



TUGAS AKHIR - SS141501

**PENGELOMPOKAN RUMAH TANGGA
DI PULAU MADURA BERDASARKAN
FAKTOR-FAKTOR YANG MEMPENGARUHI
KONSUMSI AIR BERSIH MENGGUNAKAN
*SIMILARITY WEIGHT AND FILTER METHOD***

**ASTARANI WILI MARTHA
NRP 062114 4000 0107**

**Dosen Pembimbing
Dr. Dra. Ismaini Zain, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018**



TUGAS AKHIR - SS 141501

**PENGELOMPOKAN RUMAH TANGGA
DI PULAU MADURA BERDASARKAN
FAKTOR-FAKTOR YANG MEMPENGARUHI
KONSUMSI AIR BERSIH MENGGUNAKAN
*SIMILARITY WEIGHT AND FILTER METHOD***

**ASTARANI WILI MARTHA
NRP 062114 4000 0107**

**Dosen Pembimbing
Dr. Dra. Ismaini Zain, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018**

(Halaman ini sengaja dikosongkan)



FINAL PROJECT - SS 141501

**CLUSTERING HOUSEHOLD IN MADURA ISLAND
BASED ON FACTORS THAT AFFECT WATER
CONSUMPTION USING *SIMILARITY WEIGHT
AND FILTER METHOD***

**ASTARANI WILI MARTHA
SN 062114 4000 0107**

**Supervisor
Dr. Dra. Ismaini Zain, M.Si.**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018**

(Halaman ini sengaja dikosongkan)

LEMBAR PENGESAHAN
PENGELOMPOKAN RUMAH TANGGA
DI PULAU MADURA BERDASARKAN
FAKTOR-FAKTOR YANG MEMPENGARUHI
KONSUMSI AIR BERSIH MENGGUNAKAN
SIMILARITY WEIGHT AND FILTER METHOD

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember

Oleh :

Astarani Wili Martha
NRP. 062114 4000 0107

Disetujui oleh Pembimbing :
Dr. Dra. Ismaini Zain, M. Si
NIP. 19600525 198803 2 001



Mengetahui,
Kepala Departemen



Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, JULI 2018

(Halaman ini sengaja dikosongkan)

**PENGELOMPOKAN RUMAH TANGGA
DI PULAU MADURA BERDASARKAN
FAKTOR-FAKTOR YANG MEMPENGARUHI
KONSUMSI AIR BERSIH MENGGUNAKAN
SIMILARITY WEIGHT AND FILTER METHOD**

Nama Mahasiswa : Astarani Wili Martha
NRP : 062114 4000 0107
Departemen : Statistika FMKSD-ITS
Dosen Pembimbing : Dr. Dra. Ismaini Zain, M.Si.

Abstrak

Clean Water and Sanitation adalah salah satu indikator 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 menggunakan 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 menggunakan *k-modes* dihasilkan tiga kelompok. Metode ensemble SWFM menghasilkan lima kelompok dengan nilai rasio terkecil pada jumlah kelompok sebesar 0,006627.

Kata Kunci : Air Bersih, Analisis Kelompok, K-modes, Rumah Tangga, SWFM

(Halaman ini sengaja dikosongkan)

**CLUSTERING HOUSEHOLD IN MADURA ISLAND
BASED ON FACTORS THAT AFFECT WATER
CONSUMPTION USING *SIMILARITY WEIGHT AND
FILTER METHOD***

Student Name : Astarani Wili Martha
Student Number : 062114 4000 0107
Department : Statistics FMKSD-ITS
Supervisor : Dr. Dra. Ismaini Zain, M.Si

Abstract

Clean Water and Sanitation is one of the indicators on SDGs related to clean water needs. Three of the four districts on the island of Madura suffer from drought, this is certainly the spotlight for doing research to meet the needs of clean water. Madura Island there are 3097 households that need to consume clean water. However, not all households have access to clean water. This research aims to clustering households in Madura Island based on factors influencing the consumption of clean water by cluster analysis. Grouping based on numerical data and categorical data. Therefore we use grouping with Similarity Weight and Filter Method method. SWFM is one of the mixed grouping methods, numerical data grouping using ward hierarchy method and categorical data using k-modes. Cluster analysis for numerical data, there are 3 variables and two optimum groups were obtained using ward method with pseudo-F value of 1001,172. Cluster analysis of categorical scale there are 6 variables by using k-modes generated three groups. The SWFM ensemble method yields five groups. Taking five groups as it yields the smallest value of the ratio in the group number of 0,006627.

Keywords : *Clean Water, Cluster Analysis, Households, K-modes, SWFM*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT karena berkat rahmat dan berkat-Nya penulis dapat menyelesaikan laporan Tugas Akhir dengan judul

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

Penyusunan dan penulisan laporan Tugas Akhir ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Orang tua serta keluarga penulis, yang telah memberikan dukungan sehingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
2. Bapak Dr. Suhartono selaku Kepala Departemen Statistika ITS dan Bapak Dr. Sutikno, M.Si selaku Ketua Program Studi Sarjana Departemen Statistika ITS yang telah menyediakan fasilitas guna kelancaran pengerjaan Tugas Akhir ini.
3. Ibu Dr. Dra. Ismaini Zain, M.Si. selaku dosen pembimbing yang telah membimbing dan memberikan arahan serta masukan kepada penulis.
4. Bapak Prof. Dr. Drs. I Nyoman Budiantara, M.Si., dan Ibu Erma Oktania Permatasari, S.Si., M.Si., selaku dosen penguji yang telah memberikan masukan untuk kesempurnaan tugas akhir ini.
5. Bapak Dr. Bambang Widjanarko Otok, S.Si., M.Si. selaku dosen wali selama masa perkuliahan yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika FMKSD ITS.
6. Teman-teman seperjuangan TA “Pejuang Bu Is” yang saling membantu dalam pembuatan Tugas Akhir.

7. Teman-teman Statistika FMKSD ITS angkatan 2014, Respect yang selalu senantiasa memberikan semangat dan doanya kepada penulis.
8. Semua pihak yang telah membantu dalam penulisan laporan ini, yang tidak dapat penulis sebutkan satu per satu. Penulis sangat berharap hasil Tugas Akhir ini dapat bermanfaat bagi kita semua serta saran dan kritik bersifat membangun guna perbaikan di masa mendatang.

Surabaya, Juli 2018

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	Error! Bookmark not defined.
LEMBAR PENGESAHAN.....	Error! Bookmark not defined.
Abstrak.....	vii
<i>Abstract</i>	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR TABEL.....	xv
DAFTAR GAMBAR.....	xvii
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	4
1.4 Manfaat.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Statistika Deskriptif.....	7
2.2 Analisis Kelompok (<i>Cluster Analysis</i>).....	7
2.3 Metode Pengelompokan.....	9
2.3.1 Pengelompokan Data Numerik.....	9
2.3.2 Pengelompokan Data Kategorik.....	12
2.4 Pengelompokan Data Campuran Numerik dan Kategorik.....	14
2.5 Pengelompokan <i>Similarity Weight and Filter Method</i> (SWFM).....	15
2.6 Kinerja Hasil Pengelompokan.....	17
2.7 Uji Normalitas.....	20
2.8 Uji <i>Kruskal Wallis</i>	21
2.9 Faktor-Faktor yang Mempengaruhi Konsumsi Air Bersih.....	22

	Halaman
BAB III METODOLOGI PENELITIAN	25
3.1 Sumber Data	25
3.2 Variabel Penelitian.....	25
3.3 Langkah Analisis.....	33
BAB IV ANALISIS DAN PEMBAHASAN	37
4.1 Karakteristik Rumah Tangga di Pulau Madura.....	37
4.1.1 Karakteristik Variabel Penelitian Berskala Numerik.....	37
4.1.2 Karakteristik Variabel Penelitian Berskala Kategorik.....	40
4.2 Hasil Pengelompokan dengan Metode SWFM.....	49
4.2.1 Pengelompokan pada Data Numerik.....	50
4.2.2 Pengelompokan pada Data Kategorik.....	53
4.2.3 Pengelompokan pada Data Campuran	56
BAB V KESIMPULAN DAN SARAN	65
5.1 Kesimpulan.....	65
5.2 Saran	66
DAFTAR PUSTAKA	67
LAMPIRAN.....	71
BIODATA PENULIS.....	109

DAFTAR TABEL

	Halaman
Tabel 3. 1 Variabel Penelitian	25
Tabel 3. 2 Struktur Data Penelitian.....	33
Tabel 4. 1 Analisis Deskriptif Variabel Penelitian Berskala Numerik.....	38
Tabel 4. 2 Nilai <i>Pseudo-F</i> masing-masing jumlah kelompok	50
Tabel 4. 3 Karakteristik Data pada Hasil Pengelompokan Data Berskala Numerik.....	51
Tabel 4. 4 Hasil Kelompok Data Numerik pada Rumah Tangga yang Tinggal disetiap Kabupaten di Pulau Madura ..	53
Tabel 4. 5 Karakteristik Data pada Hasil Pengelompokan Data Berskala Kategorik.....	54
Tabel 4. 6 Hasil Kelompok Data Kategorik pada Rumah Tangga yang Tinggal disetiap Kabupaten di Pulau Madura ..	55
Tabel 4. 7 Nilai Rasio Hasil Pengelompokan dengan Metode SWFM	57
Tabel 4. 8 Nilai Rata-rata Data Numerik pada Hasil Pengelompokan dengan Metode SWFM.....	58
Tabel 4. 9 Karakteristik Data Kategorik Hasil Pengelompokan dengan Metode SWFM	58
Tabel 4. 10 Hasil Kelompok Data Campuran pada Rumah Tangga yang Tinggal disetiap Kabupaten di Pulau Madura.....	63
Tabel 4. 11 Hasil Uji <i>Kruskal Wallis</i> terhadap 5 Kelompok	64

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Skema Pengelompokan SWFM	15
Gambar 2. 2 Faktor-Faktor yang Mempengaruhi Kebutuhan Air Rumah Tangga	22
Gambar 3. 1 Diagram Alir Langkah Analisis	35
Gambar 4. 1 Karakteristik Data pada Variabel Numerik.....	39
Gambar 4. 2 Persentase Pendidikan Terakhir untuk Kepala Rumah Tangga	40
Gambar 4. 3 Persentase Sumber Air Minum	41
Gambar 4. 4 Persentase Sumber Air Minum dengan Kategori Baru	42
Gambar 4. 5 Persentase Sumber Air Masak	43
Gambar 4. 6 Persentase Sumber Air Masak dengan Kategori Baru	44
Gambar 4. 7 Persentase Sumber Air Mandi/Cuci	45
Gambar 4. 8 Persentase Sumber Air Mandi/Cuci dengan Kategori Baru	46
Gambar 4. 9 Persentase Sistem Saluran Sumber Air.....	47
Gambar 4. 10 Persentase Sistem Saluran Sumber Air dengan Kategori Baru	48
Gambar 4. 11 Persentase Rumah Tangga yang Membayar Air PAM	49
Gambar 4. 13 Karakteristik Hasil Pengelompokan pada Data Numerik.	52
Gambar 4. 14 Plot Nilai Rasio SWFM	57
Gambar 4. 15 Karakteristik Data Numerik pada Hasil Pengelompokan SWFM.....	62

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data Penelitian	71
Lampiran 2 <i>Syntax</i> Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 2 dengan <i>Ward</i>	72
Lampiran 3 <i>Syntax</i> Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 3 dengan <i>Ward</i>	74
Lampiran 4 <i>Syntax</i> Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 4 dengan <i>Ward</i>	76
Lampiran 5 <i>Syntax</i> Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 5 dengan <i>Ward</i>	79
Lampiran 6 <i>Syntax</i> Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 6 dengan <i>Ward</i>	82
Lampiran 7 <i>Syntax Pseudo-F</i> untuk Pengelompokan Data Berskala Numerik	85
Lampiran 8 <i>Syntax</i> Analisis Pengelompokan Data Kategorik dengan <i>K-modes</i>	93
Lampiran 9 <i>Syntax</i> Analisis Pengelompokan Data Campuran dengan Ensembel SWFM	94
Lampiran 10 <i>Syntax</i> Rasio Sw dan Sb untuk Ensembel SWFM	97
Lampiran 11 <i>Output</i> untuk Analisis Kelompok pada Data Numerik.....	100
Lampiran 12 <i>Output</i> untuk Analisis Kelompok pada Data Kategorik dengan Jumlah Kelompok 3	101
Lampiran 13 <i>Output</i> untuk Analisis Kelompok pada Data Numerik dan Data Kategorik dan Rasio Sw dan Sb dengan Ensembel SWFM.....	102
Lampiran 14 <i>Output Pseudo-F</i> masing-masing Kelompok	103
Lampiran 15 <i>Output</i> uji <i>Kruskal Wallis</i> dan Uji Normalitas ...	105
Lampiran 16 Surat Keterangan Pengambilan Data	108

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang untuk dilakukannya penelitian ini. Selanjutnya merumuskan masalah dan menetapkan tujuan penelitian. Kemudian pada bab ini juga dibahas mengenai manfaat dan batasan masalah pada penelitian yang dilakukan.

1.1 Latar Belakang

Sustainable 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. Air merupakan sumber daya terbarukan yang mempunyai peranan sangat penting dalam kehidupan makhluk hidup, terutama dalam kehidupan sehari-hari yang dimiliki oleh manusia. Hampir semua aktivitas manusia sangat membutuhkan air bersih dalam kehidupan misalnya dalam mandi, mencuci pakaian, mencuci piring, minum dan lain-lain. Penyediaan air bersih dapat memberi dampak terhadap peningkatan taraf dan mutu kehidupan masyarakat. Studi ketersediaan air bersih untuk suatu wilayah sangat diperlukan, khususnya sebagai perlindungan dan pelestarian lingkungan berkelanjutan.

Selain itu, studi akses penyediaan air bersih pada suatu penduduk atau rumah tangga tidak terlepas dari pembangunan ekonomi suatu wilayah. Pembangunan adalah kemampuan

pemerintah untuk memenuhi kebutuhan dasar (*basic needs*), antara lain yaitu pangan, papan, kesehatan, dan perlindungan (termasuk akses terhadap air bersih) (Todaro, 2000). 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. Salah satu hak dasar yang diakui adalah akses terhadap air bersih. Menurut Suparman (2007), 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.

Pulau Madura terdiri dari dataran rendah yang membentang di pesisir utara dan selatan dan memiliki 4 Kabupaten yaitu Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Pamekasan dan Kabupaten Sumenep. Dari 4 Kabupaten di Pulau Madura, 3 diantaranya termasuk kemiskinan tertinggi menurut Badan Pusat Statistik (BPS) Jawa Timur (2016), yaitu dengan jumlah penduduk miskin Kabupaten Sampang sebanyak 227.800 orang, Kabupaten Sumenep sebanyak 216.140 orang dan Kabupaten Bangkalan sebanyak 205.710 orang. 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, 4 dari 3 Kabupaten di Pulau Madura dikatakan kurang sehat. Dari penilaian kinerja PDAM, aspek pelayanan air bersih dan aspek operasional terutama pada konsumsi air domestik merupakan salah satu bobot tertinggi (BPPSPAM, 2017). 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 rumah tangga. Pengelompokan dapat dihasilkan dengan salah satu metode statistika yaitu analisis pengelompokan (*Cluster Analysis*). Metode ensemble adalah teknik pengelompokan untuk menggabungkan hasil pengelompokan dari beberapa algoritma pengelompokan untuk mendapatkan kelompok yang lebih baik (He, Xu, & Deng, 2005a). Pada penelitian ini belum diketahui jumlah kelompok yang akan dibentuk sehingga perlu digunakan metode pengelompokan hirarki. Metode hirarki digunakan untuk mengelompokkan data numerik dengan metode hirarki *agglomerative*. Sedangkan untuk data kategorik, dilakukan pengelompokan dengan menggunakan metode *K-modes*. Setelah dikelompokkan pada masing-masing skala data, dilanjutkan dengan menggunakan pengelompokan *Similarity Weight and Filter Method* (SWFM) sebagai *final cluster* dari hasil gabungan pengelompokan skala data yang berbeda.

Metode pengelompokan dengan data berskala campuran dapat digunakan dengan mentransformasi data kategorik menjadi numerik seperti penelitian Dewangan, Sharma, dan Akasapu (2010) kemudian mengelompokkan objek dengan metode pengelompokan numerik. Metode transformasi memiliki kelebihan yaitu mengurangi kompleksitas dalam komputasi. Sedangkan kekurangan dari metode transformasi harus menentukan transformasi yang tepat agar tidak kehilangan informasi dalam datanya. Selain menggunakan transformasi, dapat dilakukan metode ensemble. Metode ensemble teknik pengelompokan untuk menggabungkan hasil pengelompokan dari beberapa algoritma pengelompokan untuk mendapatkan kelompok yang lebih baik (He, Xu, & Deng, 2005a). Terdapat beberapa metode pengelompokan data numerik yang standar seperti non-hirarki dan

hirarki. Sedangkan metode kategorik yang biasa digunakan seperti *K-modes*, *ROCK* dan *squeezer*.

Sebelumnya pernah dilakukan oleh Alvionita (2017) yaitu dengan menggunakan metode ensemble *ROCK* dan *SWFM* untuk pengelompokan data campuran skala data numerik dan kategorik pada kasus aksesori jeruk. Hal ini juga pernah dilakukan oleh Ramdhany (2017) yaitu dengan menggunakan metode ensemble *ROCK* dan *Two Step Cluster* dengan hasil pengelompokan terbaik yaitu ensemble *ROCK*. Herliyasari (2017), juga pernah melakukan pengelompokan rumah tangga di Surabaya berdasarkan indikator kesejahteraan dengan menggunakan *ensemble clustering* dengan algoritma *squeezer*. Berdasarkan hasil uraian tersebut, maka akan dilakukan penelitian tentang pengelompokan faktor-faktor yang mempengaruhi kebutuhan air bersih rumah tangga di Pulau Madura menggunakan *Similarity Weight and Filter Method* (*SWFM*). Penelitian ini diharapkan bisa menjadi masukan Pemerintah dalam penanganan konsumsi air bersih pada rumah tangga di Pulau Madura.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada penelitian ini berdasarkan permasalahan yang telah diuraikan adalah sebagai berikut.

1. Bagaimana karakteristik data faktor-faktor yang mempengaruhi konsumsi air bersih rumah tangga di Pulau Madura?
2. Bagaimana pengelompokan menggunakan metode *Similarity Weight and Filter Method* (*SWFM*) untuk faktor-faktor yang mempengaruhi konsumsi air bersih rumah tangga di Pulau Madura?

1.3 Tujuan

Berdasarkan rumusan masalah tersebut, maka didapatkan beberapa tujuan dari penelitian sebagai berikut:

1. Mendeskripsikan data faktor-faktor yang mempengaruhi konsumsi air bersih rumah tangga di Pulau Madura.

2. Mengelompokkan data menggunakan metode *Similarity Weight and Filter Method* (SWFM) untuk faktor-faktor yang mempengaruhi konsumsi air bersih rumah tangga di Pulau Madura.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan beberapa manfaat bagi berbagai pihak, diantaranya sebagai berikut.

1. Memberikan wawasan keilmuan statistika mengenai penerapan metode pengelompokan menggunakan *Similarity Weight and Filter Method* (SWFM).
2. Memberikan informasi kepada pihak Pemerintah Kabupaten di Pulau Madura, yang dapat digunakan sebagai bahan pertimbangan dalam penanganan akses air bersih untuk rumah tangga.

1.5 Batasan Masalah

Batasan pada penelitian ini adalah data berdasarkan kuisioner SUSENAS tahun 2016 di Jawa Timur khususnya Kabupaten di Pulau Madura. Penelitian ini juga dibatasi dengan menggunakan total pengeluaran satu bulan terakhir dan tingkat pendidikan tertinggi terakhir untuk kepala rumah tangga.

(Halaman ini sengaja dikosongkan)

BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai teori-teori yang digunakan dalam melakukan analisis pengelompokan data numerik dengan dua skala data campuran yaitu data data numerik dan kategorik. Permasalahan pada penelitian ini yaitu pengelompokan rumah tangga di Pulau Madura dengan tingkat kebutuhan air bersih. Beberapa teori yang dibahas di bab ini adalah statistika deskriptif, analisis kelompok (*cluster analysis*), metode pengelompokan, pengelompokan data numerik, pengelompokan data kategorik, pengelompokan data campuran numerik dan kategorik, kinerja hasil pengelompokan, uji normalitas, uji *Kruskal Wallis* dan faktor yang mempengaruhi konsumsi air bersih rumah tangga.

2.1 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 (Walpole, 1995). Pada metode ini data dibedakan menjadi dua yaitu data kategorik dan numerik.

2.2 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 (Simamora, 2005). 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 (Johnson & Winchern, 2007). 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. Tahap-tahap dalam analisis kelompok adalah sebagai berikut.

