



**TUGAS AKHIR - SS141501**

**PENGELOMPOKAN KABUPATEN/KOTA  
DI JAWA TIMUR BERDASARKAN  
FAKTOR-FAKTOR KEBUTUHAN  
KONSUMSI AIR BERSIH RUMAH TANGGA  
MENGUNAKAN METODE ENSEMBEL ROCK**

**VINDA FERDHIANI  
NRP 062114 4000 0069**

**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 - SS141501**

**PENGELOMPOKAN KABUPATEN/KOTA  
DI JAWA TIMUR BERDASARKAN  
FAKTOR-FAKTOR KEBUTUHAN  
KONSUMSI AIR BERSIH RUMAH TANGGA  
MENGUNAKAN METODE ENSEMBEL ROCK**

**VINDA FERDHIANI  
NRP 062114 4000 0069**

**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**





**FINAL PROJECT - SS 141501**

**DISTRICTS/CITIES CLASSIFICATION  
IN EAST JAVA BASED ON FACTORS  
RELATED TO THE HOUSEHOLD  
WATER CONSUMPTION USING  
ROCK ENSEMBLE METHOD**

**VINDA FERDHIANI  
SN 062114 4000 0069**

**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**



**LEMBAR PENGESAHAN**

**PENGELOMPOKAN KABUPATEN/KOTA DI JAWA TIMUR  
BERDASARKAN FAKTOR-FAKTOR KEBUTUHAN  
KONSUMSI AIR BERSIH RUMAH TANGGA  
MENGUNAKAN METODE ENSEMBEL ROCK**

**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 :

**Vinda Ferdhiani**

NRP. 062114 4000 0069

Disetujui oleh Pembimbing :

**Dr. Dra. Ismaini Zain, M.Si**  
NIP. 19600525 198803 2 001



Mengetahui,  
Kepala Departemen

**Dr. Suhartono**  
NIP. 19770929 199512 1 001

SURABAYA, JULI 2018

*(Halaman ini sengaja dikosongkan)*

**PENGELOMPOKAN KABUPATEN/KOTA  
DI JAWA TIMUR BERDASARKAN FAKTOR-FAKTOR  
KEBUTUHAN KONSUMSI AIR BERSIH RUMAH TANGGA  
MENGUNAKAN METODE ENSEMBEL ROCK**

**Nama Mahasiswa : Vinda Ferdhiani**  
**NRP : 062114 4000 0069**  
**Departemen : Statistika**  
**Dosen Pembimbing : Dr. Dra. Ismaini Zain, M.Si**

**Abstrak**

*Akses penduduk terhadap air bersih dan sanitasi merupakan salah satu indikator SDGs yang akan dicapai pada tahun 2030 termasuk di Provinsi Jawa Timur. Indikator terhadap air bersih dapat dilihat dari faktor-faktor yang mempengaruhi kebutuhan konsumsi air bersih pada suatu wilayah. Terdapat dua skala dalam data yaitu numerik dan kategorik. Metode yang digunakan untuk data numerik dan kategorik adalah metode Ensemble ROCK. Unit penelitian ini adalah kabupaten atau kota. Penelitian dengan variabel numerik menggunakan metode hirarki Agglomerative serta untuk variabel kategorik menggunakan metode ROCK. Hasil analisis pengelompokan data numerik diperoleh pengelompokan sebanyak 5 kelompok terdiri dari 25 kabupaten dalam cluster pertama, 1 kabupaten pada cluster dua, 2 kabupaten dalam cluster ketiga, 9 kabupaten/kota dalam cluster keempat, dan 1 kota dalam cluster 5. Untuk analisis pengelompokan data kategorik diperoleh hasil pengelompokan sebanyak 2 cluster terdiri dari 3 kota dalam cluster pertama dan 35 kabupaten/kota dalam cluster dua. Penggabungan hasil numerik dan kategorik diperoleh nilai rasio terkecil yaitu  $4.82 \times 10^{-11}$  dengan hasil terbentuk 2 cluster yaitu 28 kabupaten masuk dalam cluster pertama dan 10 kabupaten/kota masuk dalam cluster 2.*

***Kata Kunci : Air Bersih, Analisis Klaster, Ensembel ROCK, Hirarki, SDGs***

*(Halaman ini sengaja dikosongkan)*

**DISTRICTS/CITIES CLASSIFICATION IN EAST JAVA  
BASED ON FACTORS RELATED TO THE HOUSEHOLD  
CLEAN WATER CONSUMPTION USING  
ROCK ENSEMBLE METHOD**

**Name** : Vinda Ferdhiani  
**SN** : 062114 4000 0069  
**Department** : Statistics  
**Supervisor** : Dr. Dra. Ismaini Zain, M.Si

