarkan pada perbandingan rasio antara simpangan baku dalam kelompok ( $S_w$ ) dan simpangan baku antar kelompok ( $S_B$ ). Dimana jika rasio perbandingan semakin kecil maka kinerja pengelompokan data kategorik semakin baik karena homogenitas maksimum dalam kelompok dan heterogenitas maksimum dalam kelompok

**G. Uji Normalitas**

Uji normalitas yang digunakan dalam penelitian ini menggunakan *Kolmogorov Smirnov* dan *mshapiro test*.

Pemeriksaan *Kolmogorov Smirnov* digunakan untuk memeriksa persebaran dari suatu variabel. Metode *mshapiro test* berfungsi untuk menguji apakah data berdistribusi normal multivariat. Hipotesis

$H_0$  : Data berasal dari distribusi normal

$H_1$  : Data tidak berasal dari distribusi normal.

Statistik uji *Kolmogorov Smirnov*:

$$D = \sup_x |F_n(x) - F_0(x)| \quad (13)$$

Keterangan:

$F_n(x)$  : nilai distribusi kumulatif sampel

$F_0(x)$  : nilai distribusi kumulatif bawah  $H_0$   $P(Z < Z_i)$

Daerah Kritis : Tolak  $H_0$  jika  $D > D_\alpha$

Metode *mshapiro test* dengan persamaan sebagai berikut.

$$W = \frac{\left( \sum_{i=1}^n a_i x_{(i)} \right)}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (14)$$

Dimana  $(a_1, a_2, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$ . Diambil keputusan

tolak  $H_0$  dengan ditetapkannya taraf signifikan sebesar  $\alpha$  maka  $H_0$  ditolak jika  $W$  kurang dari  $a$  [7].

**H. Uji Kruskal Wallis**

Uji *Kruskal Wallis* merupakan teknik nonparametrik yang digunakan untuk mengetahui apakah sampel berasal dari populasi yang identik. Beberapa asumsi untuk pengujian ini adalah pengamatan independen dan skala pengukuran paling tidak ordinal. Hipotesis yang digunakan adalah.

$H_0$  : Kedua populasi identik

$H_1$  : Tidak semua populasi identik

Statistik uji yang digunakan sesuai dengan persamaan 14.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^g \frac{R_i^2}{n_i} - 3(N+1) \quad (14)$$

Dimana

$R_i$  : Jumlah peringkat-peringkat yang ditetapkan bagi hasil-hasil pengamatan di kelompok ke- $i$

$n_i$  : Banyaknya pengamatan di tiap kelompok

$N$  : Banyaknya seluruh pengamatan

$K$  : Banyaknya kelompok

Pengambilan keputusan menggunakan nilai  $H$  yang dibandingkan dengan nilai  $\chi^2_{(0,05;K-1)}$ . Keputusan tolak  $H_0$  apabila nilai  $H$  lebih besar dari  $\chi^2_{(0,05;K-1)}$  atau nilai  $\chi^2_{hitung}$  lebih besar daripada  $\chi^2_{(0,05;K-1)}$  [13].

I. Faktor-Faktor yang Mempengaruhi Konsumsi Air Bersih

Menurut Winarna (2003), dalam penelitiannya menyebutkan bahwa faktor-faktor yang mempengaruhi konsumsi air bersih pelanggan rumah tangga di Kabupaten Karanganyar adalah pendapatan keluarga, pengeluaran total rumah tangga, jumlah anggota rumah tangga, pendidikan kepala keluarga dan ada atau tidaknya sumber air PDAM. Pengeluaran untuk konsumsi air tidak berbeda jauh dengan pengeluaran lain dalam rumah tangga. Penelitian yang dilakukan oleh Joachim dan Hillenbrand (2007), konsumsi air bersih rumah tangga dipengaruhi oleh tarif atau harga air, pendapatan per kapita, jumlah kepala rumah tangga, jumlah penduduk, musim atau iklim dan suhu. Selain itu, menurut Darr dan Kamen (1976), beberapa faktor yang menjadi pertimbangan dalam mempengaruhi konsumsi air rumah tangga antara lain yaitu ukuran rumah tangga, pendapatan per kapita, perkembangan wilayah, tipe meteran, pendidikan responden, kepadatan ruang dan cakupan pelayanan. Namun menurut Kementerian Perempuan (2010) tidak demikian. Seringkali perempuan banyak berhubungan dengan kebutuhan air baik untuk kepentingan individu maupun rumah tangga.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder. Data diperoleh dari SUSENAS Jawa Timur tahun 2016. Unit penelitian yang digunakan adalah rumah tangga di empat Kabupaten di pulau Madura dengan jumlah rumah tangga yang digunakan sebanyak 3.097 rumah tangga.