- a. Menentukan ukuran kemiripan yang digunakan sebagai dasar pengelompokan.
- b. Menentukan metode pengelompokan yang ingin digunakan (hirarki atau non-hirarki)
- c. Menentukan teknik pengelompokan berdasarkan metode yang telah ditetapkan.
- d. Menganalisis pengelompokan dengan metode dan teknik yang telah ditentukan.
- e. Menginterpretasikan hasil pengelompokan.

Ukuran kemiripan dan ketidakmiripan merupakan hal yang sangat mendasar dalam kelompok analisis. Algoritma pengelompokan menggunakan ukuran kemiripan atau ketidakmiripan digunakan untuk menggabungkan atau memisahkan objek dari suatu data. Ukuran kemiripan biasanya digunakan dalam algoritma pengelompokan untuk menganalisis data kategori, sedangkan ukuran ketidakmiripan digunakan oleh algoritma pengelompokan untuk menganalisis data numerik. 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 (Sharma, 1996). 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.

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2} \quad (2.1)$$

Dimana

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

Dengan jumlah variabel sebanyak m . Semakin kecil jarak antara kedua objek, maka diindikasikan semakin besar kedekatan antara kedua objek tersebut.

Dengan dilakukannya analisis kelompok, maka diharapkan mendapatkan hasil pengelompokan yang baik. Hasil pengelompokan yang baik artinya dalam suatu kelompok mempunyai homogenitas yang tinggi antara anggota dalam suatu kelompok serta memiliki heterogenitas yang tinggi antara kelompok yang satu dengan kelompok lainnya (Hair, Black, Babin, & Anderson, 2009).

2.3 Metode Pengelompokan

Tahap pengelompokan dalam analisis kelompok dibedakan menurut jenis data yang dimiliki. Pada Umumnya analisis kelompok terfokus pada data numerik, namun sering ditemukan kasus dengan data kategorik bahkan kasus dengan data campuran data numerik dan kategorik. Oleh karena itu dalam pengelompokannya menggunakan metode yang berbeda.

2.3.1 Pengelompokan Data Numerik

Pengelompokan data numerik dilakukan berdasarkan ukuran ketidakmiripan atau jarak untuk data numerik. Hasil pengelompokan disajikan dalam bentuk dendogram (diagram pohon) yang memungkinkan penelusuran objek-objek yang diamati menjadi lebih mudah dan informatif. Teknik yang digunakan untuk pengelompokan meliputi metode hirarki dan metode non hirarki.

Pengelompokan hirarki dimulai dengan dua atau lebih objek yang mempunyai kesamaan paling dekat, kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua. Analisis dilakukan hingga kelompok membentuk hirarki (tingkatan) yang jelas antar objek dari yang paling mirip sampai paling tidak mirip. Metode hirarki digunakan bila banyaknya kelompok yang akan dibentuk tidak diketahui sebelumnya dan banyaknya amatan tidak terlalu besar (Johnson & Winchern, 2007). Terdapat dua teknik pengelompokan dalam analisis kelompok hirarki yaitu teknik pembagian (*divisive*) dan teknik penggabungan (*agglomerative*).

Dalam teknik pengelompokan hirarki *agglomerative* setiap objek merupakan satu kelompok tersendiri. Lalu dua kelompok terdekat digabungkan dan seterusnya sehingga diperoleh satu kelompok yang berunsurkan semua objek. Untuk menggabungkan dua kelompok pada tahap awal, dimana tiap kelompok hanya terdiri dari satu objek, diperlukan ukuran ketidakmiripan antar objek. Bila suatu kelompok merupakan penggabungan dari beberapa kelompok sebelumnya, maka diperlukan ukuran ketidakmiripan antar kelompok, kelompok-kelompok dengan ukuran ketidakmiripan terkecil digabungkan menjadi kelompok yang baru.

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 2.2.

$$ESS = \sum_{j=1}^N (x_j - \bar{x})' (x_j - \bar{x}) \quad (2.2)$$

Menentukan jumlah kelompok optimum merupakan tahapan penting setelah proses pengelompokan. Tahapan ini disebut sebagai validasi pengelompokan (Halkidi, Batistakis, & Vizirgiannis, 2001). Indeks R^2 merupakan salah satu indeks yang dapat digunakan untuk menentukan jumlah kelompok optimum

pada pengelompokan hirarki (Sharma, 1996). 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

Sum of Square Total (SST)

$$SST = \sum_{l=1}^{m_{\text{numerik}}} \sum_{i=1}^n (x_{il} - \bar{x}_l)^2 \quad (2.3)$$

Sum of Square Within Group (SSW)

$$SSW = \sum_{c=1}^C \sum_{l=1}^{m_{\text{numerik}}} \sum_{i=1}^{n_c} (x_{ilc} - \bar{x}_{lc})^2 \quad (2.4)$$

Sum of Square Between Group (SSB)

$$SSB = SST - SSW \quad (2.5)$$

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
 $l = 1, 2, \dots, m_{\text{numerik}}$,

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

R^2 kelompok baru adalah rasio dari SSB dan SST , artinya R^2 dapat didefinisikan sebagai ukuran perbedaan antar kelompok, dengan nilai berkisaran dari 0 sampai 1. Nilai $R^2 = 0$ menunjukkan bahwa tidak terdapat perbedaan antara kelompok, sedangkan $R^2 = 1$ menunjukkan adanya perbedaan yang signifikan antar kelompok yang terbentuk. Nilai R^2 dirumuskan pada persamaan 2.6.

$$R^2 = \frac{SSB}{SST} = \frac{[SST - SSW]}{SST} \quad (2.6)$$

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

$$Pseudo - F = \frac{\left(\frac{R^2}{c-1} \right)}{\left(\frac{1-R^2}{n-c} \right)} \quad (2.7)$$

2.3.2 Pengelompokan Data Kategorik

Pengelompokan data bersifat kategorik dapat dilakukan dengan ukuran kemiripan data dengan menggunakan metode hirarki dan non-hirarki. *K-modes* merupakan kluster algoritma dengan data bersifat kategorik. Berdasarkan Kaufman dan Rousseeuw (1990), andaikan X dan Y adalah dua data dengan fitur bertipe kategorikal. 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 2.8.

$$d(X, Y) = \sum_{j=1}^r \delta(x_j, y_j) \quad (2.8)$$

Dimana r adalah jumlah variabel, sedangkan (x_j, y_j) merupakan nilai pencocokan sebagaimana pada persamaan 2.9.

$$\delta(x_j, y_j) = \begin{cases} 0 & (x_j = y_j) \\ 1 & (x_j \neq y_j) \end{cases} \quad (2.9)$$

Misalkan X merupakan kelompok data dengan variabel bersifat kategorik $A_1, A_2, A_3, \dots, A_n$ maka modus dari $X = (X_1, X_2, X_3, \dots, X_n)$ adalah data $Q = (q_1, q_2, q_3, \dots, q_n)$ dengan nilai minimal sebagaimana pada 2.10.

$$D(X, Y) = \sum_{i=1}^n d(x_i, Q) \quad (2.10)$$

Vektor Q merupakan vektor diluar bagian dari X . Misal $n_{c_{ki}}$ adalah jumlah objek yang dimiliki oleh kategori c_{kj} ke- k pada atribut A_j dan $f(A_j = c_{kj} | X) = \frac{n_{c_{kj}}}{n}$ adalah frekuensi relatif kategori $n_{c_{ki}}$ dalam X sehingga fungsi $D(X, Q)$ akan minimal jika $f(A_j = q_j | X) \geq f(A_j = c_{kj} | X)$ untuk $q_j \neq c_{kj}$ untuk semua $j = 1, 2, 3, \dots, r$. Fungsi objektif yang digunakan dalam K -modes seperti pada persamaan 2.11.

$$J = \sum_{l=1}^k \sum_{i=1}^n \sum_{j=1}^r w_{i,l} (x_{i,j}, q_{l,j}) \quad (2.11)$$

$(x_{i,j}, q_{l,j})$ merupakan nilai pencocokan antara vektor dengan modus kelompok yang diikuti, sedangkan $w_{i,l}$ merupakan nilai keanggotaan data setiap kelompok. $w_{i,l}$ memiliki nilai antara 0 sampai 1 yang diperoleh dari persamaan 2.12.

$$w_{i,l} = \begin{cases} 1 & \text{jika } d(X_i, Q_l) < d(X_i, Q_t) \\ 0 & \text{untuk } t \neq l \end{cases} \quad (2.12)$$

Dimana $l = 1, 2, 3, \dots, k$ dan $t = 1, 2, 3, \dots, k$. k adalah jumlah kelompok sedangkan n adalah jumlah data dalam kelompok. Metode K -modes dapat diterapkan pada data kategorik

menghasilkan kelompok yang lebih rinci, waktu pembentukan kelompok lebih singkat, dan unggul dalam pengelompokan pada data berdimensi besar.

2.4 Pengelompokan Data Campuran Numerik dan Kategorik

Analisis pengelompokan terhadap data campuran diawali dengan pembagian data menjadi dua subdata, yaitu data murni numerik dan data murni kategorik. Misalkan terdapat data dengan variabel berskala campuran sebanyak m , dengan $m_{numerik}$ merupakan jumlah variabel yang murni berskala numerik dan $m_{kategorik}$ merupakan jumlah variabel murni yang berskala kategorik, sehingga diperoleh $m = m_{numerik} + m_{kategorik}$. Selanjutnya dilakukan pengelompokan data sesuai dengan jenis skala data secara terpisah. Hasil pengelompokan tersebut kemudian digabungkan menggunakan metode pengelompokan *Similarity Weight and Filter Method* (SWFM) sehingga diperoleh kelompok akhir (*final cluster*).

Pengelompokan ensemble merupakan metode yang menggabungkan beberapa algoritma yang berbeda untuk mendapatkan partisi umum dari data yang bertujuan untuk menyatukan hasil pengelompokan individu (Suguna & Selvi, 2012). Pengelompokan ensemble terdiri dari dua tahap algoritma. Tahap pertama adalah melakukan pengelompokan dengan beberapa algoritma dan menyimpan hasil pengelompokan tersebut. Tahap kedua menggunakan fungsi konsensus untuk menentukan *final cluster* dari kelompok-kelompok hasil tahap pertama. Pengelompokan *ensemble* dengan data campuran dapat menggunakan algoritma CEBMDC (*Cluster Ensemble Based Mixed Data Clustering*) ditunjukkan pada Gambar 2.1.

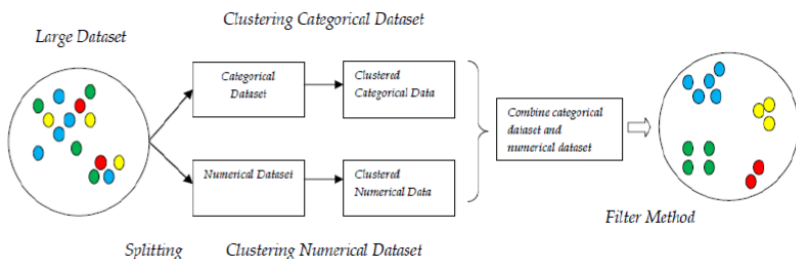
Skema pengelompokan ensemble untuk data campuran dapat menggunakan algoritma CEBMDC (*Cluster Ensemble Based Mixed Data Clustering*). Langkah-langkah dalam analisis

data campuran menggunakan metode pengelompokan ensemble yang disebut algoritma CEBMDC memiliki tahapan sebagai berikut (He, Xu, & Deng, 2005b).

- Membagi data menjadi dan subdata, yaitu murni numerik dan murni kategorik.
- Melakukan pengelompokan objek yang memiliki variabel numerik dengan algoritma pengelompokan data numerik, serta melakukan pengelompokan objek yang memiliki variabel kategorik dengan algoritma pengelompokan data kategorik.
- Menggabungkan (*combining*) hasil pengelompokan dari variabel numerik dan kategorik, yang disebut proses ensemble.
- Melakukan pengelompokan ensemble menggunakan algoritma pengelompokan data kategorik untuk mendapatkan kelompok akhir (*final cluster*).

2.5 Pengelompokan *Similarity Weight and Filter Method* (SWFM)

Pengelompokan SWFM memiliki konsep analisis yang sama dengan pengelompokan lainnya, yang membedakan adalah tahap pembentukan klaster akhir. Berikut tahapan pengelompokan SWFM pada Gambar 2.2.



Sumber : Reddy & Kavitha (2012)

Gambar 2. 1 Skema Pengelompokan SWFM

Pada 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 2.13 (Reddy & Kavitha, 2012).

$$sim(x_i, x_j) = \sum_{i \leq n_i, j \leq n_j} \frac{S_{ij}}{\max(n_i, n_j)}, i \neq j \quad (2.13)$$

Dengan

$$i, j = 1, 2, 3, \dots, n$$

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

X_i : Himpunan pengamatan ke- i dengan

$$X_i = \{x_{1i}, x_{2i}, x_{3i}, \dots, x_{ki}\}$$

X_j : Himpunan pengamatan ke- j dengan

$$X_j = \{x_{1j}, x_{2j}, x_{3j}, \dots, x_{kj}\}$$

$|X|$: 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

Untuk mengelompokkan data campuran menggunakan algoritma metode *filter* berdasarkan *similarity weight*. Pertama data dibagi menjadi dua-sub data yaitu data kategorik murni dan data numerik murni. Pada masing-masing data dilakukan pengelompokan berdasarkan jenis data untuk mendapatkan

kelompok yang sesuai. Selanjutnya hasil pengelompokan data numerik dan kategorik digabungkan untuk memperoleh pengelompokan akhir menggunakan metode *filter* sebagaimana pada persamaan 2.14 (Reddy & Kavitha, 2012).

$$F(X_i, X_j) = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w_{ij} d(X_i, X_j) \quad (2.14)$$

Dimana $d(X_i, X_j) = 1 - \text{sim}(X_i, X_j)$ dan w_{ij} merupakan bobot antara kelompok ke- i dan j yang ada pada umumnya bernilai 0.5 dan $0 < \text{sim}(X_i, X_j) \leq 1$.

Keuntungan ketika menggunakan metode SWFM antar lain penggunaan yang efisien dari *cut* dan siklus property. Bentuk klaster berdampak sangat kecil pada kinerja algoritma *clustering*, efektif untuk dimensi lebih dari 5 serta mengurangi waktu, pencarian jarak yang digunakan lebih efisien bahkan jika batas-batas kelompok tidak teratur (Reddy & Kavitha, 2012).

2.6 Kinerja Hasil Pengelompokan

Pengukuran kinerja hasil pengelompokan merupakan langkah untuk mengetahui validalitas suatu pengelompokan. Kinerja hasil pengelompokan untuk variabel dengan skala numerik dapat diketahui dari rasio nilai S_W dan S_B (Bunkers & James, 1996). Dengan menggunakan nilai rata-rata variabel, simpangan baku di dalam kelompok (S_W) dan simpangan baku antar kelompok (S_B) dapat dirumuskan dalam persamaan 2.15 dan 2.16

$$S_W = \frac{1}{C} \sum_{c=1}^C S_c \quad (2.15)$$

dengan S_c merupakan simpangan baku kelompok ke- c dan C adalah jumlah kelompok yang terbentuk.

$$S_B = \left[\frac{1}{C-1} \sum_{c=1}^C (\bar{x}_c - \bar{x})^2 \right]^{1/2} \quad (2.16)$$

dimana \bar{x}_c merupakan rata-rata kelompok ke- c dan rata-rata keseluruhan kelompok. Kinerja suatu metode pengelompokan akan semakin baik jika semakin kecil nilai rasio antara S_W dan S_B . Hal ini berarti terdapat homogenitas maksimum dalam kelompok dan heterogenitas maksimum antara kelompok (Bunkers & James, 1996).

Adapun untuk mengukur kinerja pengelompokan untuk data yang kategorik adalah dengan menggunakan tabel kontingensi yang ekuivalen dengan melakukan ANOVA (*Analysis of Variance*).

Jika terdapat sebanyak n pengamatan dengan n_k merupakan jumlah pengamatan dengan kategori ke- k dimana $k = 1, 2, 3, \dots, K$

dan $\sum_{k=1}^K n_k = n$. Selanjutnya n_{kc} merupakan jumlah pengamatan

dengan kategori ke- k dan kelompok ke- c dimana $c = 1, 2, 3, \dots, C$ dengan C adalah jumlah kelompok yang terbentuk, sehingga

$n_{.c} = \sum_{k=1}^K n_{kc}$ merupakan jumlah pengamatan pada kelompok ke- c

dan $n_{.k} = \sum_{c=1}^C n_{kc}$ merupakan jumlah pengamatan pada kategori ke-

k . Total jumlah pengamatan dapat dituliskan menjadi

$$n = \sum_{c=1}^C n_{.c} = \sum_{k=1}^K n_{.k} = \sum_{k=1}^K \sum_{c=1}^C n_{kc} .$$

Jumlah kuadrat total (*SST*) untuk variabel dengan data kategorik dapat dirumuskan seperti persamaan 2.17. Untuk total jumlah kuadrat dalam kelompok (*SSW*) dirumuskan dalam

persamaan 2.18. Serta untuk jumlah kuadrat antar kelompok (SSB) dapat dirumuskan seperti pada persamaan 2.19 (Dewi, 2012).

$$SST = \frac{n}{2} - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (2.17)$$

$$SSW = \sum_{c=1}^C \left(\frac{n_{.c}}{2} - \frac{1}{2n} \sum_{k=1}^K n_k^2 \right) = \frac{n}{2} - \frac{1}{2n} \sum_{c=1}^C \frac{1}{n_{.c}} \sum_{k=1}^K n_k^2 \quad (2.18)$$

$$SSB = \frac{1}{2} \left(\sum_{c=1}^C \frac{1}{n_{.c}} \sum_{k=1}^K n_{kc}^2 \right) - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (2.19)$$

Mean of Squares Total (MST), *Mean of Squares Within (MSW)* dan *Mean of Squares Between (MSB)* dapat dirumuskan seperti pada persamaan 2.22, 2.23, dan 2.24.

$$MST = \frac{SST}{(n-1)} \quad (2.22)$$

$$MSW = \frac{SSW}{(n-C)} \quad (2.23)$$

$$MSB = \frac{SSB}{C-1} \quad (2.24)$$

Simpangan baku dalam kelompok (S_W) dan simpangan baku antar kelompok (S_B) untuk data kategorik dapat dirumuskan pada persamaan 2.25 dan 2.26.

$$S_W = [MSW]^{1/2} \quad (2.25)$$

$$S_B = [MSB]^{1/2} \quad (2.26)$$

Selanjutnya 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.

2.7 Uji Normalitas

Uji normalitas merupakan sebuah uji yang dilakukan dengan tujuan untuk menilai sebaran data pada sebuah kelompok data atau variabel, apakah sebaran data tersebut berdistribusi normal atau tidak. Uji normalitas yang digunakan dalam penelitian ini menggunakan *Kolmogorov Smirnov* dan *mshapiro test*.

Pemeriksaan *Kolmogorov Smirnov* digunakan untuk memeriksa persebaran dari suatu variabel. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : Data berasal dari distribusi normal

H_1 : Data tidak berasal dari distribusi normal

Statistik uji:

$$D = \text{Sup}_x |F_n(x) - F_0(x)| \quad (2.27)$$

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 $D > D_\alpha$ dilihat dari nilai *p-value*. Jika *p-value* $> \alpha$ (0,05) maka data sampel berdistribusi normal.

Metode *mshapiro test* berfungsi untuk menguji apakah data berdistribusi normal multivariat. Hipotesis, statistik uji, dan keputusan yang digunakan sebagai berikut.

H_0 : Data berdistribusi normal multivariat.

H_1 : Data tidak berdistribusi normal multivariat.

Statistik uji untuk pengujian asumsi berdistribusi normal multivariat ditunjukkan dalam persamaan 2.28.

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)} \right)}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.28)$$

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 α maka H_0 ditolak jika W kurang dari a .

2.8 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 untuk pengujian *Kruskal-Wallis* adalah sebagai berikut.