**Abstract**

*Population access to clean water and sanitation is one indicator of SDGs to be achieved by 2030 including in East Java Province. Indicators of clean water can be seen from the factors that affect the needs of water consumption in a region. There are two scales in the data namely numeric and categorical. The method used for numerical and categorical data is the ROCK Ensemble method. This research unit is a district or city. Research with numerical variables using Agglomerative hierarchy method as well as for categorical variables using ROCK method. The results of numerical data grouping analysis were obtained by grouping as many as 5 groups consisting of 25 districts in the first cluster, 1 districts in cluster two, 2 districts in the third cluster, 9 districts / cities in the fourth cluster, and 1 city in cluster 5. For cluster analysis the categorical data obtained by clustering results of 2 clusters consisting of 3 cities in the first cluster and 35 districts / cities in cluster two. Cluster ensemble of numerical and categorical results obtained the smallest ratio value of  $4.82 \times 10^{-11}$  with the result of 2 clusters formed that is 28 regencies entered in the first cluster and 10 districts / cities entered in cluster 2.*

**Keywords** : *Clean Water, Cluster Analysis, Ensemble ROCK, Hierarchical, SDGs*

*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul **“Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Faktor-Faktor Kebutuhan Konsumsi Air Bersih Rumah Tangga Menggunakan Metode Ensembl ROCK”** dengan baik. Penyusunan Tugas Akhir ini tidak lepas dari banyaknya bantuan dan dukungan yang diberikan dari berbagai pihak. Penulis mengucapkan terima kasih kepada:

1. Orangtua dan seluruh keluarga tercinta yang selalu memberikan doa, nasihat, semangat, dan kasih sayang yang diberikan sehingga dapat menjadi penyemangat bagi penulis.
2. Bapak Dr. Suhartono selaku Kepala Departemen Statistika dan Bapak Dr. Sutikno, M.Si selaku Koordinator Program Studi Sarjana yang telah memberikan fasilitas untuk kelancaran penyelesaian Tugas Akhir.
3. Ibu Dr. Dra. Ismaini Zain, M.Si selaku dosen pembimbing yang dengan sabar telah memberikan bimbingan, ilmu, saran, dan dukungan selama penyusunan Tugas Akhir.
4. Bapak Prof. Drs. Nur Iriawan, M.Ikom, Ph.D dan Ibu Dr. Vita Ratnasari, S.Si, M.Si selaku dosen penguji yang telah memberikan banyak bantuan dan saran untuk Tugas Akhir ini.
5. Bapak Dr. Bambang Widjanarko Otok, S.Si., M.Si selaku dosen wali yang telah memberikan nasehat dan semangat selama masa perkuliahan.
6. Teman-teman terdekat Iqbal, Naila, Tiara, dan Fari serta teman-teman Respect 2014
7. Teman-teman seperjuangan Tugas Akhir “Pejuang Bu Is” yang saling mendukung dalam pembuatan Tugas Akhir
8. Semua pihak yang mendukung dalam penyelesaian Tugas Akhir

Penulis berharap hasil Tugas Akhir ini dapat bermanfaat bagi semua pihak yang terkait.

Surabaya, Juli 2018

Penulis

*(Halaman ini sengaja dikosongkan)*

## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b> .....	i
<b>LEMBAR PENGESAHAN</b> .....	v
<b>ABSTRAK</b> .....	vii
<b>ABSTRACT</b> .....	ix
<b>KATA PENGANTAR</b> .....	xi
<b>DAFTAR ISI</b> .....	xiii
<b>DAFTAR GAMBAR</b> .....	xv
<b>DAFTAR TABEL</b> .....	xvii
<b>DAFTAR LAMPIRAN</b> .....	xix
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
<b>BAB II TINJAUAN PUSTAKA</b> .....	5
2.1 Statistika Deskriptif.....	5
2.2 Data Numerik dan Data Kategorik.....	5
2.3 Analisis Klaster.....	6
2.3.1. Pengelompokan Data Numerik.....	7
2.3.2. Pengelompokan Data Kategorik.....	10
2.4 Pengelompokan Ensemble.....	13
2.5 Kinerja Hasil Pengelompokan.....	15

2.6	Uji <i>Kruskal-Wallis</i> .....	17
2.7	Faktor yang Mempengaruhi Pemenuhan Konsumsi Air Bersih.....	17
<b>BAB III METODOLOGI PENELITIAN</b> .....		21
3.1	Sumber Data .....	21
3.2	Variabel Penelitian.....	21
3.3	Langkah Analisis .....	26
<b>BAB IV ANALISIS DAN PEMBAHASAN</b> .....		31
4.1	Karakteristik Kabupaten atau Kota di Jawa Timur.....	31
4.2	Pengelompokkan Kabupaten atau Kota di Jawa Timur Menggunakan Metode <i>Ensemble ROCK</i> .....	41
4.2.1.	Pengelompokkan Variabel Penelitian Berskala Numerik .....	41
4.2.2.	Pengelompokkan Variabel Penelitian Berskala Kategorik.....	44
4.2.3.	Pengelompokkan Variabel Penelitian Berskala Numerik dan Kategorik Menggunakan Metode <i>Ensemble ROCK</i> .....	46
<b>BAB V KESIMPULAN DAN SARAN</b> .....		53
5.1.	Kesimpulan .....	53
5.2.	Saran .....	54
<b>DAFTAR PUSTAKA</b> .....		57
<b>LAMPIRAN</b> .....		61

## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b> Tahapan Analisis dengan Algoritma CEBMDC ....	13
<b>Gambar 2.2</b> Faktor-Faktor yang Mempengaruhi Kebutuhan Konsumsi Air Rumah Tangga .....	18
<b>Gambar 3.1</b> Diagram Alir Penelitian.....	29
<b>Gambar 4.1</b> Karakteristik Variabel Data Numerik.....	33
<b>Gambar 4.2</b> Kepadatan Penduduk.....	33
<b>Gambar 4.3</b> Persentase Tingkat Pendidikan Terakhir Kepala Rumah Tangga.....	34
<b>Gambar 4.4</b> Rata-rata Pengeluaran Sebulan Terakhir.....	35
<b>Gambar 4.5</b> Sumber Air Minum Utama.....	36
<b>Gambar 4.6</b> Kategori Baru Sumber Air Minum Utama.....	36
<b>Gambar 4.7</b> Sumber Air Masak Utama.....	37
<b>Gambar 4.8</b> Kategori Baru Sumber Air Masak Utama.....	38
<b>Gambar 4.9</b> Sumber Air Mandi/Cuci Utama.....	39
<b>Gambar 4.10</b> Kategori Baru Sumber Air Mandi/Cuci Utama....	40
<b>Gambar 4.11</b> Penggunaan Air PAM.....	40
<b>Gambar 4.12</b> Karakteristik Hasil Pengelompokan Data Numerik pada Hasil Pengelompokan <i>Ensemble</i> ROCK	49

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

	Halaman
<b>Tabel 3.1</b> Variabel Penelitian .....	21
<b>Tabel 4.1</b> Deskripsi Variabel Penelitian Berskala Numerik .....	31
<b>Tabel 4.2</b> Nilai <i>PseudoF</i> Hasil Pengelompokan Data Numerik	41
<b>Tabel 4.3</b> Nilai Rasio Pengelompokkan Variabel Penelitian Berskala Numerik.....	42
<b>Tabel 4.4</b> Hasil Pengelompokkan Variabel Penelitian Berskala Numerik.....	43
<b>Tabel 4.5</b> Karakteristik Hasil Pengelompokkan Variabel Berskala Numerik.....	43
<b>Tabel 4.6</b> Nilai Rasio Pengelompokkan Data Kategorik .....	45
<b>Tabel 4.7</b> Hasil Pengelompokkan Variabel Penelitian Berskala Kategorik.....	45
<b>Tabel 4.8</b> Karakteristik Hasil Pengelompokkan Variabel Penelitian Berskala Kategorik .....	46
<b>Tabel 4.9</b> Struktur Data dengan Variabel Baru.....	47
<b>Tabel 4.10</b> Rasio $S'_w$ dan $S'_B$ Pengelompokan Metode Ensemble ROCK.....	48
<b>Tabel 4.11</b> Hasil Pengelompokkan dengan Metode <i>Ensemble</i> ROCK.....	48
<b>Tabel 4.12</b> Karakteristik Hasil Pengelompokkan Data Numerik pada Metode <i>Ensemble</i> ROCK.....	49
<b>Tabel 4.13</b> Karakteristik Hasil Pengelompokkan Data Kategorik pada Metode <i>Ensemble</i> ROCK.....	50
<b>Tabel 4.14</b> Hasil Pengujian <i>Kruskal Wallis</i> .....	51

*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1</b> Data Penelitian .....	61
<b>Lampiran 2</b> <i>Syntax</i> Analisis Pengelompokan Data Numerik.....	63
<b>Lampiran 3</b> <i>Syntax</i> Nilai Rasio Analisis Kelompok Hirarki <i>Agglomerative</i> .....	64
<b>Lampiran 4</b> <i>Syntax</i> Analisis Metode ROCK untuk Pengelompokan Data Kategorik .....	71
<b>Lampiran 5</b> <i>Syntax</i> Analisis Pengelompokan Ensemble ROCK .	76
<b>Lampiran 6</b> <i>Output</i> untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 2 .....	79
<b>Lampiran 7</b> <i>Output</i> untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 3 .....	81
<b>Lampiran 8</b> <i>Output</i> untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 4 .....	83
<b>Lampiran 9</b> <i>Output</i> untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 5 .....	85
<b>Lampiran 10</b> <i>Output</i> untuk Pengelompokan Data Kategorik Menggunakan Metode ROCK .....	87
<b>Lampiran 11</b> <i>Output</i> untuk Pengelompokan Data Campuran Menggunakan Metode <i>Ensemble</i> ROCK.....	91
<b>Lampiran 12</b> Surat Keterangan Pengambilan Data .....	94