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah sebagai berikut. Terdapat 9 variabel penelitian diantaranya adalah tiga data berskala numerik dan enam data berskala kategorik.

Tabel 1. Variabel Penelitian

Variabel	Nama	Keterangan
X <sub>1</sub>	Total Pengeluaran RT	-
X <sub>2</sub>	Proporsi ART Perempuan	-
X <sub>3</sub>	Jumlah ART	-
X <sub>4</sub>	Pendidikan Terakhir KRT	1 : TTSD
		2 : Rendah
		3 : Menengah
		4 : Tinggi

Tabel 1. Variabel Penelitian (Lanjutan)

Variabel	Nama	Keterangan
X <sub>5</sub>	Sumber Air Minum	1 : DB

X <sub>6</sub>	Sumber Air Masak	2 : TDB
		1 : DB
X <sub>7</sub>	Sumber Air Mandi/Cuci	1 : DB
		2 : TDB
X <sub>8</sub>	Sistem Saluran Sumber Air	1 : MG
		2 : TMG
X <sub>9</sub>	Pembayaran Air PAM	1 : TMB
		2 : MB

Definisi operasional dari variabel penelitian yang digunakan sebagai berikut :

1. Total pengeluaran sebulan adalah jumlah pengeluaran yang dikeluarkan oleh rumah tangga untuk konsumsi kebutuhan lainnya.
2. Proporsi anggota rumah tangga perempuan adalah banyaknya anggota rumah tangga perempuan yang tinggal dalam satuan persen.
3. Jumlah anggota rumah tangga adalah jumlah anggota rumah tangga yang tinggal di tempat tinggal tersebut.
4. Pendidikan terakhir kepala rumah tangga merupakan pendidikan tertinggi yang ditamatkan oleh kepala rumah tangga.

Kategori :

- 1) TTSD (Tidak Tamat SD) adalah pendidikan terakhir kepala rumah tangga dengan tidak bersekolah.
- 2) Rendah adalah pendidikan terakhir kepala rumah tangga dengan tingkat pendidikan SD/setara dan SMP/setara.
- 3) Menengah adalah pendidikan terakhir kepala rumah tangga dengan tingkat pendidikan SMA/setara.
- 4) Tinggi adalah pendidikan terakhir kepala rumah tangga dengan tingkat pendidikan tinggi (PT).
5. Sumber air minum, air masak, dan air mandi merupakan sumber air utama yang digunakan oleh rumah tangga di Pulau Madura.

Kategori:

- 1) DB (Dapat dibeli) adalah sumber air yang dapat dibeli seperti kategori air bermerk, air isi ulang, ledeng meteran dan ledeng eceran.
- 2) TDB (Tidak dapat dibeli) adalah sumber air yang tidak dapat dibeli seperti air sumur bor/pompa, sumur terlindung, sumur tak terlindung, mata air terlindung, mata air tak terlindung, air hujan, air permukaan dan lainnya.
6. Sistem saluran sumber air merupakan penggunaan saluran air yang digunakan oleh rumah tangga di Pulau Madura.

Kategori:

- 1)MG (Menggunakan) adalah rumah tangga dengan menggunakan sistem saluran sumber air perpipaan atau hidran umum.
- 2)TMG (Tidak Menggunakan) adalah rumah tangga tidak menggunakan atau tidak tahu sistem saluran sumber air perpipaan/hidran umum.
7. Pembayaran air PAM merupakan rumah tangga yang membayar air PAM.