H_0 : Kedua populasi identik

H_1 : Tidak semua populasi identik

Statistik uji yang digunakan sesuai dengan persamaan 2.29.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^g \frac{R_i^2}{n_i} - 3(N+1) \quad (2.29)$$

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

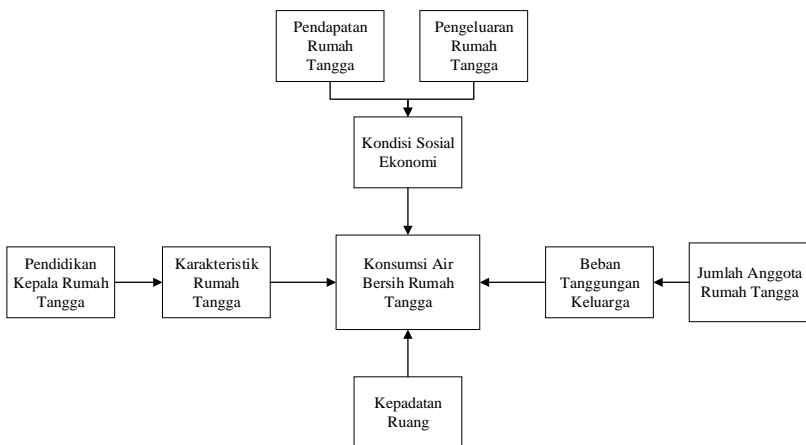
K : Banyaknya kelompok

Pengambilan keputusan menggunakan nilai H yang dibandingkan dengan nilai $\chi_{(0,05;K-1)}^2$. Keputusan tolak H_0 apabila nilai H lebih besar dari $\chi_{(0,05;K-1)}^2$ atau nilai χ_{hitung}^2 lebih besar daripada $\chi_{(0,05;K-1)}^2$ (Daniel, 1989).

2.9 Faktor-Faktor yang Mempengaruhi Konsumsi Air Bersih

Air bersih adalah air yang dipakai sehari-hari untuk keperluan mencuci, mandi, memasak dan dapat diminum setelah dimasak (Kodoatie, 2008). Air merupakan zat mutlak bagi setiap makhluk hidup dan kebersihan air adalah syarat utama bagi terjaminnya kesehatan (Dwijosaputro, 1981).

Menurut Peraturan Menteri Kesehatan RI Nomor 416/Menkes/Per/IX/1990 tentang syarat-syarat pengawasan kualitas air, air bersih adalah air yang digunakan untuk keperluan sehari-hari yang kualitasnya memenuhi syarat-syarat kesehatan dan dapat diminum apabila telah dimasak. Terdapat berbagai macam faktor-faktor yang dapat mempengaruhi konsumsi air bersih. Salah satunya adalah berdasarkan teori Keynes dalam Sukirno (2004), bahwa konsumsi masyarakat semakin dipengaruhi oleh besarnya pendapatan. Berdasarkan uraian tersebut, maka faktor-faktor yang diduga dapat mempengaruhi konsumsi air bersih yang dapat digambarkan pada Gambar 2.2 sebagai berikut.



Gambar 2. 2 Faktor-Faktor yang Mempengaruhi Konsumsi Air Rumah Tangga

Berdasarkan hasil penjelasan diatas, menurut Winarna (2003), 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 23ariff 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. Kebutuhan air bersih antara laki-laki dengan perempuan tidaklah berbeda. Namun menurut Kementerian Perempuan (2010) tidak demikian. Seringkali perempuan banyak berhubungan dengan kebutuhan air baik untuk kepentingan individu maupun rumah tangga.

(Halaman ini sengaja dikosongkan)

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai sumber data yang digunakan, variabel penelitian, dan langkah analisis yang dilakukan dalam penelitian. Langkah analisis dalam penelitian ini kemudian disajikan dalam bentuk diagram alir.

3.1 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 yaitu Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Sumenep dan Kabupaten Pamekasan dengan jumlah rumah tangga yang digunakan sebanyak 3.097 rumah tangga.

3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini ditunjukkan oleh Tabel 3.1. Terdapat 9 variabel diantaranya adalah tiga data berskala numerik dan enam data berskala kategorik.

Tabel 3. 1 Variabel Penelitian

	Variabel	Keterangan
A. Data Numerik	X_1	Total Pengeluaran sebulan
	X_2	Proporsi Anggota Rumah Tangga Perempuan
	X_3	Jumlah Anggota Rumah Tangga
B. Data Kategorik	X_4	Tingkat Pendidikan KRT
	X_5	Sumber Air Minum
	X_6	Sumber Air Masak
	X_7	Sumber Air Mandi/Cuci
	X_8	Sistem saluran sumber air
	X_9	Pembayaran Air PAM

Definisi operasional dari variabel penelitian yang digunakan sebagai berikut:

X_1 = Total pengeluaran sebulan adalah jumlah pengeluaran yang dikeluarkan oleh rumah tangga untuk konsumsi kebutuhan lainnya.

X_2 = Proporsi anggota rumah tangga perempuan adalah banyaknya anggota rumah tangga perempuan yang tinggal dalam satuan persen.

X_3 = Jumlah anggota rumah tangga adalah jumlah keseluruhan anggota rumah tangga.

X_4 = Tingkat pendidikan kepala rumah tangga merupakan pendidikan tertinggi yang ditamatkan oleh kepala rumah tangga.

Kategori:

- 1) Kode 1 (satu) merupakan kepala rumah tangga dengan tingkat pendidikan tidak bersekolah.
- 2) Kode 2 (dua) merupakan kepala rumah tangga dengan tingkat pendidikan rendah yaitu pendidikan SD/setara dan SMP/setara.
- 3) Kode 3 (tiga) merupakan kepala rumah tangga dengan tingkat pendidikan menengah yaitu pendidikan SMA/setara.
- 4) Kode 4 (empat) merupakan kepala rumah tangga dengan tingkat pendidikan tinggi yaitu perguruan tinggi (PT).

X_5 = Sumber air minum utama merupakan sumber air minum yang digunakan oleh rumah tangga di Pulau Madura.

Kategori dari SUSENAS:

- 1) Kode 1 (satu) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu air kemasan bermerk. Air kemasan bermerk adalah air yang diproduksi dan didistribusikan oleh suatu perusahaan dalam bentuk botol dan kemasan gelas.
- 2) Kode 2 (dua) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu air isi ulang. Air isi

ulang adalah air yang diproduksi melalui proses penjernihan dan tidak memiliki merk.

- 3) Kode 3 (tiga) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu ledeng meteran. Ledeng meteran adalah air yang diproduksi melalui proses penjernihan dan penyehatan sebelum dialirkan kepada konsumen melalui suatu instansi berupa saluran air.
- 4) Kode 4 (empat) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu ledeng eceran. Ledeng eceran adalah air yang diproduksi melalui proses penjernihan dan penyehatan (air PAM) namun disalurkan ke konsumen melalui pedagang air keliling/pikulan.
- 5) Kode 5 (lima) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu air sumur bor/pompa yang merupakan air tanah yang cara pengambilannya dengan pompa tangan, pompa listrik atau kincir angin termasuk sumur artesis.
- 6) Kode 6 (enam) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu sumur terlindung yang merupakan sumur galian bila lingkaran sumur/perigi tersebut dilindungi oleh tembok paling sedikit 0,8 meter di atas tanah dan 3 meter ke bawah tanah, serta ada lantai semen sejauh 1 meter dari lingkaran sumur.
- 7) Kode 7 (tujuh) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu sumur tak terlindung yang merupakan sumur yang tidak memenuhi syarat sebagai sumur terlindung.
- 8) Kode 8 (delapan) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu mata air terlindung yang merupakan sumber air permukaan tanah dimana air timbul dengan sendirinya. Dikategorikan sebagai terlindung bila mata air tersebut terlindung dari air bekas pakai.

- 9) Kode 9 (sembilan) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu mata air tak terlindung yang merupakan sumber air permukaan tanah dimana air timbul dengan sendirinya. Dikategorikan sebagai tak terlindung bila mata air tersebut tidak terlindung atau tercemar.
 - 10) Kode 10 (sepuluh) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu air permukaan. Air permukaan adalah apabila rumah tangga menggunakan air dari sungai, danau, waduk, kolam, irigasi sebagai sumber utama air bersih.
 - 11) Kode 11 (sebelas) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu air hujan. Air hujan adalah apabila rumah tangga menggunakan air hujan sebagai sumber air bersih.
 - 12) Kode 12 (dua belas) merupakan rumah tangga dengan sumber air minum yang digunakan yaitu lainnya adalah sumber air selain yang tersebut di atas, seperti air laut yang disuling.
- X_6 = Sumber air masak merupakan sumber air masak yang digunakan oleh rumah tangga di Pulau Madura.
- Kategori dari SUSENAS:
- 1) Kode 1 (satu) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu air kemasan bermerk. Air kemasan bermerk adalah air yang diproduksi dan didistribusikan oleh suatu perusahaan dalam bentuk botol dan kemasan gelas.
 - 2) Kode 2 (dua) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu air isi ulang. Air isi ulang adalah air yang diproduksi melalui proses penjernihan dan tidak memiliki merk.
 - 3) Kode 3 (tiga) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu ledeng meteran. Ledeng meteran adalah air yang diproduksi melalui proses

penjernihan dan penyehatan sebelum dialirkan kepada konsumen melalui suatu instansi berupa saluran air.

- 4) Kode 4 (empat) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu ledeng eceran. Ledeng eceran adalah air yang diproduksi melalui proses penjernihan dan penyehatan (air PAM) namun disalurkan ke konsumen melalui pedagang air keliling/pikulan.
- 5) Kode 5 (lima) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu air sumur bor/pompa yang merupakan air tanah yang cara pengambilannya dengan pompa tangan, pompa listrik atau kincir angin termasuk sumur artesis.
- 6) Kode 6 (enam) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu sumur terlindung yang merupakan sumur galian bila lingkaran sumur/perigi tersebut dilindungi oleh tembok paling sedikit 0,8 meter di atas tanah dan 3 meter ke bawah tanah, serta ada lantai semen sejauh 1 meter dari lingkaran sumur.
- 7) Kode 7 (tujuh) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu sumur tak terlindung yang merupakan sumur yang tidak memenuhi syarat sebagai sumur terlindung.
- 8) Kode 8 (delapan) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu mata air terlindung yang merupakan sumber air permukaan tanah dimana air timbul dengan sendirinya. Dikategorikan sebagai terlindung bila mata air tersebut terlindung dari air bekas pakai.
- 9) Kode 9 (sembilan) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu mata air tak terlindung yang merupakan sumber air permukaan tanah dimana air timbul dengan sendirinya. Dikategorikan sebagai tak terlindung bila mata air tersebut tidak terlindung atau tercemar.

- 10) Kode 10 (sepuluh) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu air permukaan. Air permukaan adalah apabila rumah tangga menggunakan air dari sungai, danau, waduk, kolam, irigasi sebagai sumber utama air bersih.
- 11) Kode 11 (sebelas) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu air hujan. Air hujan adalah apabila rumah tangga menggunakan air hujan sebagai sumber air bersih.
- 12) Kode 12 (dua belas) merupakan rumah tangga dengan sumber air masak yang digunakan yaitu lainnya adalah sumber air selain yang tersebut di atas, seperti air laut yang disuling.

$X_7 =$ Sumber air mandi/cuci merupakan sumber air mandi/cuci yang digunakan oleh rumah tangga di Pulau Madura.

Kategori dari SUSENAS:

- 1) Kode 1 (satu) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu air kemasan bermerk. Air kemasan bermerk adalah air yang diproduksi dan didistribusikan oleh suatu perusahaan dalam bentuk botol dan kemasan gelas.
- 2) Kode 2 (dua) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu air isi ulang. Air isi ulang adalah air yang diproduksi melalui proses penjernihan dan tidak memiliki merk.
- 3) Kode 3 (tiga) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu ledeng meteran. Ledeng meteran adalah air yang diproduksi melalui proses penjernihan dan penyehatan sebelum dialirkan kepada konsumen melalui suatu instansi berupa saluran air.
- 4) Kode 4 (empat) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu ledeng eceran. Ledeng eceran adalah air yang diproduksi melalui proses penjernihan dan penyehatan (air PAM)

namun disalurkan ke konsumen melalui pedagang air keliling/pikulan.

- 5) Kode 5 (lima) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu air sumur bor/pompa yang merupakan air tanah yang cara pengambilannya dengan pompa tangan, pompa listrik atau kincir angin termasuk sumur artesis.
- 6) Kode 6 (enam) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu sumur terlindung yang merupakan sumur galian bila lingkaran sumur/perigi tersebut dilindungi oleh tembok paling sedikit 0,8 meter di atas tanah dan 3 meter ke bawah tanah, serta ada lantai semen sejauh 1 meter dari lingkaran sumur.
- 7) Kode 7 (tujuh) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu sumur tak terlindung yang merupakan sumur yang tidak memenuhi syarat sebagai sumur terlindung.
- 8) Kode 8 (delapan) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu mata air terlindung yang merupakan sumber air permukaan tanah dimana air timbul dengan sendirinya. Dikategorikan sebagai terlindung bila mata air tersebut terlindung dari air bekas pakai.
- 9) Kode 9 (sembilan) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu mata air tak terlindung yang merupakan sumber air permukaan tanah dimana air timbul dengan sendirinya. Dikategorikan sebagai tak terlindung bila mata air tersebut tidak terlindung atau tercemar.
- 10) Kode 10 (sepuluh) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu air permukaan. Air permukaan adalah apabila rumah tangga menggunakan air dari sungai, danau, waduk, kolam, irigasi sebagai sumber utama air bersih.

- 11) Kode 11 (sebelas) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu air hujan. Air hujan adalah apabila rumah tangga menggunakan air hujan sebagai sumber air bersih.
- 12) Kode 12 (dua belas) merupakan rumah tangga dengan sumber air mandi/cuci yang digunakan yaitu lainnya adalah sumber air selain yang tersebut di atas, seperti air laut yang disuling.

$X_8 =$ Sistem saluran air pada sumber air (minum, masak, dan mandi/cuci) merupakan penggunaan saluran air yang digunakan oleh rumah tangga di Pulau Madura.

Kategori:

- 1) Kode 1 (satu) merupakan rumah tangga dengan penggunaan saluran air perpipaan. Perpipaan adalah bila air yang digunakan menggunakan pipa dari sumber air sampai rumah.
- 2) Kode 2 (dua) merupakan rumah tangga dengan penggunaan saluran air hidran. Hidran umum/terminal air adalah sarana penyediaan air bersih yang sumbernya berasal dari air permukaan yang dialirkan melalui perpipaan ke tempat atau distribusi yang bersifat komunal.
- 3) Kode 5 (lima) merupakan rumah tangga yang tidak menggunakan saluran air perpipaan maupun hidran umum/terminal air.
- 4) Kode 8 (delapan) merupakan rumah tangga yang tidak tahu menggunakan saluran air perpipaan maupun hidran umum/terminal air.

$X_9 =$ Pembayaran air PAM merupakan rumah tangga yang membayar air PAM.

Kategori:

- 1) Kode 1 (satu) merupakan rumah tangga yang tidak menggunakan dan tidak membayar air PAM.

- 2) Kode 2 (dua) merupakan rumah tangga yang menggunakan dan membayar air PAM.

Tabel 3.2 menunjukkan struktur data yang digunakan dalam penelitian ini.

Tabel 3. 2 Struktur Data Penelitian

Nomor urut Rumah Tangga	$X_{1,i}$	$X_{2,i}$	$X_{3,i}$	$X_{4,i}$...	$X_{9,i}$
1	$x_{1,1}$	$x_{2,1}$	$x_{3,1}$	$x_{4,1}$...	$x_{9,1}$
2	$x_{1,2}$	$x_{2,2}$	$x_{3,2}$	$x_{4,2}$...	$x_{9,2}$
3	$x_{1,3}$	$x_{2,3}$	$x_{3,3}$	$x_{4,3}$...	$x_{9,3}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
3.097	$x_{1,3.097}$	$x_{2,3.097}$	$x_{3,3.097}$	$x_{4,3.097}$...	$x_{9,3.097}$

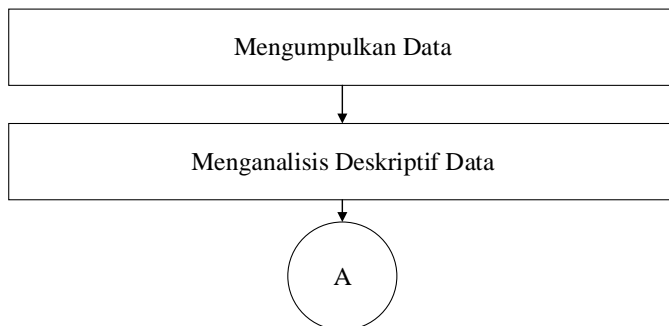
3.3 Langkah Analisis

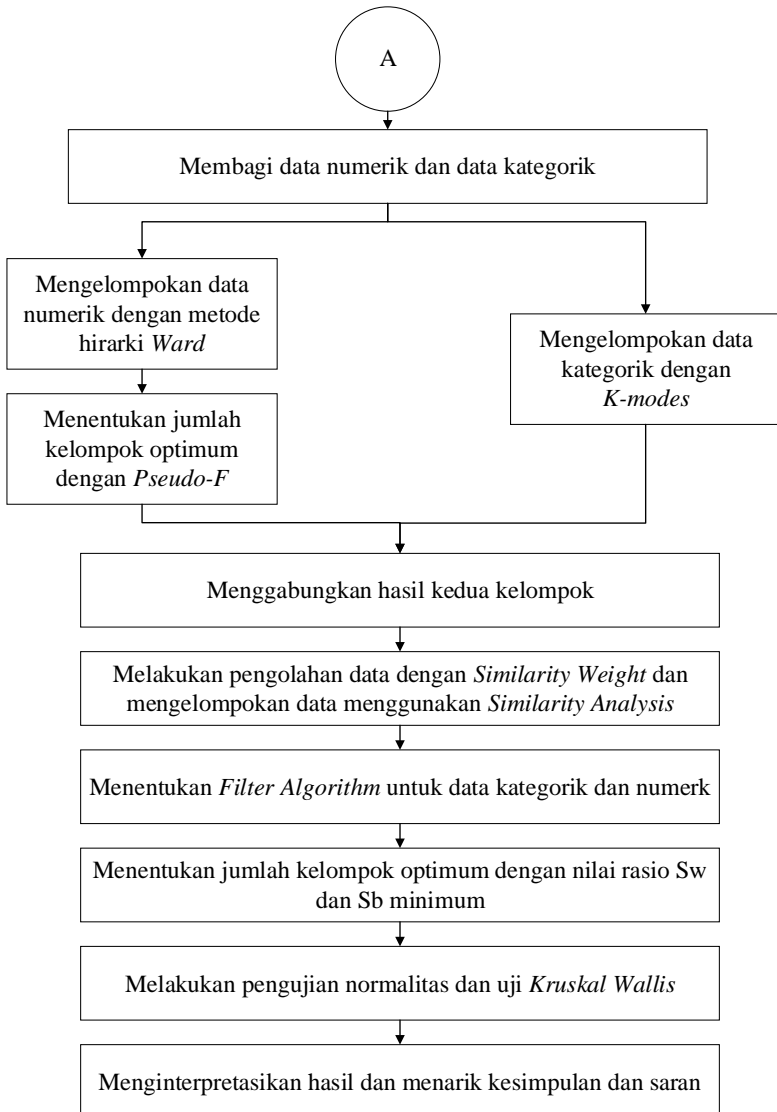
Berdasarkan hasil penentuan variabel yang digunakan dalam penelitian. Langkah-langkah yang dilakukan untuk mengelompokan data kategorik dan numerik pada Kabupaten/Kota di Jawa Timur menggunakan *Similarity Weight and Filter Method* (SWFM) adalah sebagai berikut.

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 histogram dan *pie chart*.
2. Mengelompokan rumah tangga di Pulau Madura menggunakan metode *Similarity Weight and Filter Method* (SWFM).
 - a. Membagi variabel penelitian menjadi data kategorik dan data numerik.
 - b. Melakukan klaster pada data numerik dengan metode hirarki *Ward*. Jarak yang digunakan adalah jarak *Euclidean*.

- c. Menentukan jumlah kelompok optimum, menghitung dan memilih kinerja pengelompokan berdasarkan nilai *Pseudo-F* pada masing-masing hasil pengelompokan numerik.
 - d. Mengelompokkan variabel yang berskala kategorik dengan menggunakan metode *K-modes*.
 - e. Mengkombinasikan hasil klaster 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 S_w dan S_b terkecil.
3. Melakukan pengujian dengan uji normal multivariat dan uji *Kruskal Wallis*.
 4. Menginterpretasikan hasil dan menarik kesimpulan dan saran.

Langkah analisis secara umum digambarkan dalam diagram alir pada Gambar 3.1.





Gambar 3. 1 Diagram Alir Langkah Analisis

(Halaman ini sengaja dikosongkan)

BAB IV

ANALISIS DAN PEMBAHASAN

Bab ini membahas mengenai karakteristik dan pengelompokan rumah tangga di Pulau Madura berdasarkan faktor-faktor yang mempengaruhi konsumsi air bersih. Pengelompokan dilakukan dengan pengelompokan data numerik, pengelompokan data kategorik dan terakhir dengan pengelompokan gabungan menggunakan SWFM. Sebelum dilakukan karakteristik data dan pengelompokan, variabel penelitian berskala numerik dengan banyaknya hasil *outlier* yang mendekati 10% diatasi dengan imputasi *mean*. Setelah diatasi *outlier*, data berskala numerik dapat diolah dengan statistika deskriptif dan pengelompokan data.

4.1 Karakteristik Rumah Tangga di Pulau Madura

Sebelum dilakukan analisis lebih lanjut, dilakukan eksplorasi data variabel yang akan diteliti menjadi hal penting untuk dilakukan terlebih dahulu. Tujuan dilakukan eksplorasi data adalah untuk mengetahui karakteristik rumah tangga di Pulau Madura faktor-faktor yang mempengaruhi konsumsi air bersih. Pada penelitian ini digunakan data numerik yang akan dilakukan analisis deskriptif dalam ukuran pemusatan data dan *boxplot* sedangkan data kategorik dilakukan melihat persentase dengan bentuk grafik histogram dan *pie-chart*.

4.1.1 Karakteristik Variabel Penelitian Berskala Numerik

Pada penelitian ini digunakan variabel penelitian berskala numerik berdasarkan faktor-faktor yang mempengaruhi konsumsi air bersih rumah tangga dengan tiga variabel penelitian berskala numerik yang digunakan antara lain adalah total pengeluaran rumah tangga sebulan (X_1), proporsi anggota rumah tangga perempuan (X_2), dan jumlah anggota rumah tangga (X_3). Hasil analisis statistika deskriptif dapat ditampilkan pada Tabel 4.1.

Tabel 4. 1 Analisis Deskriptif Variabel Penelitian Berskala Numerik

Variabel	Rata-rata	Varians	Nilai Tengah	Min.	Maks.
Total pengeluaran RT (juta)	2.414	2.068	1.929	0.179	29.612
Proporsi Anggota RT Perempuan (persen)	0,55	0,22	0,5	0	1
Jumlah anggota rumah tangga (orang)	4	2	4	1	14

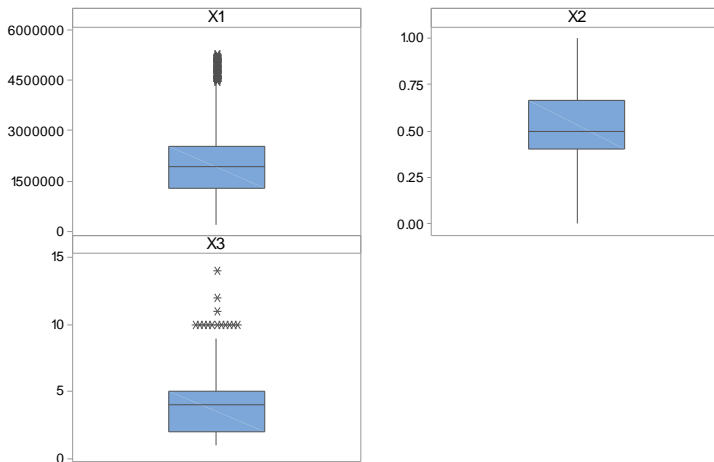
Variabel total pengeluaran rumah tangga memiliki rata-rata sebesar 2.414 juta rupiah per rumah tangga dengan total pengeluaran minimal 179 ribu rupiah dengan rumah tangga yang tinggal berada di Kabupaten Bangkalan. Total pengeluaran terbesar sebesar 29.612 juta rupiah dengan rumah tangga yang berada di Kabupaten Sumenep dan nilai persebaran data dari total pengeluaran sebesar 2.068 juta rupiah.

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. Nilai proporsi terbesar sebesar 1 persen dengan proporsi anggota rumah tangga perempuan terbanyak adalah Kabupaten Sumenep dan nilai persebaran data dari proporsi anggota rumah tangga perempuan sebesar 0,22 persen.

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 lebih banyak tinggal di Kabupaten Sumenep. Kemudian jumlah anggota rumah tangga terbanyak sebesar 14 orang, dengan rumah tangga yang tinggal berada di Kabupaten Pamekasan. Nilai varian sebesar 2 orang pada jumlah anggota rumah tangga memiliki keberagaman yang cukup kecil.

Selanjutnya, dapat dilihat pada *boxplot* dengan mendeteksi *outlier* jika jumlah *outlier* tersebut mendekati atau lebih dari 10%. Penelitian ini telah dilakukan imputasi *mean* dalam mengatasi *outlier* pada data dengan jumlah *outlier* yang mendekati 10%.

Hal ini berpengaruh pada hasil pengelompokan, jika tidak diatasi *outlier* tersebut. Variabel penelitian berskala numerik dengan jumlah *outlier* yang paling banyak antara lain adalah variabel total pengeluaran rumah tangga (X_1) dengan banyaknya *outlier* sebesar 6% sehingga dapat dihasilkan *boxplot* untuk masing-masing variabel dengan ditampilkan pada Gambar 4.1.

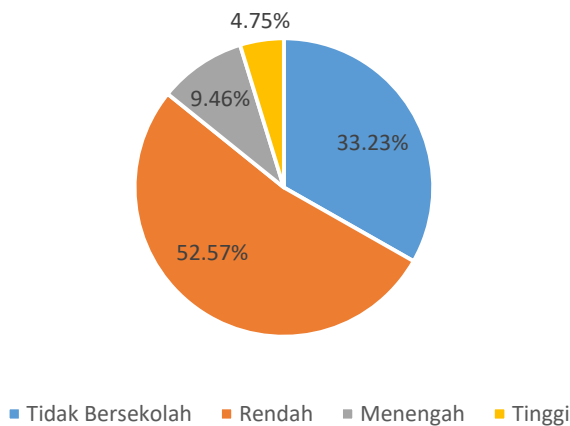


Gambar 4. 1 Karakteristik Data pada Variabel Numerik

Dari tiga variabel yang berskala numerik, terdapat tiga variabel yang terlihat dalam aspek *outlier* yaitu variabel total pengeluaran rumah tangga (X_1) dan jumlah anggota rumah tangga (X_3). Adapun yang tidak memiliki *outlier* yaitu variabel proporsi anggota rumah tangga perempuan (X_2).

4.1.2 Karakteristik Variabel Penelitian Berskala Kategorik

Pada penelitian ini digunakan variabel penelitian berskala kategorik dengan enam variabel antara lain adalah pendidikan terakhir kepala rumah tangga, sumber air minum, sumber air masak, sumber air mandi/cuci, sistem saluran sumber air dan rumah tangga yang membayar air PAM. Berikut hasil persentase dari variabel penelitian pada pendidikan terakhir kepala rumah tangga ditampilkan pada Gambar 4.2.



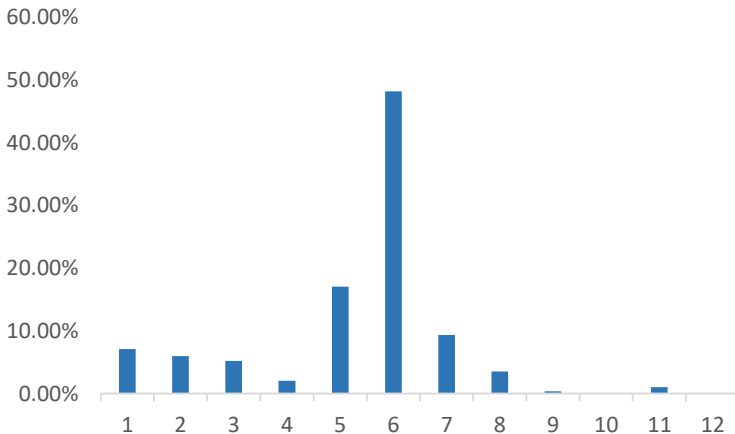
Sumber: SUSENAS 2016 (diolah)

Gambar 4. 2 Persentase Pendidikan Terakhir untuk Kepala Rumah Tangga

Hasil karakteristik pada Gambar 4.2 menyajikan proporsi pendidikan terakhir untuk kepala rumah tangga dengan proporsi terbesar pada kategori pendidikan rendah sebesar 1.628 orang (52.57%) diikuti dengan kategori tidak bersekolah sebesar 1.029 orang (33.23%). Sedangkan kategori pendidikan menengah sebesar 293 orang (9.46%) dan paling sedikit adalah kategori pendidikan tinggi sebesar 147 orang (4.75%). Maka jenjang pendidikan untuk kepala rumah tangga di Pulau Madura masih

termasuk rendah dengan jenjang pendidikan terakhir paling banyak adalah pendidikan rendah dan tidak bersekolah.

Variabel penelitian pada sumber air minum dapat diolah berdasarkan hasil persentase. Adapun hasil persentase dengan ditampilkan pada Gambar 4.3.



Sumber: SUSENAS 2016 (diolah)

Kategori:

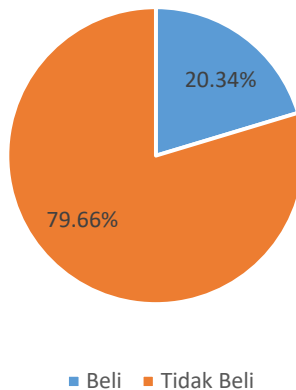
- | | |
|----------------------|-----------------------------|
| 1 = Air bermerk | 7 = Sumur tak terlindung |
| 2 = Air isi ulang | 8 = Mata air terlindung |
| 3 = Ledeng Meteran | 9 = Mata air tak terlindung |
| 4 = Ledeng Eceran | 10 = Air permukaan |
| 5 = Sumur bor/pompa | 11 = Air hujan |
| 6 = Sumur terlindung | 12 = Lainnya |

Gambar 4. 3 Persentase Sumber Air Minum

Karakteristik pada Gambar 4.3, dengan menyajikan sumber air minum yang digunakan oleh rumah tangga di Pulau Madura terbanyak digunakan adalah sumber air yang digunakan sumur terlindung sebesar 1.493 rumah tangga (48,24%) diikuti dengan air sumur bor/pompa sebesar 529 rumah tangga (17.08%), sumur tak terlindung sebanyak 289 rumah tangga (9.33%). Kemudian untuk

sumber air bermerk sebanyak 220 rumah tangga (7,10%), air isi ulang sebanyak 185 rumah tangga (5,97%). Sedangkan sumber air minum dengan air ledeng meteran sebanyak 160 rumah tangga (6,20%), ledeng eceran sebanyak 64 rumah tangga (2,07%), mata air terlindung sebanyak 109 rumah tangga (3,52%) dan mata air tak terlindung sebanyak 12 rumah tangga (0,39%). Sedangkan sumber air yang paling sedikit dikonsumsi adalah air hujan sebanyak 32 rumah tangga (1,03%), air permukaan sebanyak 2 (0,06%) dan lainnya sebanyak 1 (0,03%).

Dengan melihat hasil persentase dari sumber air minum menurut data SUSENAS 2016, maka kategori dapat di *recording* dengan kategori baru. Hal ini dilakukan agar menghasilkan pengelompokan yang efektif. Berikut hasil kategori baru ditampilkan pada Gambar 4.4.

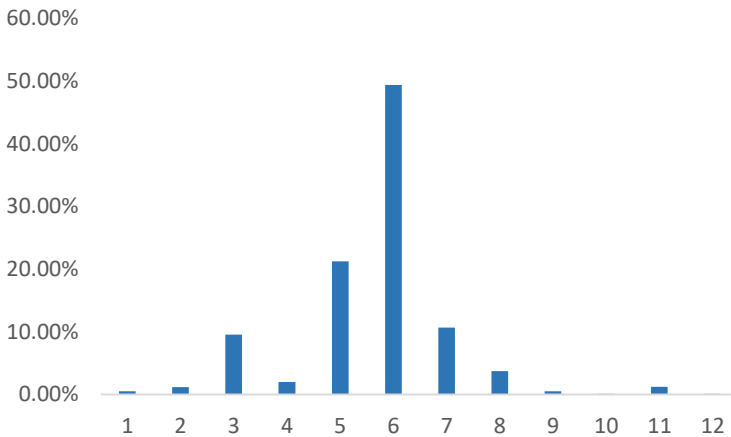


Gambar 4. 4 Persentase Sumber Air Minum dengan Kategori Baru

Berdasarkan Gambar 4.4, terbentuk kategori baru yaitu kategori air minum yang dapat dibeli dan tidak dapat dibeli. Sumber air minum yang dapat dibeli terdiri dari air bermerk, air isi ulang, ledeng meteran, dan ledeng eceran. Selebihnya termasuk kategori air tidak dapat dibeli. Kategori air minum yang dapat

dibeli lebih sedikit digunakan oleh rumah tangga sebesar 630 (20,34%) rumah tangga, daripada air yang tidak dapat dibeli sebanyak 2467 (79,66%) rumah tangga.

Variabel penelitian pada sumber air masak dapat diolah berdasarkan hasil persentase. Adapun hasil persentase dengan ditampilkan pada Gambar 4.4.



Sumber: SUSENAS 2016 (diolah)

Kategori :

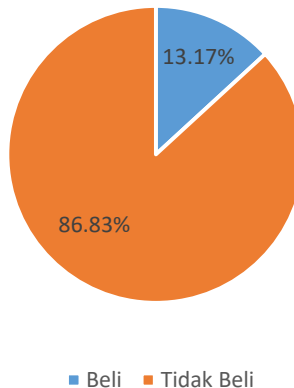
- | | |
|----------------------|-----------------------------|
| 1 = Air bermerk | 7 = Sumur tak terlindung |
| 2 = Air isi ulang | 8 = Mata air terlindung |
| 3 = Ledeng Meteran | 9 = Mata air tak terlindung |
| 4 = Ledeng Eceran | 10 = Air permukaan |
| 5 = Sumur bor/pompa | 11 = Air hujan |
| 6 = Sumur terlindung | 12 = Lainnya |

Gambar 4.5 Persentase Sumber Air Masak

Karakteristik pada Gambar 4.5, dengan menyajikan sumber air masak yang digunakan oleh rumah tangga di Pulau Madura terbanyak digunakan sumber air yang digunakan sumur terlindung sebesar 1.528 rumah tangga (49,34%) diikuti dengan air sumur bor/pompa sebesar 658 rumah tangga (21,25%) dan sumur tak

terlindung sebanyak 330 rumah tangga (10,66%). Kemudian untuk sumber air bermerk sebanyak 16 rumah tangga (0,52%), air isi ulang sebanyak 36 rumah tangga (1,16%). Sedangkan sumber air masak dengan air ledeng meteran sebanyak 295 rumah tangga (9,53%), ledeng eceran sebanyak 61 rumah tangga (1,97%), mata air terlindung sebanyak 115 rumah tangga (3,71%) dan mata air tak terlindung sebanyak 15 rumah tangga (0,48%). Sedangkan sumber air yang paling sedikit dikonsumsi adalah air hujan sebanyak 38 rumah tangga (1,23%), air permukaan sebanyak 2 (0,06%) dan lainnya sebanyak 3 (0,10%).

Dengan melihat hasil persentase dari sumber air masak menurut data SUSENAS 2016, maka kategori dapat di *recording* dengan kategori baru. Hal ini dilakukan agar menghasilkan pengelompokan yang efektif. Berikut hasil kategori baru ditampilkan pada Gambar 4.6.

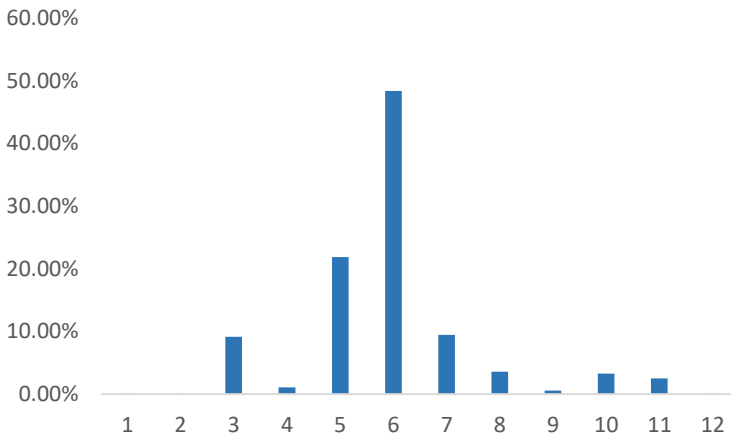


Gambar 4. 6 Persentase Sumber Air Masak dengan Kategori Baru

Pada Gambar 4.6, terbentuk kategori baru yaitu kategori air masak yang dapat dibeli dan tidak dapat dibeli. Sumber air masak yang dapat dibeli terdiri dari air bermerk, air isi ulang, ledeng meteran, dan ledeng eceran. Selebihnya termasuk kategori air tidak

dapat dibeli. Kategori air masak yang dapat dibeli lebih sedikit digunakan oleh rumah tangga sebesar 408 (13,17%) rumah tangga, daripada air yang tidak dapat dibeli sebanyak 2.689 (86,83%) rumah tangga.

Variabel penelitian pada sumber air mandi/cuci dapat diolah berdasarkan hasil persentase. Adapun hasil persentase dengan ditampilkan pada Gambar 4.7.



Sumber: SUSENAS 2016 (diolah)

Kategori :

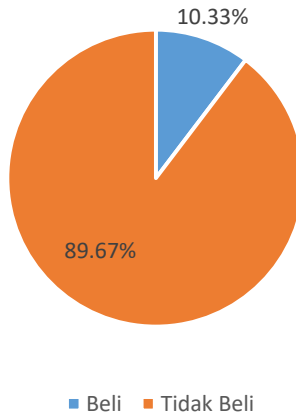
- | | |
|----------------------|-----------------------------|
| 1 = Air bermerk | 7 = Sumur tak terlindung |
| 2 = Air isi ulang | 8 = Mata air terlindung |
| 3 = Ledeng Meteran | 9 = Mata air tak terlindung |
| 4 = Ledeng Eceran | 10 = Air permukaan |
| 5 = Sumur bor/pompa | 11 = Air hujan |
| 6 = Sumur terlindung | 12 = Lainnya |

Gambar 4. 7 Persentase Sumber Air Mandi/Cuci

Karakteristik pada Gambar 4.7, dengan menyajikan sumber air mandi/cuci yang digunakan oleh rumah tangga di Pulau Madura dengan sumber air mandi/cuci terbanyak adalah sumur terlindung sebesar 1.497 rumah tangga (48,34%) diikuti dengan air sumur

bor/pompa sebesar 677 rumah tangga (21,86%) dan sumur tak terlindung sebanyak 293 rumah tangga (9,46%). Kemudian untuk sumber air hujan sebanyak 78 rumah tangga (2,52%), air permukaan sebanyak 102 (3,29%), ledeng meteran sebanyak 283 rumah tangga (9,11%), ledeng eceran sebanyak 34 rumah tangga (1,10%), mata air terlindung sebanyak 111 rumah tangga (3,59%) dan mata air tak terlindung sebanyak 17 rumah tangga (0,55%). Sedangkan sumber air paling sedikit yang dikonsumsi adalah air bermerk sebanyak 2 rumah tangga (0,06%), air isi ulang sebanyak 1 rumah tangga (0,03%), dan lainnya sebanyak 2 rumah tangga (0,06%).

Dengan melihat hasil persentase dari sumber air mandi/cuci menurut data SUSENAS 2016, maka kategori dapat di *recording* dengan kategori baru. Hal ini dilakukan agar menghasilkan pengelompokan yang efektif. Berikut hasil kategori baru ditampilkan pada Gambar 4.8.

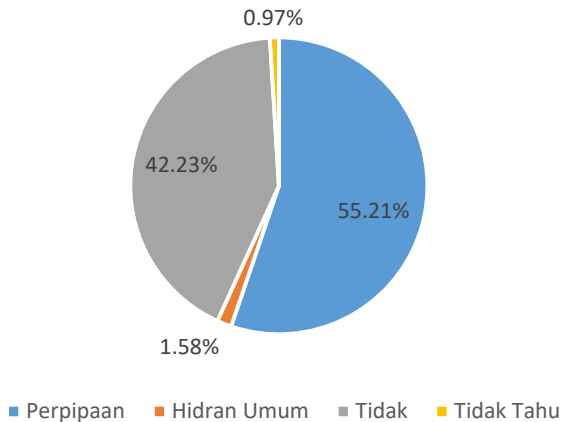


Gambar 4. 8 Persentase Sumber Air Mandi/Cuci dengan Kategori Baru

Karakteristik Gambar 4.8, terbentuk kategori baru yaitu kategori air mandi/cuci yang dapat dibeli dan tidak dapat dibeli. Sumber air mandi/cuci yang dapat dibeli terdiri dari air bermerk, air isi ulang, ledeng meteran, dan ledeng eceran. Selebihnya

termasuk kategori air tidak dapat dibeli. Kategori air mandi/cuci yang dapat dibeli lebih sedikit digunakan oleh rumah tangga sebesar 320 (10,33%) rumah tangga, daripada air yang tidak dapat dibeli sebanyak 2.777 (89,67%) rumah tangga.

Variabel penelitian pada sistem saluran sumber air dapat diolah berdasarkan hasil persentase. Adapun hasil persentase dengan ditampilkan pada Gambar 4.7.