Kategori :

- 1) TMB (Tidak Membayar) adalah rumah tangga yang tidak menggunakan dan tidak membayar air PAM.

- 2) MB (Membayar) adalah rumah tangga yang menggunakan dan membayar air PAM.

### C. Langkah Analisis

Langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut.

1. Membuat statistika deskriptif pada data untuk variabel berskala numerik dengan menentukan nilai maksimum dan nilai minimum serta *mean* dari setiap variabel. Sedangkan untuk data berskala kategorik dengan grafik.
2. Mengelompokkan rumah tangga di Pulau Madura menggunakan metode *Similarity Weight and Filter Method*.
  - a. Membagi variabel penelitian menjadi data kategorik dan data numerik.
  - b. Melakukan kluster pada data numerik dengan metode hirarki *agglomerative ward*. Jarak yang digunakan adalah jarak *Euclidean* dan pautan yang digunakan adalah *average linkage*.
  - c. Menentukan jumlah kelompok optimum, menghitung dan memilih kinerja pengelompokan berdasarkan nilai *pseudo-F* maksimum pada masing-masing hasil pengelompokan numerik.
  - d. Mengelompokkan variabel yang berskala kategorik dengan menggunakan metode *K-modes*.
  - e. Mengkombinasikan hasil kluster pada tahap c dan d.
  - f. Melakukan pengolahan data dengan menggunakan *Similarity Weight Method*.
  - g. Mengelompokkan data menggunakan *Similarity Analysis*.
  - h. Menentukan *Filter Algorithm* untuk data kategorik dan numerik.
  - i. Mengelompokkan data menggunakan *Similarity Weight and Filter Method*.
  - j. Menentukan jumlah kelompok optimum berdasarkan nilai rasio Sw dan Sb terkecil.
3. Melakukan pengujian dengan uji normalitas dan uji kruskal wallis.
4. Menginterpretasikan hasil dan menarik kesimpulan dan saran.

## IV. ANALISIS DAN PEMBAHASAN

### A. Karakteristik Variabel Penelitian Berskala Numerik

Pada penelitian ini digunakan variabel numerik berdasarkan faktor-faktor yang mempengaruhi kebutuhan air bersih rumah tangga antara lain adalah total pengeluaran untuk air sebulan, total pengeluaran rumah tangga ( $X_1$ ), proporsi anggota rumah tangga perempuan ( $X_2$ ), dan jumlah anggota rumah tangga ( $X_3$ ). Hasil analisis statistika deskriptif ditampilkan pada Tabel 2.

Tabel 2.  
Karakteristik Data pada Ketiga Variabel Prediktor

Variabel	Rata-rata	Varians	Nilai tengah	Min.	Maks.
$X_1$ (juta)	2.414	2.068	1.929	0.179	29.612
$X_2$ (persen)	0,55	0,22	0,5	0	1
$X_3$ (orang)	4	2	4	1	14

Variabel total pengeluaran rumah tangga memiliki rata-rata sebesar sebesar 2.414 juta rupiah per rumah tangga dengan total pengeluaran minimal 179 ribu rupiah dengan rumah

tangga yang tinggal berada di Kabupaten Bangkalan. Sedangkan total pengeluaran terbesar sebesar 29.612 rupiah dengan rumah tangga yang berada di Kabupaten Sumenep.

Proporsi anggota rumah tangga perempuan yang tinggal memiliki rata-rata 0,55 persen dengan proporsi terkecil sebesar 0 atau tidak ada anggota rumah tangga yang tinggal. Sedangkan proporsi terbesar sebesar 1 persen dengan proporsi anggota rumah tangga perempuan terbanyak adalah Kabupaten Sumenep.

Jumlah anggota rumah tangga memiliki rata-rata sebesar 4 orang yang tinggal dengan jumlah anggota rumah tangga minimal sebesar 1 orang, anggota rumah yang beranggotakan 1 orang terbanyak berada di Kabupaten Sumenep. Sedangkan jumlah anggota rumah tangga terbanyak sebesar 14 orang, dengan rumah tangga yang tinggal berada di Kabupaten Pamekasan.