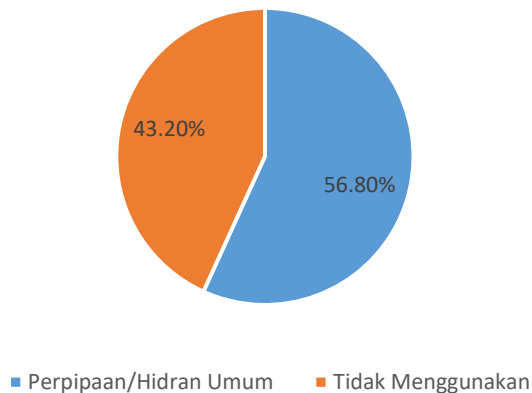
Sumber: SUSENAS 2016 (diolah)

Gambar 4. 9 Persentase Sistem Saluran Sumber Air

Pada karakteristik variabel ini menyajikan dengan penggunaan saluran sumber air yang paling banyak digunakan oleh rumah tangga di Pulau Madura adalah saluran dengan perpipaan sebesar 1.710 rumah tangga (55,21%). Tetapi rumah tangga di Pulau Madura juga masih banyak yang tidak menggunakan sistem saluran sumber air baik melalui perpipaan maupun hidran umum sebanyak 1.308 rumah tangga (42,32%), dan rumah tangga yang tidak tahu menggunakan saluran sumber air sebesar 30 rumah tangga (0,97%). Hal ini berkaitan dengan penyaluran air bersih yang kurang merata dengan seiringnya kekeringan yang melanda empat kabupaten di Pulau Madura. Selain itu penggunaan saluran

air dengan hidran umum paling sedikit sebesar 49 rumah tangga (1.58%).

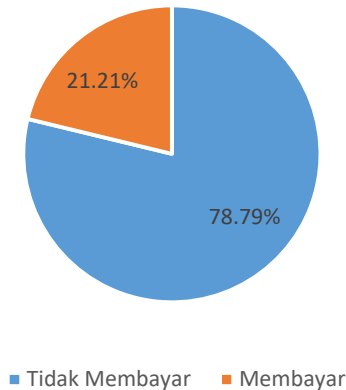
Dengan melihat hasil persentase dari sistem saluran sumber air menurut data SUSENAS 2016, maka kategori dapat di *recording* dengan kategori baru. Hal ini dilakukan agar menghasilkan pengelompokan yang efektif. Berikut hasil kategori baru ditampilkan pada Gambar 4.10.



Gambar 4. 10 Persentase Sistem Saluran Sumber Air dengan Kategori Baru

Karakteristik Gambar 4.10, terbentuk kategori baru yaitu kategori saluran sumber air perpipaan/hidran umum dan tidak menggunakan saluran sumber air. Sistem saluran sumber air perpipaan/hidran umum terdiri dari rumah tangga yang menggunakan sistem saluran perpipaan maupun hidran umum. Selebihnya termasuk kategori tidak menggunakan yaitu tidak maupun tidak tahu menggunakan sistem saluran air. Kategori perpipaan/hidran umum digunakan oleh rumah tangga sebesar 1.759 (78,79%) rumah tangga. Namun, yang tidak menggunakan sistem saluran air sebanyak 1.338 (43,20%) rumah tangga

Variabel penelitian pada pembayaran air PAM dapat diolah berdasarkan hasil persentase dengan *pie chart*. Adapun hasil persentase dengan ditampilkan pada Gambar 4.11.



Sumber: SUSENAS 2016 (diolah)

Gambar 4. 11 Persentase Rumah Tangga yang Membayar Air PAM

Pada karakteristik variabel penelitian pembayaran air PAM menyajikan dengan rumah tangga yang membayar air PAM di Pulau Madura. Dari Gambar 4.11, bahwa rumah tangga yang paling banyak tidak membayar air PAM sebesar 2.439 rumah tangga (78,79%). Sedangkan yang membayar air PAM hanya sebanyak 656 rumah tangga (21,21%). Hal ini membuktikan bahwa rumah tangga di Pulau Madura kurang menggunakan air PAM, sehingga PDAM di tiga dari empat Kabupaten di Pulau Madura termasuk dalam kategori kurang sehat.

4.2 Hasil Pengelompokan dengan Metode SWFM

Pada analisis ini dijelaskan mengenai analisis yang dilakukan untuk mengelompokkan rumah tangga di Pulau Madura. Analisis data berskala numerik dilakukan dengan menggunakan

metode hirarki *agglomerative* dengan teknik pengelompokan *ward*. Sedangkan untuk data berskala kategorik digunakan pengelompokan dengan metode *k-modes*.

4.2.1 Pengelompokan pada Data Numerik

Pengelompokan untuk data berskala numerik dilakukan menggunakan metode hirarki *ward* dengan setiap objek pengamatan sebagai suatu kelompok dengan anggota tunggal. Tahap selanjutnya adalah membentuk matriks jarak antar objek pengamatan. Jarak yang digunakan dalam penelitian ini adalah jarak *Euclidean* yang diperoleh dari 3.097 objek pengamatan. 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*. Tabel 4.2 merupakan hasil pengelompokan data numerik berdasarkan nilai *Pseudo F-statistic* dari masing-masing jumlah kelompok yang digunakan.

Tabel 4. 2 Nilai *Pseudo-F* masing-masing jumlah kelompok

Jumlah Kelompok	<i>Pseudo F</i>
2	1001,172
3	500,4242
4	333,5083
5	250,0504
6	199,9756

Tabel 4.2 menyajikan nilai *pseudo-F* pada masing-masing jumlah kelompok. Jika nilai *pseudo-F* semakin besar maka semakin baik hasil pengelompokannya atau jumlah kelompok tersebut optimum. Nilai pada pengelompokan dengan jumlah kelompok sebanyak 2 sebesar 1001,172 yang artinya pengelompokan jumlah klaster sebanyak 2 adalah jumlah kelompok optimum, karena memiliki nilai *pseudo-F* yang lebih besar. Berdasarkan pembagian jumlah anggota kelompok dengan metode hirarki *ward* untuk jumlah 2 kelompok adalah Kelompok 1 : 1.782 rumah tangga

Kelompok 2 : 1.315 rumah tangga

Sedangkan karakteristik data yang digunakan adalah nilai rata-rata dari masing-masing kelompok yang terbentuk. Berikut hasil dari rata-rata jumlah kelompok 2 pada Tabel 4.3 adalah sebagai berikut.

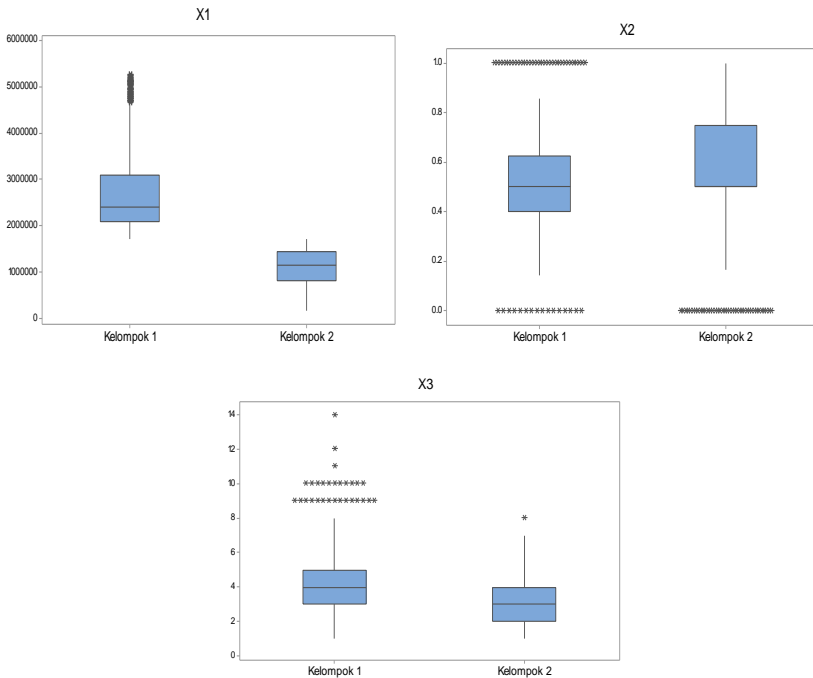
Tabel 4. 3 Karakteristik Data pada Hasil Pengelompokan Data Berskala Numerik

Variabel	Kelompok 1	Kelompok 2
Total Pengeluaran Rumah Tangga (juta)	3,36	1,13
Proporsi Anggota Rumah Tangga Perempuan (persen)	0,51	0,61
Jumlah Anggota Rumah Tangga (orang)	4	3

Tabel 4.3, menunjukkan nilai rata-rata dari anggota kelompok masing-masing. Berdasarkan Tabel 4.3, 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.

Selanjutnya dilakukan pengecekan nilai *outlier* dengan *boxplot* untuk mengetahui nilai *outlier* pada masing-masing kelompok. Berikut hasil karakteristik dengan melihat *boxplot* yang ditampilkan pada Gambar 4.13.



Gambar 4. 12 Karakteristik Hasil Pengelompokan pada Data Numerik.

Gambar 4.12, menunjukkan *boxplot* dari anggota kelompok pertama, kedua dan ketiga. Dari tiga variabel hasil pengelompokan berskala numerik, menunjukkan pada kelompok 2 hanya memiliki sedikit nilai *outlier*.

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 numerik dapat ditampilkan pada Tabel 4.4.

Tabel 4. 4 Hasil Kelompok Data Numerik pada Rumah Tangga yang Tinggal disetiap Kabupaten di Pulau Madura

Kabupaten	Kelompok 1	Kelompok 2	Total
Bangkalan	503	243	746
Sampang	473	285	758
Pamekasan	326	431	757
Sumenep	480	356	836
Total	1782	1315	3097

Tabel 4.4 menunjukkan hasil pengelompokan rumah tangga yang tinggal disetiap Kabupaten di Pulau Madura. Berdasarkan Tabel 4.4 didapatkan rumah tangga yang jumlahnya lebih sedikit dalam kelompok 2 (perekonomian rendah) adalah rumah tangga yang tinggal di Kabupaten Bangkalan. Sebaliknya, rumah tangga yang tinggal dengan jumlah yang sedikit dalam kelompok 1 (perekonomian tinggi) adalah rumah tangga yang tinggal di Kabupaten Pamekasan.

4.2.2 Pengelompokan pada Data Kategorik

Pengelompokan pada data variabel berskala kategorik dengan melakukan metode *k-modes* dengan setiap kelompok objek pengamatan sabagai 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 jumlah kelompok dengan pembagian kelompok adalah.

Kelompok 1 : 1.975 rumah tangga

Kelompok 2 : 723 rumah tangga

Kelompok 3 : 399 rumah tangga

Karakteristik variabel berskala kategorik dari hasil pengelompokan dengan metode non-hirarki *k-modes* dapat ditunjukkan melalui nilai yang paling dominan atau nilai modus. Perbandingan karakteristik rumah tangga berdasarkan variabel kategorik antar kelompok disajikan pada Tabel 4.5.

Tabel 4. 5 Karakteristik Data pada Hasil Pengelompokan Data Berskala Kategorik

Variabel		Kel 1	Kel 2	Kel 3
Pendidikan KRT	Tidak Bersekolah	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%
Sumber Air Minum	Dapat dibeli	9,47%	7,19%	97,99%
	Tidak dapat dibeli	90,53%	92,81%	2,01%
Sumber Air Masak	Dapat dibeli	0,96%	0,69%	96,24%
	Tidak dapat dibeli	99,04%	99,31%	3,76%
Sumber Air Mandi/Cuci	Dapat dibeli	0,15%	0,14%	79,20%
	Tidak dapat dibeli	98,85%	99,86%	20,80%
Sistem Saluran Sumber Air	Perpipaan atau Hidran Umum	70,78%	0,00%	90,48%
	Tidak Menggunakan	29,22%	100,00%	9,52%
Pembayaran Air PAM	Tidak Membayar	87,80%	89,21%	15,29%
	Membayar	12,20%	10,79%	84,71%

Tabel 4.5 merupakan penyajian persentase dari kategori yang paling dominan dari variabel-variabel yang ada pada kelompok. Berdasarkan Tabel 4.5, 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 dapat diketahui 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.

Dengan mengetahui hasil karakteristik dari pengelompokan variabel berskala kategorik, maka dapat dilihat dari masing-masing kelompok menunjukkan hasil karakteristik berdasarkan perekonomian rumah tangga di Pulau Madura. Hal tersebut terlihat dari karakteristik rumah tangga, berikut hasil dari kriteria masing-masing kelompok.

Kelompok 1 : Perekonomian rendah

Kelompok 2 : Perekonomian sedang

Kelompok 3 : Perekonomian tinggi.

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 kategorik dapat ditampilkan pada Tabel 4.6.

Tabel 4. 6 Hasil Kelompok Data Kategorik pada Rumah Tangga yang Tinggal di setiap Kabupaten di Pulau Madura

Kabupaten	Kelompok 1	Kelompok 2	Kelompok 3	Total
Bangkalan	463	205	78	746
Sampang	499	157	102	758
Pamekasan	595	59	103	757
Sumenep	418	302	116	836
Total	1975	723	399	3.097

Berdasarkan Tabel 4.6, didapatkan rumah tangga yang tinggal di Kabupaten Bangkalan lebih sedikit (19,55%) jumlah rumah tangga yang tinggal dalam kelompok 3 atau kriteria rumah

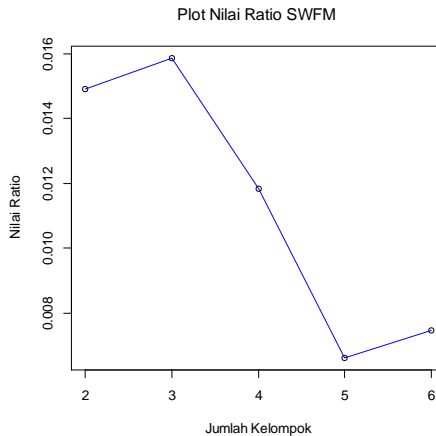
tangga dengan ciri-ciri perekonomian tinggi. Rumah tangga yang tinggal di Kabupaten Sumenep dalam kelompok 1 atau ciri-ciri rumah tangga dengan perekonomian rendah lebih sedikit jumlahnya (21,16%). Dari masing-masing jumlah rumah tangga yang tinggal dalam masing-masing kelompok ini tersebar merata untuk kriteria atau ciri-ciri dari kelompok 1, kelompok 2 dan kelompok 3.

4.2.3 Pengelompokan pada Data Campuran

Pengelompokan data campuran menggunakan teknik SWFM (*Similarity Weight and Filter Method*) dengan melakukan pengelompokan masing-masing pada hasil pengelompokan sebelumnya (data berskala numerik dan data berskala 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 untuk data berskala numerik yang diperoleh pada subbab 4.2.1 dan hasil pengelompokan untuk data berskala kategorik diperoleh dari analisis diperoleh dinyatakan sebagai variabel berskala kategorik dari analisis pada subbab 4.2.2. Hasil pengelompokan yang telah diperoleh dinyatakan sebagai variabel kategorik yang nantinya akan diolah dengan metode SWFM.

Tahap selanjutnya membentuk matriks jarak antar objek pengamatan menggunakan jarak *similarity weight*, yang kemudian dibentuk nilai jarak F antar objek pengamatan. Setelah itu dilakukan penggabungan kelompok dengan jarak terdekat dan pembaruan matriks jarak. Jumlah kelompok yang dibentuk antara dua sampai enam kelompok yang dilakukan dengan *software R*. Hasil pengelompokan terbaik ditentukan dari nilai rasio Sw dan Sb terkecil. Berikut ini merupakan plot nilai rasio pada Gambar 4.14.



Gambar 4. 13 Plot Nilai Rasio SWFM

Hasil pengelompokan terbaik yang disajikan pada Gambar 4.13 menunjukkan bahwa nilai rasio S_w dan S_b terkecil pada pengelompokan dengan jumlah kelompok 5 sebesar 0,006627. Nilai rasio S_w dan S_b sebesar 0,006627 menunjukkan bahwa simpangan baku dalam kelompok bernilai 0,006627 kali dari simpangan baku kelompok lain sehingga variansi data dalam kelompok juga lebih kecil dibanding variansi kelompok lain. Berikut hasil nilai rasio pada masing-masing jumlah kelompok data berskala campuran.

Tabel 4. 7 Nilai Rasio Hasil Pengelompokan dengan Metode SWFM

Jumlah Kelompok	Rasio S_w dan S_b
2	0,014908
3	0,015859
4	0,011841
5	0,006627
6	0,007455

Hasil pengelompokan terbaik yang disajikan pada Tabel 4.6 dapat terlihat bahwa nilai rasio terkecil berada pada jumlah

kelompok 5. Pembagian jumlah anggota kelompok dengan metode SWFM untuk jumlah 5 kelompok adalah.

Kelompok 1 : 1063 rumah tangga

Kelompok 2 : 1359 rumah tangga

Kelompok 3 : 276 rumah tangga

Kelompok 4 : 272 rumah tangga

Kelompok 5 : 127 rumah tangga

Setelah didapatkan banyaknya kelompok optimum, selanjutnya adalah menentukan anggota pada setiap kelompok yang terbentuk dan menentukan karakteristik variabel dari anggota hasil pengelompokan pada analisis SWFM. Karakteristik data numerik dan kategorik pada masing-masing kelompok disajikan pada Tabel 4.8 dan Tabel 4.9 sebagai berikut.

Tabel 4. 8 Nilai Rata-rata Data Numerik pada Hasil Pengelompokan dengan Metode SWFM

Variabel	Karakteristik pada Kelompok ke-				
	1	2	3	4	5
Total Pengeluaran RT (juta)	3.30	1.71	1.20	4.41	1.23
Proporsi ART Perempuan (persen)	0,51	0,58	0,57	0,51	0,64
Jumlah ART (orang)	4	3	3	4	3

Tabel 4. 9 Karakteristik Data Kategorik Hasil Pengelompokan dengan Metode SWFM

Variabel	Karakteristik pada Kelompok ke-				
	1	2	3	4	5
Pendidikan KRT	Rendah (44,03%)	Rendah (50,99%)	Rendah (90,22%)	Rendah (52,21%)	Rendah (59,84%)
Sumber Air Minum	Tidak dapat dibeli (84,01%)	Tidak dapat dibeli (95,66%)	Tidak dapat dibeli (96,38%)	Dapat dibeli (98,53%)	Dapat dibeli (96,85%)

Tabel 4.9 Karakteristik Data Kategorik Hasil Pengelompokan dengan Metode SWFM (Lanjutan)

Variabel	Karakteristik Kelompok ke-				
	1	2	3	4	5
Sumber Air Masak	Tidak Beli (98,40%)	Tidak dapat dibeli (99,48%)	Tidak dapat dibeli (100,00%)	Dapat dibeli (96,69%)	Dapat dibeli (95,28%)
Sumber Air Mandi/Cuci	Tidak Beli (99,91%)	Tidak Beli (99,85%)	Tidak dapat dibeli (99,64%)	Dapat dibeli (84,19%)	Dapat dibeli (68,50%)
Sistem Saluran Sumber Air	Perpipaan atau Hidran umum (76,39%)	Tidak Menggunakan (56,88%)	Tidak Menggunakan (100,00%)	Perpipaan atau Hidran Umum (94,12%)	Perpipaan atau Hidran Umum (82,68%)
Pembayaran Air PAM	Tidak Membayar (87,02%)	Tidak Membayar (88,67%)	Tidak Membayar (90,22%)	Membayar (79,04%)	Membayar (96,85%)

Berdasarkan hasil ringkasan yang disajikan pada Tabel 4.8 dan Tabel 4.9 maka karakteristik rumah tangga pada tiap kelompok dapat dideskriptifkan antara lain adalah.

1. Kelompok 1

Terdapat 1.063 rumah tangga yang termasuk pada kelompok 1 dengan karakteristik pendidikan terakhir kepala rumah tangga adalah berpendidikan rendah. Dari sumber air minum, masak dan mandi/cuci yang digunakan oleh rumah tangga ini tergolong air yang tidak dapat dibeli. Selain itu karakteristik sistem saluran sumber air bersih yang digunakan sebagian besar menggunakan perpipaan atau hidran umum. Karakteristik untuk pembayaran air PAM sebagian besar tidak melakukan pembayaran. Rata-rata total pengeluaran rumah tangga sebesar 3.30 juta. Proporsi anggota rumah tangga perempuan yang tinggal dalam satu rumah rata-rata proporsinya sebesar 51 persen. Rata-rata jumlah anggota rumah tangga dalam satu rumah sebesar 4 orang.

2. Kelompok 2

Kelompok 2 beranggotakan 1.359 rumah tangga dengan pendidikan terakhir kepala rumah tangga sebagian besar berpendidikan rendah. Dengan sumber air minum, masak dan mandi/cuci yang digunakan oleh rumah tangga ini tergolong air yang tidak dapat dibeli. Selain itu karakteristik sistem saluran sumber air bersih yang digunakan sebagian besar tidak menggunakan perpipaan atau hidran umum. Karakteristik untuk pembayaran air PAM sebagian besar tidak melakukan pembayaran. Rata-rata total pengeluaran rumah tangga sebesar 1.71 juta. Proporsi anggota rumah tangga perempuan yang tinggal dalam satu rumah rata-rata proporsinya sebesar 58 persen. Rata-rata jumlah anggota rumah tangga yang tinggal dalam satu rumah sebesar 3 orang.

3. Kelompok 3

Kelompok 3 merupakan kelompok dengan jumlah anggota rumah tangga sebanyak 276 rumah tangga. Sebagian besar rumah tangga berpendidikan rendah untuk pendidikan terakhir kepala rumah tangga. Dengan sumber air minum, masak dan mandi/cuci yang digunakan oleh rumah tangga ini tergolong air yang tidak dapat dibeli. Selain itu karakteristik sistem saluran sumber air bersih yang digunakan sebagian besar tidak menggunakan perpipaan atau hidran umum. Karakteristik untuk pembayaran air PAM sebagian besar tidak melakukan pembayaran. Rata-rata total pengeluaran rumah tangga sebesar 1.20 juta. Proporsi anggota rumah tangga perempuan yang tinggal dalam satu rumah rata-rata proporsinya sebesar 57 persen. Rata-rata jumlah anggota rumah tangga yang tinggal dalam satu rumah sebesar 3 orang.

4. Kelompok 4

Kelompok 4 memiliki anggota sebanyak 272 rumah tangga dengan sebagian besar pendidikan terakhir kepala rumah tangga adalah berpendidikan rendah. Dengan sumber air minum, masak dan mandi/cuci yang digunakan oleh rumah tangga ini tergolong air yang dapat dibeli. Selain itu karakteristik sistem saluran

sumber air bersih yang digunakan sebagian besar menggunakan perpipaan atau hidran umum. Karakteristik untuk pembayaran air PAM sebagian besar melakukan pembayaran. Rata-rata total pengeluaran rumah tangga sebesar 4.41 juta. Proporsi anggota rumah tangga perempuan yang tinggal dalam satu rumah rata-rata proporsinya sebesar 51 persen. Rata-rata jumlah anggota rumah tangga yang tinggal dalam satu rumah sebesar 4 orang.

5. Kelompok 5

Kelompok 5 adalah kelompok yang memiliki anggota rumah tangga paling sedikit yaitu sebanyak 127 rumah tangga. Sebagian besar pendidikan terakhir kepala rumah tangga adalah berpendidikan rendah. Dengan sumber air minum, masak dan mandi/cuci yang digunakan oleh rumah tangga ini tergolong air yang dapat dibeli. Selain itu karakteristik sistem saluran sumber air bersih yang digunakan sebagian besar menggunakan perpipaan atau hidran umum. Karakteristik untuk pembayaran air PAM sebagian besar melakukan pembayaran. Rata-rata total pengeluaran rumah tangga sebesar 1.23 juta. Kemudian proporsi anggota rumah tangga perempuan yang tinggal dalam satu rumah rata-rata proporsinya sebesar 64 persen. Rata-rata jumlah anggota rumah tangga yang tinggal dalam satu rumah sebesar 3 orang.

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.

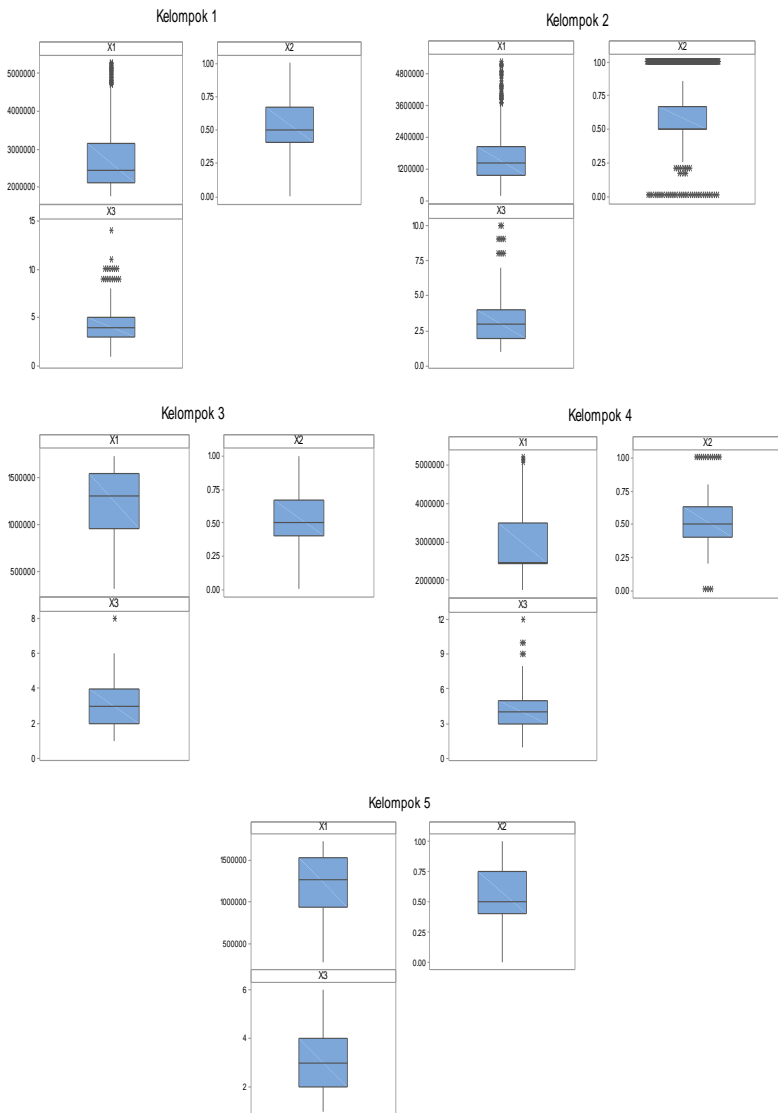
kelompok 2 : Perekonomian sedang

kelompok 3 : Perekonomian sangat rendah

kelompok 4 : Perekonomian paling tinggi

kelompok 5 : Perekonomian rendah.

Selanjutnya dilakukan pengecekan nilai *outlier* dengan *boxplot* pada masing-masing kelompok yang ditampilkan pada Gambar 4.14.



Gambar 4. 14 Karakteristik Data Numerik pada Hasil Pengelompokan SWFM.

Dari tiga variabel hasil pengelompokan dengan SWFM pada Gambar 4.14, menunjukkan untuk kelompok 3 dan kelompok 5 memiliki *outlier* paling sedikit di antara kelompok-kelompok lain.

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 4.10.

Tabel 4. 10 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 4.10 merupakan penyajian hasil kelompok rumah tangga yang tinggal disetiap Kabupaten dengan menggunakan metode SWFM berdasarkan perekonomian rumah tangga. Berdasarkan Tabel 4.10, 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.

4.3 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 data tidak berdistribusi normal (Lampiran 15). Dengan pemeriksaan normal multivariat diperoleh hasil *mshapiro test* yaitu nilai *p-value* adalah $2,2 \times 10^{-16}$ lebih kecil dari nilai α sebesar 0,05 maka didapatkan tidak berdistribusi normal (Lampiran 15).

Uji *Kruskal Wallis* digunakan pada tiga variabel penelitian berskala numerik. Hipotesis yang digunakan pada pengujian *Kruskal Wallis* adalah sebagai berikut.

H_0 : Tidak ada perbedaan rata-rata antar kelompok

H_1 : adanya perbedaan rata-rata antar kelompok

Ringkasan uji *Kruskal Wallis* pada variabel penelitian ditampilkan pada Tabel 4.11.

Tabel 4. 11 Hasil Uji Kruskal Wallis terhadap 5 Kelompok

Variabel	<i>Chi-Square</i>	df	<i>Chi-Square</i> Tabel
Total Pengeluaran RT	1408,271	4	9,49
Proporsi ART Perempuan	60,302	4	9,49
Jumlah ART	442,726	4	9,49

Uji *Kruskal Wallis* ini menggunakan hipotesis yang ada pada Bab 2.8 dengan nilai α sebesar 0,05. Dari pengujian pada Tabel 4.11, tiga variabel penelitian yang digunakan diantaranya memiliki 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.

BAB V

KESIMPULAN DAN SARAN

Bab ini dijelaskan mengenai kesimpulan dan saran yang dirangkum dari bab sebelumnya. Kesimpulan diperoleh dari analisis berdasarkan algoritma dan interpretasi. Kemudian dijelaskan beberapa saran untuk perbaikan dan pengembangan penelitian lebih lanjut agar sesuai.

5.1 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 1782 rumah tangga dan kriteria perekonomian rendah sebanyak 1315 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 1975, 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 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.

5.2 Saran

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

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 yang diduga terkait dengan faktor konsumsi air bersih rumah tangga. Selain itu, variabel yang digunakan sebaiknya dicoba menggunakan *kruskal wallis* multivariat.

DAFTAR PUSTAKA

- Alvionita. (2017). *Metode Ensemble ROCK dan SWFM untuk Pengelompokan Data Campuran Numerik dan Kategorik pada Kasus Akses Jeruk*. Surabaya: Statistika, Institut Teknologi Sepuluh Nopember.
- BPPSPAM. (2017). *Buku Kinerja PDAM 2017*. Jakarta: Kementerian Pekerjaan Umum dan Perumahan Rakyat.
- Bunkers, M., & James, R. M. (1996). Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. *J. Climate*.
- Daniel, W. W. (1989). *Statistik Nonparametrik Terapan*. (A. T. Widodo, Penerj.) Jakarta: PT. Gramedia.
- Darr, P. S., & Kamen, C. (1976). *The Demand for Urban Water*. Martinus Nijhoff Social Division. Leiden.
- Departemen Kesehatan RI. (1990). *Peraturan Menteri Kesehatan No. 416/MEN.KES/PER/IX/1990 Tentang Syarat-syarat dan Pengawasan Kualitas Air*. Jakarta: Departemen Kesehatan RI.
- Dewangan, R. R., Sharma, L. K., & Akasapu, A. K. (2010). Fuzzy Clustering Technique for Numerical and Categorical dataset. *International Journal on Computer Science and Engineering*, 75-80.
- Dewi, A. (2012). *Metode Cluster Ensemble untuk Pengelompokan Desa Pedesaan di Provinsi Riau*. Surabaya: Statistika, Institut Teknologi Sepuluh Nopember.
- Dwijosaputro. (1981). *Dasar-Dasar Mikrobiologi*. Jakarta: Djambatan.
- ESCAP-UNDP. (2002). *ESCAP-UNDP Initiative for The Achievement of Millenium Development Goals in Asia And The Pacific*. Bangkok: UN ESCAP.
- Guha, S., Rastogi, R., & Shim, K. (2000). ROCK : A Robust Clustering Algorithm for Categorical Attributes. *Proceeding of the 15th International Conference on Data Engineering*.

- Hair, J. F., Black, W. C., Babin, J., & Anderson, E. (2009). *Multivariate Data Analysis* (seventh ed.). New Jersey: Prentice Hall Inc.
- Halkidi, Batistakis, & Vizirgiannis. (2001). On Clustering Validation Techniques. *Journal of Intelligent Systems*, 17:2/3, hal 107-145.
- He, Z., Xu, X., & Deng, S. (2005a). A Cluster Ensemble Method For Clustering Categorical Data. *Information Fusion*, hal 143-151.
- He, Z., Xu, X., & Deng, S. (2005b). Clustering Mixed Numeric and Categorical Data: A Cluster Ensemble Approach. *Departement of Computer Science and Engineering, Harbin Institute of Technology*.
- Herliyasari, R. R. (2017). *Metode Ensemble Clustering dengan Algoritma Squeezer untuk Pengelompokan Rumah Tangga di Surabaya Berdasarkan Indikator Kesejahteraan*. Surabaya: Statistika, Institut Teknologi Sepuluh Nopember.
- Joachim, S., & Hillenbrand, T. (2007). Determinants of Residential Water Demand in Germany. *Working Paper Sustainability and Innovation*, No. S 3.
- Johnson, R. A., & Winchern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). New Jersey: Pearson Prentice Hall.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data- An Introduction to Cluster Analysis*. Wiley.
- Kodoatie, R. (2008). *Pengelolaan Sumber Daya Air terpadu*. Yogyakarta: Andi Yogyakarta.
- Ramdhan, R. (2017). *Pengelompokan Desa di Kabupaten Bondowoso Berdasarkan Data Campuran Numerik dan Kategorik Menggunakan Metode Ensembl ROCK*. Surabaya: Statistika, Institut Teknologi Sepuluh Nopember.

- Reddy, M., & Kavitha, B. (2012). *Clustering the Mixed Numerical and Categorical Dataset using Similarity Weight and Filter Method* (Vol. 5 no. 1).
- Sharma, S. (1996). *Applied Multivariate Techniques*. Canada: John Wiley & Sons, Inc.
- Simamora, B. (2005). *Analisis Multivariat Pemasaran*. Jakarta: PT Gramedia Pustaka Umum.
- Suguna, J., & Selvi, M. A. (2012). Ensemble Fuzzy Clustering for Mixed for Numerical and Categorical Data. *International Journal of Computer Application*, 42, 19.
- Sukirno, S. (2004). *Pengantar Teori Makro Ekonomi*. Jakarta: PT. Raja Grafindo Persada.
- Suparman, S. (2007). Kaum Perempuan Paling Peduli. Artikel dalam Percik, Media Informasi Air Minum dan Penyehatan Lingkungan. hal. 9-11.
- SUSENAS. (2015). *Survei Sosial Ekonomi Nasional 2015 Maret (KOR)*. Dipetik Mei 31, 2018, dari Microdata BPS: <http://microdata.bps.go.id/mikrodata/index.php/catalog/657/datafile/F2/V239>
- Todaro, M. P. (2000). *Pengembangan Ekonomi*. (H. Munandar, Penerj.) Jakarta: Bumi Aksara.
- Walpole, R. E. (1995). *Pengantar Metode Statistika Ke-3*. (B. Sumantri, Penerj.) Jakarta: PT. Gramedia Pustaka Utama.
- Winarna, S. (2003). *Analisis Konsumsi Air Bersih Pelanggan Rumah tangga Berdasarkan Faktor-faktor yang Mempengaruhinya*. Semarang: Universitas Diponegoro.

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1 Data Penelitian

NO	No. urut RT	KAB	X1	X2	...	X9
1	352621020	26	1795077	0.5	...	2
2	352621021	26	2331168	0.333333	...	1
3	352621022	26	2899283	0.666667	...	1
4	352621023	26	1832265	0.333333	...	1
5	352621024	26	2277622	0.5	...	1
6	352621025	26	2652265	0.571429	...	1
7	352621026	26	965735.1	1	...	1
8	352621027	26	789391.3	1	...	1
9	352621028	26	2013211	0.714286	...	1
10	352621029	26	1525015	0.5	...	1
11	352621030	26	2362372	0.166667	...	1
12	352621031	26	1896172	0.6	...	1
13	352621032	26	3518176	0.571429	...	1
14	352621033	26	1514436	0.5	...	1
15	352621034	26	2332429	0.666667	...	1
...
3090	352924109	29	2894051	0.333333	...	2
3091	352924110	29	1121345	0.5	...	1
3092	352924111	29	4228993	0.666667	...	2
3093	352924112	29	2752833	0.666667	...	2
3094	352924113	29	3464310	0.333333	...	2
3095	352924114	29	3313617	0.75	...	2
3096	352924115	29	3578261	0.5	...	2
3097	352924116	29	6391324	0.666667	...	2

Lampiran 2 Syntax Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 2 dengan Ward

```

data<-read.csv("F:/Kuliah/Semester 8/TA/data SUSENAS
Excel/data (numerik).csv",sep=","header=T)
cluster.2ward = function(data){
  k = 2
  datanumerik = data.frame(data$X1, data$X2, data$X3)
  d = dist(datanumerik, method = "euclidean")
  # Analisis Kluster Hirarki
  fit.ward = hclust(d, method = "ward")
  # Memotong Dendogram untuk k Kluster
  ward = cutree(fit.ward, k=2)
  hasil.kelompok.numerik = data.frame(ward)
  # Menghitung sum of square total  $y = (x_{il} - \bar{x}_l)^2$ 
  y = c((data$X1 - mean(data$X1))^2, (data$X2 -
  mean(data$X2))^2, (data$X3 - mean(data$X3))^2)
  SST = sum (y)
  # Menghitung sum of square within group (ward)
  kelompok.ward = hasil.kelompok.numerik$ward
  data.ward = data.frame(ward, data$X1, data$X2, data$X3)
  data.ward.sort = data.ward[order(data.ward$ward),]
  a = length(data.ward.sort$ward[data.ward.sort$ward==1])
  b = length(data.ward.sort$ward[data.ward.sort$ward==2])
  data.c1.ward = data.ward.sort[1:a,]
  data.c2.ward = data.ward.sort[(a+1):3097,]
  xlc1bar.ward = matrix(c(rep(mean(data.c1.ward$data.X1),a),
  rep(mean(data.c1.ward$data.X2),a),
  rep(mean(data.c1.ward$data.X3),a)), ncol = 3)
  xlc2bar.ward = matrix(c(rep(mean(data.c2.ward$data.X1),b),
  rep(mean(data.c2.ward$data.X2),b),
  rep(mean(data.c2.ward$data.X3),b)), ncol = 3)
  xil.c1.ward =
  matrix(c(data.c1.ward$data.X1,data.c1.ward$data.X2,
  data.c1.ward$data.X3), ncol = 3)

```

```

xil.c2.ward =
  matrix(c(data.c2.ward$data.X1,data.c2.ward$data.X2,
           data.c2.ward$data.X3), ncol = 3)
x1.ward = sum ((xil.c1.ward-xlc1bar.ward)^2)
x2.ward = sum ((xil.c2.ward-xlc2bar.ward)^2)
SSW.ward.c2 = sum (x1.ward,x2.ward)
SSB.ward.c2 = SST - SSW.ward.c2
R.square.ward.c2 = SSB.ward.c2/SST
Hasil.ward = c(SST, SSW.ward.c2, R.square.ward.c2)
Hasil = rbind(Hasil.ward)
Hasil2 = data.frame(hasil.kelompok.numerik)
print("Hasil Pengelompokan Hirarki")
print(hasil.kelompok.numerik)
print("-----")
print(" Untuk 2 kelompok yang terbentuk ")
print("Metode  SST  SSW  R-square")
print(Hasil)
write.csv(Hasil2,file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
kelompok2 ward.csv")
write.csv(Hasil,file="F:/Kuliah/Semester 8/TA/Data Run/Hasil 2
ward.csv")
}

```

Lampiran 3 *Syntax* Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 3 dengan Ward

```

data<-read.csv("F:/Kuliah/Semester 8/TA/data SUSENAS
Excel/data (numerik).csv",sep="," ,header=T)
cluster.3ward = function(data){
  k = 3
  datanumerik = data.frame(data$X1, data$X2, data$X3)
  d = dist(datanumerik, method = "euclidean")
  # Analisis Kluster Hirarki
  fit.ward = hclust(d, method = "ward")
  # Memotong Dendogram untuk k Kluster
  ward = cutree(fit.ward, k=3)
  hasil.kelompok.numerik = data.frame(ward)
  # Menghitung sum of square total  $y = (x_i - \bar{x})^2$ 
  y = c((data$X1-mean(data$X1))^2,(data$X2-
mean(data$X2))^2,(data$X3-mean(data$X3))^2)
  SST = sum (y)
  # Menghitung sum of square within group (ward)
  kelompok.ward = hasil.kelompok.numerik$ward
  data.ward = data.frame(ward, data$X1, data$X2, data$X3)
  data.ward.sort = data.ward[order(data.ward$ward),]
  a = length(data.ward.sort$ward[data.ward.sort$ward==1])
  b = length(data.ward.sort$ward[data.ward.sort$ward==2])
  c = length(data.ward.sort$ward[data.ward.sort$ward==3])
  data.c1.ward = data.ward.sort[1:a,]
  data.c2.ward = data.ward.sort[(a+1):(a+b),]
  data.c3.ward = data.ward.sort[(a+b+1):3097,]
  xlc1bar.ward = matrix(c(rep(mean(data.c1.ward$data.X1),a),
                        rep(mean(data.c1.ward$data.X2),a),
                        rep(mean(data.c1.ward$data.X3),a)), ncol = 3)
  xlc2bar.ward = matrix(c(rep(mean(data.c2.ward$data.X1),b),
                        rep(mean(data.c2.ward$data.X2),b),
                        rep(mean(data.c2.ward$data.X3),b))), ncol = 3)
  xlc3bar.ward = matrix(c(rep(mean(data.c3.ward$data.X1),c),