### B. Karakteristik Variabel Penelitian Berskala Kategorik

Pada penelitian ini digunakan variabel kategorik antara lain adalah pendidikan kepala rumah tangga, sumber air minum, sumber air masak, sumber air mandi/cuci, penggunaan saluran sumber air dan pembayaran air PAM.

Variabel kategorik pada penelitian ini pendidikan terakhir kepala rumah tangga adalah pendidikan rendah sebesar 52,57%. Sedangkan sumber air minum, masak dan mandi/cuci yang digunakan lebih banyak menggunakan sumber air yang tidak dapat dibeli. Penggunaan saluran sumber air yang lebih banyak digunakan adalah perpipaan/ hidran umum dengan sebagian besar rumah tangga tidak membayar air PAM.

### C. Pengelompokan Data Berskala Numerik

Pengelompokan untuk data numerik dilakukan menggunakan teknik *average linkage* dengan setiap objek pengamatan sebagai suatu kelompok dengan anggota tunggal.

Jumlah kelompok optimum berada pada 5 kelompok dengan metode *ward*. Jumlah kelompok yang dibentuk antara dua sampai enam kelompok dengan dilakukan menggunakan *software R*. Setelah didapatkan hasil pengelompokan dari dua sampai enam kelompok, kemudian didapatkan nilai *Pseudo F-statistic*. Hasil nilai *pseudo-F* didapatkan jumlah kelompok optimum dengan terbentuk 2 kelompok dengan nilai *pseudo-F* terbesar adalah 1001,172, karena nilai *pseudo-F* lebih besar dari jumlah kelompok lain.

Pembagian jumlah anggota kelompok dengan metode hirarki *ward* untuk jumlah 2 kelompok berturut-turut sebanyak 1.782 dan 1.315 rumah tangga.

Sedangkan karakteristik digunakan nilai rata-rata dari masing-masing kelompok yang terbentuk. Berikut hasil dari rata-rata jumlah kelompok 3 adalah sebagai berikut.

Tabel 3.  
Karakteristik Data Hasil Pengelompokan Variabel Berskala Numerik

Variabel	Kelompok 1	Kelompok 2
$X_1$ (juta)	3.36	1.13
$X_2$ (persen)	0,51	0,61
$X_3$ (orang)	4	3

Tabel 3 menunjukkan nilai rata-rata dari masing-masing anggota kelompok. Berdasarkan Tabel 3, kelompok 1 terlihat memiliki nilai rata-rata tertinggi pada variabel total

pengeluaran rumah tangga dan jumlah anggota rumah tangga. Kemudian untuk kelompok 2 terlihat memiliki nilai rata-rata terkecil untuk variabel total pengeluaran rumah tangga dan jumlah anggota rumah tangga. Tetapi nilai rata-rata untuk variabel proporsi anggota rumah tangga perempuan untuk kelompok 2 lebih besar. Dengan mengetahui hasil karakteristik dari pengelompokan variabel berskala numerik, maka dapat dilihat dari masing-masing kelompok menunjukkan hasil karakteristik berdasarkan perekonomian rumah tangga di Pulau Madura. Hal tersebut terlihat dari karakteristik rumah tangga pada kelompok 2 yang dominan termasuk perekonomian rendah. Kemudian untuk karakteristik kelompok 1 yang dominan termasuk perekonomian tinggi.

Selain mengetahui karakteristik dari masing-masing kelompok per rumah tangga di Pulau Madura, perlu diketahui juga masing-masing hasil kelompok rumah tangga yang tinggal di kabupaten di pulau Madura berdasarkan kriteria atau ciri-ciri berdasarkan karakteristik yang terbentuk. Hasil kelompok rumah tangga dengan perekonomian rendah lebih banyak adalah rumah tangga yang tinggal di kabupaten Sumenep.