```



```

        rep(mean(data.c3.ward$data.X2),c),
        rep(mean(data.c3.ward$data.X3),c)), ncol = 3)
xil.c1.ward =
  matrix(c(data.c1.ward$data.X1,data.c1.ward$data.X2,
    data.c1.ward$data.X3), ncol = 3)
xil.c2.ward =
  matrix(c(data.c2.ward$data.X1,data.c2.ward$data.X2,
    data.c2.ward$data.X3), ncol = 3)
xil.c3.ward =
  matrix(c(data.c3.ward$data.X1,data.c3.ward$data.X2,
    data.c3.ward$data.X3), ncol = 3)
x1.ward = sum ((xil.c1.ward-xlc1bar.ward)^2)
x2.ward = sum ((xil.c2.ward-xlc2bar.ward)^2)
x3.ward = sum ((xil.c3.ward-xlc3bar.ward)^2)
SSW.ward.c3 = sum (x1.ward, x2.ward, x3.ward)
SSB.ward.c3 = SST - SSW.ward.c3
R.square.ward.c3 = SSB.ward.c3/SST
Hasil.ward = c(SST, SSW.ward.c3, R.square.ward.c3)
Hasil = rbind(Hasil.ward)
Hasil3 = data.frame(hasil.kelompok.numerik)
print("Hasil Pengelompokan Hirarki")
print(hasil.kelompok.numerik)
print("-----")
print(" Untuk 3 kelompok yang terbentuk ")
print("Metode SST SSW R-square")
print(Hasil)
write.csv(Hasil3,file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
kelompok3 ward.csv")
write.csv(Hasil, file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
3 ward.csv")
}

```



```

        rep(mean(data.c2.ward$data.X3),b)), ncol = 3)
xlc3bar.ward = matrix(c(rep(mean(data.c3.ward$data.X1),c),
        rep(mean(data.c3.ward$data.X2),c),
        rep(mean(data.c3.ward$data.X3),c)), ncol = 3)
xlc4bar.ward = matrix(c(rep(mean(data.c4.ward$data.X1),d),
        rep(mean(data.c4.ward$data.X2),d),
        rep(mean(data.c4.ward$data.X3),d)), ncol = 3)
xil.c1.ward =
    matrix(c(data.c1.ward$data.X1, data.c1.ward$data.X2,
        data.c1.ward$data.X3), ncol = 3)
xil.c2.ward =
    matrix(c(data.c2.ward$data.X1, data.c2.ward$data.X2,
        data.c2.ward$data.X3), ncol = 3)
xil.c3.ward =
    matrix(c(data.c3.ward$data.X1, data.c3.ward$data.X2,
        data.c3.ward$data.X3), ncol = 3)
xil.c4.ward =
    matrix(c(data.c4.ward$data.X1, data.c4.ward$data.X2,
        data.c4.ward$data.X3), ncol = 3)
x1.ward = sum ((xil.c1.ward-xlc1bar.ward)^2)
x2.ward = sum ((xil.c2.ward-xlc2bar.ward)^2)
x3.ward = sum ((xil.c3.ward-xlc3bar.ward)^2)
x4.ward = sum ((xil.c4.ward-xlc4bar.ward)^2)
SSW.ward.c4 = sum (x1.ward, x2.ward, x3.ward, x4.ward)
SSB.ward.c4 = SST - SSW.ward.c4
R.square.ward.c4 = SSB.ward.c4/SST
Hasil.ward = c(SST, SSW.ward.c4, R.square.ward.c4)
Hasil = rbind(Hasil.ward)
Hasil4 = data.frame(hasil.kelompok.numerik)
print("Hasil Pengelompokan Hirarki")
print(hasil.kelompok.numerik)
print("-----")
print(" Untuk 4 kelompok yang terbentuk ")
print("Metode  SST  SSW  R-square")
print(Hasil)

```

```
write.csv(Hasil4,file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
kelompok4 ward.csv")
write.csv(Hasil, file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
4 ward.csv")
}
```

Lampiran 5 *Syntax* Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 5 dengan *Ward*

```

data<-read.csv("F:/Kuliah/Semester 8/TA/data SUSENAS
Excel/data (numerik).csv",sep="," ,header=T)
cluster.5ward = function(data){
  k = 5
  datanumerik = data.frame(data$X1, data$X2, data$X3)
  d = dist(datanumerik, method = "euclidean")
  # Analisis Kluster Hirarki
  fit.ward = hclust(d, method = "ward")
  # Memotong Dendogram untuk k Kluster
  ward = cutree(fit.ward, k=5)
  hasil.kelompok.numerik = data.frame(ward)
  # Menghitung sum of square total  $y = (x_i - \bar{x})^2$ 
  y = c((data$X1 - mean(data$X1))^2, (data$X2 -
mean(data$X2))^2, (data$X3 - mean(data$X3))^2)
  SST = sum (y)
  # Menghitung sum of square within group (ward)
  kelompok.ward = hasil.kelompok.numerik$ward
  data.ward = data.frame(ward, data$X1, data$X2, data$X3)
  data.ward.sort = data.ward[order(data.ward$ward),]
  a = length(data.ward.sort$ward[data.ward.sort$ward==1])
  b = length(data.ward.sort$ward[data.ward.sort$ward==2])
  c = length(data.ward.sort$ward[data.ward.sort$ward==3])
  d = length(data.ward.sort$ward[data.ward.sort$ward==4])
  e = length(data.ward.sort$ward[data.ward.sort$ward==5])
  data.c1.ward = data.ward.sort[1:a,]
  data.c2.ward = data.ward.sort[(a+1):(a+b),]
  data.c3.ward = data.ward.sort[(a+b+1):(a+b+c),]
  data.c4.ward = data.ward.sort[(a+b+c+1):(a+b+c+d),]
  data.c5.ward = data.ward.sort[(a+b+c+d+1):3097,]
  xlc1bar.ward = matrix(c(rep(mean(data.c1.ward$data.X1),a),
rep(mean(data.c1.ward$data.X2),a),
rep(mean(data.c1.ward$data.X3),a)), ncol = 3)

```

```

xlc2bar.ward = matrix(c(rep(mean(data.c2.ward$data.X1),b),
                        rep(mean(data.c2.ward$data.X2),b),
                        rep(mean(data.c2.ward$data.X3),b)), ncol = 3)
xlc3bar.ward = matrix(c(rep(mean(data.c3.ward$data.X1),c),
                        rep(mean(data.c3.ward$data.X2),c),
                        rep(mean(data.c3.ward$data.X3),c)), ncol = 3)
xlc4bar.ward = matrix(c(rep(mean(data.c4.ward$data.X1),d),
                        rep(mean(data.c4.ward$data.X2),d),
                        rep(mean(data.c4.ward$data.X3),d)), ncol = 3)
xlc5bar.ward = matrix(c(rep(mean(data.c5.ward$data.X1),e),
                        rep(mean(data.c5.ward$data.X2),e),
                        rep(mean(data.c5.ward$data.X3),e)), ncol = 3)
xil.c1.ward =
  matrix(c(data.c1.ward$data.X1, data.c1.ward$data.X2,
          data.c1.ward$data.X3), ncol = 3)
xil.c2.ward =
  matrix(c(data.c2.ward$data.X1, data.c2.ward$data.X2,
          data.c2.ward$data.X3), ncol = 3)
xil.c3.ward =
  matrix(c(data.c3.ward$data.X1, data.c3.ward$data.X2,
          data.c3.ward$data.X3), ncol = 3)
xil.c4.ward =
  matrix(c(data.c4.ward$data.X1, data.c4.ward$data.X2,
          data.c4.ward$data.X3), ncol = 3)
xil.c5.ward =
  matrix(c(data.c5.ward$data.X1, data.c5.ward$data.X2,
          data.c5.ward$data.X3), ncol = 3)
x1.ward = sum ((xil.c1.ward-xlc1bar.ward)^2)
x2.ward = sum ((xil.c2.ward-xlc2bar.ward)^2)
x3.ward = sum ((xil.c3.ward-xlc3bar.ward)^2)
x4.ward = sum ((xil.c4.ward-xlc4bar.ward)^2)
x5.ward = sum ((xil.c5.ward-xlc5bar.ward)^2)
SSW.ward.c5 = sum (x1.ward, x2.ward, x3.ward, x4.ward,
x5.ward)
SSB.ward.c5 = SST - SSW.ward.c5

```

```
R.square.ward.c5 = SSB.ward.c5/SST
Hasil.ward = c(SST, SSW.ward.c5, R.square.ward.c5)
Hasil = rbind(Hasil.ward)
Hasil5 = data.frame(hasil.kelompok.numerik)
print("Hasil Pengelompokan Hirarki")
print(hasil.kelompok.numerik)
print("-----")
print(" Untuk 5 kelompok yang terbentuk ")
print("Metode SST SSW R-square")
print(Hasil)
write.csv(Hasil5,file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
kelompok5 ward.csv")
write.csv(Hasil, file="F:/Kuliah/Semester 8/TA/Data Run/Hasil 5
ward.csv")
}
```

Lampiran 6 *Syntax* Analisis Pengelompokan Data Numerik untuk Jumlah Kelompok 6 dengan *Ward*

```

data<-read.csv("F:/Kuliah/Semester 8/TA/data SUSENAS
Excel/data (numerik).csv",sep="," ,header=T)
cluster.6ward = function(data){
  k = 6
  datanumerik = data.frame(data$X1, data$X2, data$X3)
  d = dist(datanumerik, method = "euclidean")
  fit.ward = hclust(d, method = "ward")
  # Memotong Dendogram untuk k Kluster
  ward = cutree(fit.ward, k=k)
  hasil.kelompok.numerik = data.frame(ward)
  # Menghitung sum of square total  $y = (x_i - \bar{x})^2$ 
  y = c((data$X1-mean(data$X1))^2,(data$X2-
mean(data$X2))^2,(data$X3-mean(data$X3))^2)
  SST = sum (y)
  # Menghitung sum of square within group (ward)
  kelompok.ward = hasil.kelompok.numerik$ward
  data.ward = data.frame(ward, data$X1, data$X2, data$X3)
  data.ward.sort = data.ward[order(data.ward$ward),]
  a = length(data.ward.sort$ward[data.ward.sort$ward==1])
  b = length(data.ward.sort$ward[data.ward.sort$ward==2])
  c = length(data.ward.sort$ward[data.ward.sort$ward==3])
  d = length(data.ward.sort$ward[data.ward.sort$ward==4])
  e = length(data.ward.sort$ward[data.ward.sort$ward==5])
  f = length(data.ward.sort$ward[data.ward.sort$ward==6])
  data.c1.ward = data.ward.sort[1:a,]
  data.c2.ward = data.ward.sort[(a+1):(a+b),]
  data.c3.ward = data.ward.sort[(a+b+1):(a+b+c),]
  data.c4.ward = data.ward.sort[(a+b+c+1):(a+b+c+d),]
  data.c5.ward = data.ward.sort[(a+b+c+d+1):(a+b+c+d+e),]

```



```

data.c6.ward = data.ward.sort[(a+b+c+d+e+1):3097,]
xlc1bar.ward = matrix(c(rep(mean(data.c1.ward$data.X1),a),
                        rep(mean(data.c1.ward$data.X2),a),
                        rep(mean(data.c1.ward$data.X3),a)), ncol = 3)
xlc2bar.ward = matrix(c(rep(mean(data.c2.ward$data.X1),b),
                        rep(mean(data.c2.ward$data.X2),b),
                        rep(mean(data.c2.ward$data.X3),b)), ncol = 3)
xlc3bar.ward = matrix(c(rep(mean(data.c3.ward$data.X1),c),
                        rep(mean(data.c3.ward$data.X2),c),
                        rep(mean(data.c3.ward$data.X3),c)), ncol = 3)
xlc4bar.ward = matrix(c(rep(mean(data.c4.ward$data.X1),d),
                        rep(mean(data.c4.ward$data.X2),d),
                        rep(mean(data.c4.ward$data.X3),d)), ncol = 3)
xlc5bar.ward = matrix(c(rep(mean(data.c5.ward$data.X1),e),
                        rep(mean(data.c5.ward$data.X2),e),
                        rep(mean(data.c5.ward$data.X3),e)), ncol = 3)
xlc6bar.ward = matrix(c(rep(mean(data.c6.ward$data.X1),f),
                        rep(mean(data.c6.ward$data.X2),f),
                        rep(mean(data.c6.ward$data.X3),f)), ncol = 3)

xil.c1.ward =
  matrix(c(data.c1.ward$data.X1, data.c1.ward$data.X2,
           data.c1.ward$data.X3), ncol = 3)
xil.c2.ward =
  matrix(c(data.c2.ward$data.X1, data.c2.ward$data.X2,
           data.c2.ward$data.X3), ncol = 3)
xil.c3.ward =
  matrix(c(data.c3.ward$data.X1, data.c3.ward$data.X2,
           data.c3.ward$data.X3), ncol = 3)
xil.c4.ward =
  matrix(c(data.c4.ward$data.X1, data.c4.ward$data.X2,
           data.c4.ward$data.X3), ncol = 3)
xil.c5.ward =
  matrix(c(data.c5.ward$data.X1, data.c5.ward$data.X2,

```

```

data.c5.ward$X3), ncol = 3)
xil.c6.ward =
  matrix(c(data.c6.ward$X1, data.c6.ward$X2,
    data.c6.ward$X3), ncol = 3)
x1.ward = sum ((xil.c1.ward-xlc1bar.ward)^2)
x2.ward = sum ((xil.c2.ward-xlc2bar.ward)^2)
x3.ward = sum ((xil.c3.ward-xlc3bar.ward)^2)
x4.ward = sum ((xil.c4.ward-xlc4bar.ward)^2)
x5.ward = sum ((xil.c5.ward-xlc5bar.ward)^2)
x6.ward = sum ((xil.c6.ward-xlc6bar.ward)^2)
SSW.ward.c6 = sum (x1.ward, x2.ward, x3.ward, x4.ward,
x5.ward, x6.ward)
SSB.ward.c6 = SST - SSW.ward.c6
R.square.ward.c6 = SSB.ward.c6/SST
Hasil.ward = c(SST, SSW.ward.c6, R.square.ward.c6)
Hasil = rbind(Hasil.ward)
Hasil5 = data.frame(hasil.kelompok.numerik)
print("Hasil Pengelompokan Hirarki")
print(hasil.kelompok.numerik)
print("-----")
print(" Untuk 6 kelompok yang terbentuk ")
print("Metode SST SSW R-square")
print(Hasil)
write.csv(Hasil5,file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
kelompok6 ward.csv")
write.csv(Hasil, file="F:/Kuliah/Semester 8/TA/Data Run/Hasil
6 ward.csv")
}

```

Lampiran 7 Syntax *Pseudo-F* untuk Pengelompokan Data Berskala Numerik

```
#Pseudo-F k2
data <- read.csv("F:/Kuliah/Semester 8/TA/Data Run/ward
                k2.csv")
pseudoF <- function(data, nc)
{
  n <- dim(data)[1]
  p <- dim(data)[2]
  X <- data[,1:(p-1)]
  Group <- data[,p]

  p <- dim(X)[2]
  Mean.X <- matrix(ncol = p, nrow = (nc+1))
  for (i in 1:nc)
  {
    for (j in 1:p)
    {
      Mean.X[i,j] <- mean(X[which(Group==i),j])
      Mean.X[(nc+1),j] <- mean(X[,j])
    }
  }

  SST <- matrix(ncol=p, nrow=n)
  for (i in 1:n)
  {
    for (j in 1:p)
    {
      SST[i,j] <- (X[i,j] - Mean.X[(nc+1),j])^2
    }
  }
  SST <- sum(sum(SST))

  SSE <- matrix(ncol=p, nrow=n)
  for (i in 1:n)
```

```

{
  for (j in 1:p)
  {
    for (k in 1:nc)
    {
      if (Group[i]==k)
      {
        SSE[i,j] <- (X[i,j] - Mean.X[k,j])^2
      }
    }
  }
}
SSE <- sum(sum(SSE))

Rsqr <- (SST-SSE)/SST
pseudoF <- (Rsqr/(2-1))/((1-Rsqr)/(n-2))
list(Rsqr=Rsqr, pseudoF=pseudoF)
}
pseudoF(data,3097)

#Pseudo-F k3
data <- read.csv("F:/Kuliah/Semester 8/TA/Data Run/ward
                 k3.csv")
pseudoF <- function(data, nc)
{
  n <- dim(data)[1]
  p <- dim(data)[2]
  X <- data[,1:(p-1)]
  Group <- data[,p]

  p <- dim(X)[2]
  Mean.X <- matrix(ncol = p, nrow = (nc+1))
  for (i in 1:nc)
  {
    for (j in 1:p)

```

```

{
  Mean.X[i,j] <- mean(X[which(Group==i),j])
  Mean.X[(nc+1),j] <- mean(X[,j])
}
}
SST <- matrix(ncol=p, nrow=n)
for (i in 1:n)
{
  for (j in 1:p)
  {
    SST[i,j] <- (X[i,j] - Mean.X[(nc+1),j])^2
  }
}
SST <- sum(sum(SST))

SSE <- matrix(ncol=p, nrow=n)
for (i in 1:n)
{
  for (j in 1:p)
  {
    for (k in 1:nc)
    {
      if (Group[i]==k)
      {
        SSE[i,j] <- (X[i,j] - Mean.X[k,j])^2
      }
    }
  }
}
SSE <- sum(sum(SSE))
Rsqr <- (SST-SSE)/SST
pseudoF <- (Rsqr/(3-1))/((1-Rsqr)/(n-3))
list(Rsqr=Rsqr, pseudoF=pseudoF)
}
pseudoF(data,3097)

```

```

#Pseudo-F k4
data <- read.csv("F:/Kuliah/Semester 8/TA/Data Run/ward
                k4.csv")
pseudoF <- function(data, nc)
{
  n <- dim(data)[1]
  p <- dim(data)[2]
  X <- data[,1:(p-1)]
  Group <- data[,p]

  p <- dim(X)[2]
  Mean.X <- matrix(ncol = p, nrow = (nc+1))
  for (i in 1:nc)
  {
    for (j in 1:p)
    {
      Mean.X[i,j] <- mean(X[which(Group==i),j])
      Mean.X[(nc+1),j] <- mean(X[,j])
    }
  }

  SST <- matrix(ncol=p, nrow=n)
  for (i in 1:n)
  {
    for (j in 1:p)
    {
      SST[i,j] <- (X[i,j] - Mean.X[(nc+1),j])^2
    }
  }
  SST <- sum(sum(SST))

  SSE <- matrix(ncol=p, nrow=n)
  for (i in 1:n)
  {
    for (j in 1:p)

```

```

{
  for (k in 1:nc)
  {
    if (Group[i]==k)
    {
      SSE[i,j] <- (X[i,j] - Mean.X[k,j])^2
    }
  }
}
}
SSE <- sum(sum(SSE))

Rsqr <- (SST-SSE)/SST
pseudoF <- (Rsqr/(4-1))/((1-Rsqr)/(n-4))
list(Rsqr=Rsqr, pseudoF=pseudoF)
}
pseudoF(data,3097)

#Pseudo-F k5
data <- read.csv("F:/Kuliah/Semester 8/TA/Data Run/ward
k5.csv")
pseudoF <- function(data, nc)
{
  n <- dim(data)[1]
  p <- dim(data)[2]
  X <- data[,1:(p-1)]
  Group <- data[,p]

  p <- dim(X)[2]
  Mean.X <- matrix(ncol = p, nrow = (nc+1))
  for (i in 1:nc)
  {
    for (j in 1:p)
    {
      Mean.X[i,j] <- mean(X[which(Group==i),j])
    }
  }
}

```

```

    Mean.X[(nc+1),j] <- mean(X[,j])
  }
}

SST <- matrix(ncol=p, nrow=n)
for (i in 1:n)
{
  for (j in 1:p)
  {
    SST[i,j] <- (X[i,j] - Mean.X[(nc+1),j])^2
  }
}
SST <- sum(sum(SST))

SSE <- matrix(ncol=p, nrow=n)
for (i in 1:n)
{
  for (j in 1:p)
  {
    for (k in 1:nc)
    {
      if (Group[i]==k)
      {
        SSE[i,j] <- (X[i,j] - Mean.X[k,j])^2
      }
    }
  }
}
SSE <- sum(sum(SSE))

Rsqr <- (SST-SSE)/SST
pseudoF <- (Rsqr/(5-1))/((1-Rsqr)/(n-5))
list(Rsqr=Rsqr, pseudoF=pseudoF)
}
pseudoF(data,3097)