**D. Pengelompokan Data Berskala Kategorik**

Pengelompokan pada data variabel kategorik dengan melakukan metode *k-modes* dengan setiap kelompok objek pengamatan sebagai suatu kelompok tunggal dengan anggota tunggal. Penelitian ini menggunakan *software R* untuk melakukan analisis *k-modes*. Nilai *k* atau jumlah kelompok yang digunakan pada analisis ini ditentukan sebesar 3. Pengelompokan data lebih banyak masuk kedalam jumlah kelompok 1, sehingga masing-masing pembagian jumlah kelompok 3 berturut-turut sebanyak 1.975, 723, dan 399 rumah tangga.

Hasil pengelompokan berdasarkan karakteristik dengan persentase kategori yang paling dominan antara lain.

Tabel 4.

Karakteristik Data Hasil Pengelompokan Variabel Berskala Kategorik				
Variabel	Kel 1	Kel 2	Kel 3	
X <sub>4</sub>	TTSD	48,81%	0,00%	16,29%
	Rendah	39,59%	86,86%	54,64%
	Menengah	8,15%	11,34%	12,53%
	Tinggi	3,44%	1,80%	16,54%
X <sub>5</sub>	DB	9,47%	7,19%	97,99%
	TDB	90,53%	92,81%	2,01%
X <sub>6</sub>	DB	0,96%	0,69%	96,24%
	TDB	99,04%	99,31%	3,76%
X <sub>7</sub>	DB	0,15%	0,14%	79,20%
	TDB	98,85%	99,86%	20,80%
X <sub>8</sub>	MG	70,78%	0,00%	90,48%
	TMG	29,22%	100,00%	9,52%
X <sub>9</sub>	TMB	87,80%	89,21%	15,29%
	MB	12,20%	10,79%	84,71%

Tabel 4 merupakan penyajian persentase dari kategori yang paling dominan dari variabel-variabel yang ada pada kelompok. Kelompok 1 dapat dilihat bahwa sebagian besar tidak bersekolah dengan penggunaan sumber air minum, air masak dan air mandi/cuci yang paling banyak adalah air yang tidak dapat dibeli kemudian sistem saluran sumber air yang digunakan adalah perpipaan atau hidran umum dan untuk rumah tangga tidak melakukan pembayaran air PAM. Kelompok 2 dapat dilihat sebagian besar berpendidikan rendah

dengan sumber air minum, air masak dan air mandi/cuci paling banyak digunakan adalah air yang tidak dapat dibeli, kemudian saluran sumber air tidak menggunakan sistem perpipaan maupun hidran umum dan untuk rumah tangga tidak melakukan pembayaran air PAM. Kelompok 3 sebagian besar berpendidikan rendah dengan penggunaan sumber air minum, air masak dan air mandi/cuci yang digunakan adalah air yang dapat dibeli kemudian sistem saluran sumber air menggunakan perpipaan atau hidran umum dan untuk rumah tangga melakukan pembayaran air PAM.

Berdasarkan hasil karakteristik pengelompokan data kategorik, didapatkan kriteria atau ciri-ciri dari masing-masing kelompok berturut-turut adalah perekonomian rendah, perekonomian sedang, dan perekonomian tinggi. Hasil kelompok rumah tangga dengan perekonomian rendah lebih banyak adalah rumah tangga yang tinggal di kabupaten Pamekasan.

**E. Pengelompokan Data Campuran**

Pengelompokan data campuran menggunakan SWFM (*Similaruty Weight and Filter Method*) dengan melakukan pengelompokan masing-masing pada hasil pengelompokan sebelumnya (data numerik dan data kategorik). Tahap pertama yang dilakukan dalam analisis pengelompokan ensemble SWFM untuk data campuran adalah dengan melakukan pengelompokan masing-masing jenis data menggunakan metodenya masing-masing.

Hasil pengelompokan terbaik didapatkan nilai rasio terkecil berada pada jumlah kelompok lima sebesar 0,006672. Pembagian jumlah anggota kelompok dengan metode SWFM untuk jumlah 5 kelompok berturut-turut adalah 1.063, 1.395, 276, 272, dan 127 rumah tangga.

Karakteristik variabel penelitian berskala numerik dan kategorik dari anggota hasil pengelompokan pada analisis SWFM pada masing-masing kelompok disajikan pada Tabel 5 dan Tabel 6 sebagai berikut.

Tabel 5.

Karakteristik Data Numerik pada Hasil Pengelompokan SWFM					
Variabel	Karakteristik Kelompok ke-				
	1	2	3	4	5
X <sub>1</sub> (juta)	3.30	1.71	1.20	4.41	1.23
X <sub>2</sub> (persen)	0,51	0,58	0,57	0,51	0,64
X <sub>3</sub> (orang)	4	3	3	4	3

Tabel 5 menunjukkan nilai rata-rata dari masing-masing anggota kelompok. Berdasarkan Tabel 5, kelompok 4 dan kelompok 3 terlihat memiliki nilai rata-rata tertinggi pada variabel total pengeluaran rumah tangga dan jumlah anggota rumah tangga. Kemudian untuk kelompok 3 dan kelompok 5 terlihat memiliki nilai rata-rata terkecil untuk variabel total pengeluaran rumah tangga dan jumlah anggota rumah tangga. Akan tetapi nilai rata-rata untuk variabel proporsi anggota rumah tangga perempuan untuk kelompok 5 dan kelompok 2 lebih besar dibandingkan dengan kelompok lainnya.

Tabel 6.

Karakteristik Data Kategorik pada Hasil Pengelompokan SWFM	
Variabel	Karakteristik kelompok ke-

	1	2	3	4	5
X <sub>4</sub>	Rendah (44,03%)	Rendah (50,99%)	Rendah (90,22%)	Rendah (52,21%)	Rendah (50,84%)
X <sub>5</sub>	TDB (84,01%)	TDB (95,66%)	TDB (96,38%)	DB (98,53%)	DB (96,85%)
X <sub>6</sub>	TDB (98,40%)	TDB (99,48%)	TDB (100,00%)	DB (96,69%)	DB (95,28%)
X <sub>7</sub>	TDB (99,91%)	TDB (99,85%)	TDB (99,64%)	DB (84,19%)	DB (68,50%)
X <sub>8</sub>	MG (76,39%)	TMG (56,88%)	TMG (100,00%)	MG (94,12%)	MG (82,68%)
X <sub>9</sub>	TMB (87,02%)	TMB (88,67%)	TMB (90,22%)	MB (79,02%)	MB (96,85%)

Tabel 6 merupakan penyajian persentase dari kategori yang paling dominan dari variabel-variabel yang ada pada hasil kelompok. Dari hasil kelompok didapatkan dominan pendidikan kepala rumah tangga adalah berpendidikan rendah. Kelompok 1, kelompok 2 dan kelompok 3 dapat dilihat bahwa sebagian besar penggunaan sumber air minum, air masak dan air mandi/cuci adalah air yang tidak dapat dibeli dan sebagian besar tidak membayar air PAM. Kelompok 2 dan kelompok 3 sebagian besar tidak menggunakan sistem saluran air. Kelompok 4 dan kelompok 5 sebagian besar menggunakan air yang dapat dibeli untuk sumber air minum, air masak dan air mandi/cuci dan sebagian besar rumah tangga membayar air PAM. Rumah tangga yang berada di kelompok 1, kelompok 4 dan kelompok 5 cenderung menggunakan sistem saluran air.

Dengan mengetahui hasil karakteristik dari pengelompokan menggunakan metode SWFM, maka dapat dilihat dari masing-masing kelompok menunjukkan hasil karakteristik berdasarkan perekonomian rumah tangga di Pulau Madura. Hal tersebut terlihat dari karakteristik rumah tangga yaitu  
 Kelompok 1 : Perekonomian tinggi.  
 Kelompok 2 : Perekonomian sedang.  
 Kelompok 3 : Perekonomian sangat rendah.  
 Kelompok 4 : Perekonomian paling tinggi.  
 Kelompok 5 : Perekonomian rendah.

Selain mengetahui karakteristik dari kelompok per rumah tangga di Pulau Madura, didapatkan masing-masing kriteria atau ciri-ciri berdasarkan perekonomian rumah tangga yang tinggal di Kabupaten di Pulau Madura. Berikut hasil kelompok rumah tangga yang tinggal di kabupaten di Pulau Madura untuk pengelompokan data campuran dapat ditampilkan pada Tabel 7.