```



```

#Pseudo-F k6
data <- read.csv("F:/Kuliah/Semester 8/TA/Data Run/ward
k6.csv")
pseudoF <- function(data, nc)
{
  n <- dim(data)[1]
  p <- dim(data)[2]
  X <- data[,1:(p-1)]
  Group <- data[,p]

  p <- dim(X)[2]
  Mean.X <- matrix(ncol = p, nrow = (nc+1))
  for (i in 1:nc)
  {
    for (j in 1:p)
    {
      Mean.X[i,j] <- mean(X[which(Group==i),j])
      Mean.X[(nc+1),j] <- mean(X[,j])
    }
  }

  SST <- matrix(ncol=p, nrow=n)
  for (i in 1:n)
  {
    for (j in 1:p)
    {
      SST[i,j] <- (X[i,j] - Mean.X[(nc+1),j])^2
    }
  }
  SST <- sum(sum(SST))

  SSE <- matrix(ncol=p, nrow=n)
  for (i in 1:n)
  {
    for (j in 1:p)

```

```
{
  for (k in 1:nc)
  {
    if (Group[i]==k)
    {
      SSE[i,j] <- (X[i,j] - Mean.X[k,j])^2
    }
  }
}
SSE <- sum(sum(SSE))

Rsqr <- (SST-SSE)/SST
pseudoF <- (Rsqr/(6-1))/((1-Rsqr)/(n-6))
list(Rsqr=Rsqr, pseudoF=pseudoF)
}
pseudoF(data,3097)
```

Lampiran 8 *Syntax* Analisis Pengelompokan Data Kategorik dengan *K-modes*

```
data <- read.csv("F:/Kuliah/Semester 8/TA/data SUSENAS  
Excel/data kategorik.csv")  
kmode<-kmodes(data, 3, iter.max=10, weighted=FALSE)  
kmode  
cluster.output<-cbind(data, kmode$cluster)  
write.csv(cluster.output, file = "F:/Kuliah/Semester 8/TA/Data  
Run/Hasil kmodes.csv", row.names = TRUE)
```

Lampiran 9 *Syntax* Analisis Pengelompokan Data Campuran dengan Ensembel SWFM

```
Num<-read.csv("F:/Kuliah/Semester 8/TA/Data Run/ ward
k2.csv", sep=",", header=T)
Kat<-read.csv("F:/Kuliah/Semester 8/TA/Data Run/data kmodes
k3.csv", sep=",", header=T)

sij <- function(x,y)
{
sij <- length(intersect(x,y))/length(union(x,y))
p <- length(x)-1
q <- length(y)-1
m <- max (p,q)
sm <- sum (sij/m)
return(sm)
}
data1 = c(Num[1,],Kat[1,])
data2 = c(Num[2,],Kat[2,])
data3 = c(Num[3,],Kat[3,])
data4 = c(Num[4,],Kat[4,])
data5 = c(Num[5,],Kat[5,])
data6 = c(Num[6,],Kat[6,])
data7 = c(Num[7,],Kat[7,])
data8 = c(Num[8,],Kat[8,])
data9 = c(Num[9,],Kat[9,])
data10 = c(Num[10,],Kat[10,])
data11 = c(Num[11,],Kat[11,])
data12 = c(Num[12,],Kat[12,])
data13 = c(Num[13,],Kat[13,])
data14 = c(Num[14,],Kat[14,])
data15 = c(Num[15,],Kat[15,])
:
data3087 = c(Num[3087,],Kat[3087,])
data3088 = c(Num[3088,],Kat[3088,])
```

```

data3089 = c(Num[3089,],Kat[3089,])
data3090 = c(Num[3090,],Kat[3090,])
data3091 = c(Num[3091,],Kat[3091,])
data3092 = c(Num[3092,],Kat[3092,])
data3093 = c(Num[3093,],Kat[3093,])
data3094 = c(Num[3094,],Kat[3094,])
data3095 = c(Num[3095,],Kat[3095,])
data3096 = c(Num[3096,],Kat[3096,])
data3097 = c(Num[3097,],Kat[3097,])
data=rbind(data1, data2, data3, data4, data5, data6, data7, data8,
data9, data10, data11, data12, data13, data14, data15, ...,
data3087, data3088, data3089, data3090, data3091, data3092,
data3093, data3094, data3095, data3096, data 3097)
sim <- matrix (1, nrow = 3097, ncol = 3097)
rownames(sim) <- c(1:3097)
colnames(sim) <- c(1:3097)
for (i in 1:(nrow(sim)-1))
{
for (j in (i+1):nrow(sim))
{
x <- data[i,]
y <- data[j,]
sim [i,j] <- sij(x,y)
sim [j,i] <- sim[i,j]
}
}
sim
wi <- 0.5
F <- wi*(1-sim)
jarak <- as.dist(F)
fit <- hclust(jarak,method = "ward")
hasil.k2 <- cutree(fit, k=2)
hasil.k3 <- cutree(fit, k=3)
hasil.k4 <- cutree(fit, k=4)
hasil.k5 <- cutree(fit, k=5)

```

```
hasil.k6 <- cutree(fit, k=6)
hasil.kelompok <-
data.frame(hasil.k2,hasil.k3,hasil.k4,hasil.k5,hasil.k6)
SWFM <- data.frame(hasil.kelompok)
write.csv(sim,file <- "F:/Kuliah/Semester 8/TA/Data Run/sim
ward3.csv")
write.csv(SWFM,file <- "F:/Kuliah/Semester 8/TA/Data Run/Data
Hasil SWFM ave.csv")
SWFM <- read.csv("F:/Kuliah/Semester 8/TA/Data Run/Data
Hasil SWFM ave.csv", header=TRUE,sep=",")
SWFM
```

Lampiran 10 Syntax Rasio Sw dan Sb untuk Ensembel SWFM

```

model2 <- read.table("F:/Kuliah/Semester 8/TA/Data Run/k2
ward3.csv",sep="," ,header=TRUE)
model3 <- read.table("F:/Kuliah/Semester 8/TA/Data Run/k3
ward3.csv",sep="," ,header=TRUE)
model4 <- read.table("F:/Kuliah/Semester 8/TA/Data Run/k4
ward3.csv",sep="," ,header=TRUE)
model5 <- read.table("F:/Kuliah/Semester 8/TA/Data Run/k5
ward3.csv",sep="," ,header=TRUE)
model6 <- read.table("F:/Kuliah/Semester 8/TA/Data Run/k6
ward3.csv",sep="," ,header=TRUE)

model.2 <- aov(c~numerik+kategori, data=model2)
model.3 <- aov(c~numerik+kategori, data=model3)
model.4 <- aov(c~numerik+kategori, data=model4)
model.5 <- aov(c~numerik+kategori, data=model5)
model.6 <- aov(c~numerik+kategori, data=model6)

SSW.2 = sum((model.2$residuals)^2)
SSW.3 = sum((model.3$residuals)^2)
SSW.4 = sum((model.4$residuals)^2)
SSW.5 = sum((model.5$residuals)^2)
SSW.6 = sum((model.6$residuals)^2)

summary.2 = c(summary(model.2))
summary.3 = c(summary(model.3))
summary.4 = c(summary(model.4))
summary.5 = c(summary(model.5))
summary.6 = c(summary(model.6))

summary.22=matrix(unlist(summary.2),3,5)
summary.33=matrix(unlist(summary.3),3,5)
summary.44=matrix(unlist(summary.4),3,5)
summary.55=matrix(unlist(summary.5),3,5)

```

```

summary.66=matrix(unlist(summary.6),3,5)

SSB.2 = sum(summary.22[1:2,3])
SSB.3 = sum(summary.33[1:2,3])
SSB.4 = sum(summary.44[1:2,3])
SSB.5 = sum(summary.55[1:2,3])
SSB.6 = sum(summary.66[1:2,3])

SW.2 = sqrt(SSW.2/(3095-2))
SW.3 = sqrt(SSW.3/(3095-3))
SW.4 = sqrt(SSW.4/(3095-4))
SW.5 = sqrt(SSW.5/(3095-5))
SW.6 = sqrt(SSW.6/(3095-6))

SB.2 = sqrt(SSB.2/(2-1))
SB.3 = sqrt(SSB.3/(3-1))
SB.4 = sqrt(SSB.4/(4-1))
SB.5 = sqrt(SSB.5/(5-1))
SB.6 = sqrt(SSB.6/(6-1))

Ratio.2 = SW.2 / SB.2
Ratio.3 = SW.3 / SB.3
Ratio.4 = SW.4 / SB.4
Ratio.5 = SW.5 / SB.5
Ratio.6 = SW.6 / SB.6

Ratio.SWFM = rbind (Ratio.2, Ratio.3, Ratio.4, Ratio.5, Ratio.6)
u = c(2,3,4,5,6)
z = Ratio.SWFM
plot(u, z, main = " Plot Nilai Ratio SWFM",
xlab = "Jumlah Kelompok", ylab = "Nilai Ratio")
lines(u,z,col = "Blue")
print("Hasil Pengelompokan SWFM")
print(hasil.kelompok)
print("-----")

```



```
print("Nilai Ratio untuk Setiap Nilai K")
print(Ratio.SWFM)
write.csv(Ratio.SWFM,file="F:/Kuliah/Semester 8/TA/Data
Run/Rasio SWFM3 ward.csv")
write.csv(hasil.kelompok,file="F:/Kuliah/Semester 8/TA/Data
Run/Hasil Kelompok SWFM3 ward.csv")
```

Lampiran 11 *Output* untuk Analisis Kelompok pada Data Numerik

NO	No. urut RT	KAB	ward.k2	ward.k3	ward.k4	ward.k5	ward.k6
1	352621020	26	1	1	1	1	1
2	352621021	26	1	1	1	1	2
3	352621022	26	1	2	2	2	3
4	352621023	26	1	1	1	1	1
5	352621024	26	1	1	1	1	2
6	352621025	26	1	1	1	1	2
7	352621026	26	2	3	3	3	4
8	352621027	26	2	3	3	3	4
9	352621028	26	1	1	1	1	1
10	352621029	26	2	3	4	4	5
...
3090	352924109	29	1	2	2	2	3
3091	352924110	29	2	3	4	4	5
3092	352924111	29	1	2	2	5	6
3093	352924112	29	1	2	2	2	3
3094	352924113	29	1	2	2	2	3
3095	352924114	29	1	2	2	2	3
3096	352924115	29	1	2	2	2	3
3097	352924116	29	1	1	1	1	2

Lampiran 12 Output untuk Analisis Kelompok pada Data Kategorik dengan Jumlah Kelompok 3

NO	No. urut RT	KAB	X4	X5	X6	X7	X8	X9	Kmodes
1	352621020	26	2	2	2	2	1	2	1
2	352621021	26	2	2	2	2	1	1	1
3	352621022	26	2	2	2	2	1	1	1
4	352621023	26	2	2	2	2	1	1	1
5	352621024	26	2	2	2	2	1	1	1
6	352621025	26	1	2	2	2	1	1	1
7	352621026	26	1	2	2	2	1	1	1
8	352621027	26	1	2	2	2	1	1	1
9	352621028	26	3	2	2	2	1	1	1
10	352621029	26	2	2	2	2	1	1	1
11	352621030	26	2	2	2	2	1	1	1
12	352621031	26	2	2	2	2	1	1	1
13	352621032	26	2	2	2	2	1	1	1
14	352621033	26	2	2	2	2	1	1	1
15	352621034	26	3	2	2	2	1	1	1
...
3090	352924109	29	2	2	2	2	2	2	2
3091	352924110	29	1	2	2	2	2	1	1
3092	352924111	29	2	2	2	2	2	2	2
3093	352924112	29	2	2	2	2	2	2	2
3094	352924113	29	2	2	2	2	2	2	2
3095	352924114	29	1	2	2	2	2	2	1
3096	352924115	29	2	2	2	2	2	2	2
3097	352924116	29	3	2	2	2	2	2	2

Lampiran 13 Output untuk Analisis Kelompok pada Data Numerik dan Data Kategorik dan Rasio Sw dan Sb dengan Ensembel SWFM

NO	No. urut RT	KAB	hasil.k2	hasil.k3	hasil.k4	hasil.k5	hasil.k6
1	352621020	26	1	1	1	1	1
2	352621021	26	1	1	1	1	1
3	352621022	26	1	1	1	1	1
4	352621023	26	1	1	1	1	1
5	352621024	26	1	1	1	1	1
6	352621025	26	1	1	1	1	1
7	352621026	26	2	2	2	2	2
8	352621027	26	2	2	2	2	2
9	352621028	26	1	1	1	1	1
10	352621029	26	2	2	2	2	2
11	352621030	26	1	1	1	1	1
12	352621031	26	1	1	1	1	1
13	352621032	26	1	1	1	1	1
14	352621033	26	2	2	2	2	2
...
3090	352924109	29	2	2	2	2	2
3091	352924110	29	2	2	2	2	2
3092	352924111	29	2	2	2	2	2
3093	352924112	29	2	2	2	2	2
3094	352924113	29	2	2	2	2	2
3095	352924114	29	1	1	1	1	1
3096	352924115	29	2	2	2	2	2
3097	352924116	29	2	2	2	2	2

	2	3	4	5	6
Rasio Sw dan Sb	0,014908	0,015859	0,011841	0,006627	0,007455

Lampiran 14 Output *Pseudo-F* masing-masing Kelompok

```
#PseudoF k2
> pseudoF(data,3097)
$Rsq
[1] 0.2444165

$pseudoF
[1] 1001.172

#PseudoF k3
> pseudoF(data,3097)
$Rsq
[1] 0.2444165

$pseudoF
[1] 500.4242

#PseudoF k4
> pseudoF(data,3097)
$Rsq
[1] 0.2444165

$pseudoF
[1] 333.5083

#PseudoF k5
> pseudoF(data,3097)
$Rsq
[1] 0.2444165

$pseudoF
[1] 250.0504
```

```
#PseudoF k6  
> pseudoF(data,3097)  
$Rsq  
[1] 0.2444165  
  
$pseudoF  
[1] 199.9756
```

Lampiran 15 Output uji *Kruskal Wallis* dan Uji Normalitas

Test Statistics^{a,b}

	Total_Pengeluaran_RT	Proporsi_ART_Perempuan	Jumlah_ART
Chi-Square	1408.271	60.302	442.726
df	4	4	4
Asymp. Sig.	.000	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: Cluster

Uji Normal Multivariat

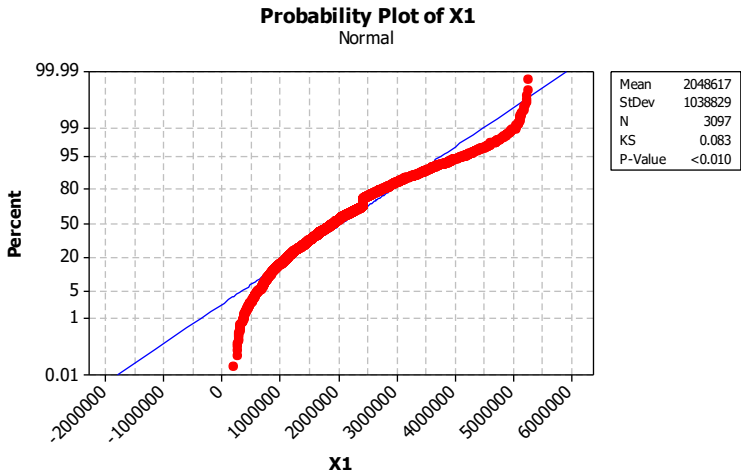
mshapiro.test(X)

Shapiro-Wilk normality test

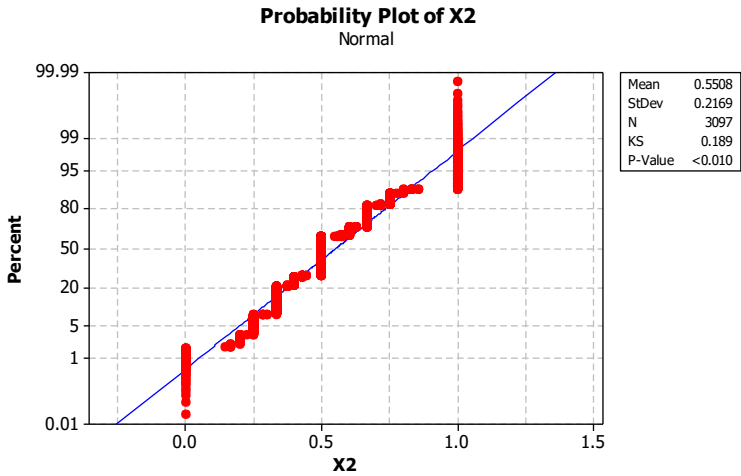
data: Z

W = 0.9639, p-value < 2.2e-16

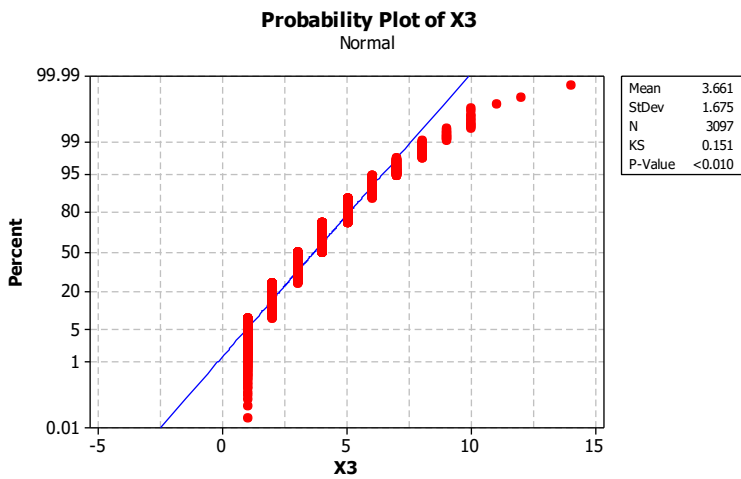
Uji Normalitas (Total Pengeluaran RT)



Uji Normalitas (Proporsi ART Perempuan)



Uji Normalitas (Jumlah ART)



Lampiran 16 Surat Keterangan Pengambilan Data



**BADAN PUSAT STATISTIK
PROVINSI JAWA TIMUR**



SURAT KETERANGAN

Yang bertanda tangan dibawah ini :

N a m a : Thomas Wunang Tjahjo, M.Sc, M.Eng.
N I P : 19700329 1992 11 1 001
Jabatan : Kepala Bidang Integrasi Pengolahan dan Diseminasi Statistik

Dengan ini menerangkan bahwa :

N a m a : Astarani Wili Martha
Fakultas/Program Studi : Fakultas Matematika, Komputasi Dan Sains Data / Statistika
N.R.P : 06211440000107
Alamat Rumah : Griya Mapan Sentosa CE 38 Waru, Sidoarjo
Akademi / Universitas : Institut Teknologi Sepuluh Nopember (ITS)
Telp (031) 594 3352, (031) 599 4251-55
Fax (031) 592 2940

Di berikan kesempatan menggunakan data Badan Pusat Statistik (BPS) Provinsi Jawa Timur, dengan syarat menyebut judul publikasi dan sumbernya serta tidak untuk tujuan komersil. Data ini digunakan dalam rangka menyusun Tugas Akhir / Skripsi / Thesis / Disertasi dengan judul :

" Metode Similarity Weight And Filter Method Untuk Pengelompokan Rumah Tangga di Pulau Madura Berdasarkan Faktor-Faktor Yang mempengaruhi Kebutuhan Air Bersih "

Demikian surat keterangan ini dibuat dan agar dipergunakan sebagaimana mestinya



Surabaya, 6 Juni 2018
An. Kepala BPS Provinsi Jawa Timur
Kepala Bidang JPDS

Thomas Wunang Tjahjo, M.Sc, M.Eng.

BIODATA PENULIS



Penulis dengan nama lengkap Astarani Wili Martha dengan nama panggilan Vivi lahir di Surabaya, 9 Maret 1995. Penulis menempuh pendidikan formal di SD Al-Muslim Sidoarjo, SMP Al-Muslim Sidoarjo dan SMA Dharma Wanita Surabaya. Kemudian penulis diterima sebagai mahasiswa Departemen Statistika ITS melalui jalur Program Kemitraan Mandiri (PKM) pada tahun 2014. Selama masa perkuliahan, penulis

aktif di berbagai kepanitiaan salah satunya sebagai *Liaison Office* (LO) di acara Pekan Raya Statistika (PRS) tahun 2016. Selain itu, penulis juga aktif dalam organisasi yang menaungi HIMASTA-ITS yaitu *Statistics Computer Course* (SCC) HIMASTA-ITS sebagai staff *Public Relation* (PR) dan juga penulis tergabung dalam UKM ITS *Foreign Language Society* (IFLS) sebagai staff Kewirausahaan dan tim pengajar Bahasa Korea. Penulis juga pernah diberi kesempatan untuk kerja praktik atau magang di Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Jawa Timur. Apabila pembaca ingin memberi kritik dan saran kepada penulis serta diskusi lebih mengenai Tugas Akhir ini, maka dapat menghubungi penulis melalui email astarani.wili@gmail.com atau nomor telepon 081232179733.