Tabel 7.  
 Hasil Kelompok Data Campuran pada Rumah Tangga yang Tinggal disetiap Kabupaten di Pulau Madura

Kabupaten	Kelompok ke-					Total
	1	2	3	4	5	
Bangkalan	307	291	70	61	17	746
Sampang	300	292	64	80	22	758
Pamekasan	255	355	44	56	47	757
Sumenep	201	421	98	75	41	836
Total	1.063	1.359	276	272	127	3.097

Tabel 7 merupakan penyajian hasil kelompok rumah tangga yang tinggal disetiap Kabupaten dengan menggunakan metode SWFM berdasarkan perekonomian rumah tangga. Berdasarkan Tabel 7, jumlah rumah tangga di Kabupaten Sumenep lebih

sedikit berada pada kelompok 1 dengan kriteria perekonomian tinggi. Rumah tangga di Kabupaten Pamekasan lebih sedikit jumlahnya untuk kriteria perekonomian paling tinggi.

F. Pengujian Beda Rata-rata Antar Kelompok

Setelah diperoleh kelompok yang optimum, kemudian dilakukan pengujian perbedaan rata-rata antar kelompok. Pengujian dilakukan untuk mengetahui apakah hasil pengelompokan telah memberikan kelompok-kelompok yang berbeda secara signifikan. Dalam menguji perbedaan antar klaster memperhatikan variabel yang berskala numerik. Pengujian dilakukan secara nonparametrik dengan uji *Kruskal Wallis*. Hal ini dilakukan karena variabel penelitian yang digunakan tidak berdistribusi normal per variabel dengan pemeriksaan *Kolmogorov smirnov*, dan hasil *mshapiro test* yaitu nilai *p-value* adalah  $2,2 \times 10^{-16}$  maka didapatkan data tidak berdistribusi normal. Uji *Kruskal Wallis* digunakan pada tiga variabel penelitian berskala numerik. Hipotesis yang digunakan pada pengujian *Kruskal Wallis* adalah sebagai berikut.

H<sub>0</sub> : Tidak ada perbedaan rata-rata antar kelompok

H<sub>1</sub> : adanya perbedaan rata-rata antar kelompok

Ringkasan uji *Kruskal Wallis* pada variabel penelitian ditampilkan pada Tabel 8.

Tabel 8.  
 Hasil Uji *Kruskal Wallis* terhadap 5 Kelompok

Variabel	Chi-Square	df	Chi-Square Tabel
X <sub>1</sub>	1408,271	4	9,49
X <sub>2</sub>	60,302	4	9,49
X <sub>3</sub>	442,726	4	9,49

Uji *Kruskal Wallis* ini menggunakan nilai  $\alpha$  sebesar 0,05. Dari pengujian pada Tabel 8, tiga variabel penelitian yang digunakan diantaranya nilai *chi-square* yang lebih besar dari nilai *chi-square* tabel, sehingga dapat disimpulkan tiga variabel tersebut signifikan berbeda antara rumah tangga di lima kelompok.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Setelah dilakukan analisis dan pembahasan mengenai pengelompokan rumah tangga di pulau Madura menggunakan tiga metode, dapat diambil kesimpulan bahwa hasil karakteristik data numerik dan kategorik didapatkan rata-rata total pengeluaran sebesar 2.414 juta dan sebagian besar sumber air yang digunakan adalah sumber air yang tidak dapat dibeli dengan saluran air yang digunakan sebagian menggunakan perpipaan/hidran umum. Rumah tangga di pulau Madura sebagian besar tidak membayar air PAM.

Pengelompokan data berskala numerik menggunakan metode hirarki *agglomerative ward* terbentuk dua kelompok berdasarkan hasil nilai *pseudo-f*, dengan masing-masing kriteria yaitu perekonomian tinggi (kelompok 1) dan perekonomian rendah (kelompok 2). Dengan jumlah rumah tangga yang tinggal dalam kriteria perekonomian sebanyak 1.782 rumah tangga dan kriteria perekonomian rendah sebanyak 1.315 rumah tangga. Hasil kelompok dengan jumlah

paling banyak pada rumah tangga berkriteria perekonomian rendah adalah rumah tangga yang tinggal di Kabupaten Sumenep.

Pada hasil pengelompokan data berskala kategorik menggunakan metode *k-modes*, terbentuk tiga kelompok dengan masing-masing kriteria kelompok perekonomian rendah, sedang dan tinggi. Jumlah rumah tangga yang tinggal dalam kelompok 1, kelompok 2 dan kelompok 3 berturut-turut sebesar 1.975, 723, dan 399 rumah tangga. Hasil kelompok dengan jumlah paling banyak pada rumah tangga dengan ciri-ciri perekonomian rendah adalah rumah tangga yang tinggal di Kabupaten Pamekasan.

Pengelompokan data berskala campuran yaitu numerik dan kategorik menggunakan metode SWFM terbentuk lima kelompok berdasarkan nilai rasio  $S_w$  dan  $S_b$ , dengan kriteria berturut-turut adalah perekonomian tinggi, perekonomian sedang, perekonomian sangat rendah, perekonomian paling tinggi dan perekonomian rendah. Hasil kelompok dengan jumlah paling banyak pada rumah tangga dengan ciri-ciri perekonomian rendah dan sangat rendah adalah rumah tangga yang tinggal di Kabupaten Pamekasan dan Sumenep.

#### B. Saran

Penelitian yang telah dilakukan ini masih terdapat beberapa perbaikan dan pengembangan selanjutnya, diantaranya adalah.

##### 1. Untuk Pemerintah Kabupaten di Pulau Madura

Hasil penelitian menunjukkan kesimpulan, rumah tangga yang tinggal di Kabupaten Pamekasan dan Sumenep merupakan rumah tangga dengan ciri-ciri perekonomian

rendah maupun sangat rendah, sehingga perlu diperhatikan.

2. Untuk perbaikan dan pengembangan pada penelitian selanjutnya perlu penambahan variabel penelitian yang digunakan. Selain itu, variabel yang digunakan sebaiknya dicoba menggunakan uji *Kruskal Wallis* multivariat.

#### DAFTAR PUSTAKA

- [1] M. Todaro, *Pengembangan Ekonomi*. Jakarta: Bumi Aksara, 2000.
- [2] ESCAP-UNDP, "ESCAP-UNDP Initiative for The Achievement of Millenium Development Goals in Asia and The Pacific," Bangkok, 2002.
- [3] S. Suparman, "Kaum Perempuan Paling Peduli." Percik, Media Informasi Air Minum dan Penyehatan Lingkungan, pp. 9–11, 2007.
- [4] BPPSPAM, "Buku Kinerja PDAM 2017," Jakarta, 2017.
- [5] R. Walpole, *Pengantar Metode Statistika*, 3rd ed. Jakarta: Bumi Aksara, 1995.
- [6] B. Simamora, *Analisis Multivariat Pemasaran*. Jakarta: Gramedia Pustaka Umum, 2005.
- [7] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey: Person Prentice Hall, 2007.
- [8] S. Sharma, *Applied Multivariate Techniques*. Canada: John Wiley & Sons Inc, 1996.
- [9] B. Halkidi and Vizirgiannis, "On Clustering Validation Techniques," *J. Intell. Syst.*, vol. 17, no. 2/3, pp. 107–145, 2001.
- [10] L. Kaufman and P. Rousseeuw, *Finding Groups in Data-An Introduction to Cluster Analysis*. Wiley, 1990.
- [11] M. Reddy and B. Kavitha, "Clustering the Mixed Numerical and Categorical Dataset using Similarity Weight and Filter Method," vol. 5, no. 1, 2012.
- [12] M. Bunkers and R. James, "Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique.," 1996.
- [13] W. Daniel, *Statistik Nonparametrik Terapan (Terjemahan)*. Jakarta: Gramedia, 1989.