*(Halaman ini sengaja dikosongkan)*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Air bersih dibutuhkan dalam pemenuhan kebutuhan manusia untuk melakukan segala kegiatan sehingga perlu diketahui bagaimana air dikatakan bersih dari segi kualitas dan bisa digunakan dalam jumlah yang memadai dalam kegiatan sehari-hari manusia. Banyaknya populasi manusia yang semakin meningkat menyebabkan kebutuhan manusia terhadap air semakin meningkat pula. Agar kelangsungan hidup manusia dapat berjalan lancar, air bersih juga harus tersedia dalam jumlah yang memadai sesuai dengan aktifitas manusia pada tempat tertentu dan kurun waktu tertentu (Gabriel, 2001).

SDGs atau *Sustainable Development Goals* adalah sebuah kesepakatan pembangunan baru pengganti MDGs (*Millenium Development Goals*) yang berlaku pada tahun 2015 sampai 2030 mendatang. Akses penduduk terhadap air bersih dan sanitasi merupakan salah satu indikator dari SDGs. Target yang akan dicapai adalah setiap penduduk dapat mengakses air bersih. Oleh karena itu diperlukan suatu studi yang dapat mempersiapkan pencapaian target SDGs khususnya pada bidang akses terhadap air minum. Hasil SUSENAS (Survei Sosial Ekonomi Nasional) tahun 2016 menunjukkan bahwa terdapat rumah tangga dengan pengeluaran nol atau tanpa mengeluarkan biaya konsumsi air bersih. Apabila akses terhadap kebutuhan air bersih tidak dapat terpenuhi maka target akan sulit dipenuhi.

Jawa Timur merupakan salah satu provinsi dengan jumlah penduduk yang banyak yang terdiri dari 38 kabupaten/kota. Berdasarkan data sementara yang dihimpun Pusat Pengendali Operasi Badan Nasional Penanggulangan Bencana (BNPB), Jawa Timur mengalami kekeringan yang melanda 588 desa di 171 kecamatan dan 23 kabupaten/kota. Permasalahan kebutuhan air bersih yang muncul yaitu terjadinya penyebaran air yang tidak merata pada setiap daerah pada kabupaten/kota. Permasalahan

lainnya yaitu kurang tercukupi konsumsi air bersih setiap rumah tangga pada setiap kabupaten/kota berdasarkan faktor-faktor yang mempengaruhi rumah tangga untuk konsumsi air bersih. Pesatnya pertumbuhan ekonomi dan urbanisasi saat ini telah menyebabkan semakin tingginya konflik di antara berbagai penggunaan air. Menurut data BPS (Badan Pusat Statistik) Provinsi Jawa Timur, Pengeluaran konsumsi rumahtangga di Jawa Timur dari tahun 2013 sampai tahun 2016 mengalami peningkatan pada tiap tahunnya. Pengeluaran seseorang untuk konsumsi dan tabungan dipengaruhi oleh pendapatannya. Semakin besar pendapatan seseorang maka akan semakin banyak tingkat konsumsinya pula (Keynes, 1936). Hal tersebut dapat mempengaruhi konsumsi masyarakat di Jawa Timur dalam memperoleh air bersih.

Oleh karena itu untuk mengetahui wilayah kabupaten/kota di Jawa Timur yang memiliki ciri dengan telah tercukupinya kebutuhan air bersih maupun belum maka dilakukan pengelompokan kabupaten/kota di Jawa Timur. Terdapat beberapa faktor yang dipertimbangkan dalam pengelompokan kabupaten atau kota. Metode yang akan digunakan dalam pengelompokan penelitian ini yaitu metode Ensemble ROCK. Metode ensemble adalah teknik pengelompokan untuk menggabungkan hasil pengelompokan dari beberapa algoritma pengelompokan untuk mendapatkan kelompok yang lebih baik (He, Xu, & Deng, 2005).

Metode ensemble ROCK pernah dilakukan oleh Alvionita (2017) yaitu tentang metode ensemble ROCK dan SWFM untuk pengelompokan data campuran numerik dan kategorik pada kasus akses jeruk dan diperoleh pengelompokan terbaik adalah metode ensemble ROCK. Metode ensemble juga pernah dilakukan oleh Ramdhany (2017) yaitu tentang metode Ensemble ROCK dan *two step cluster* pada kasus pengelompokan desa di Kabupaten Bondowoso dan diperoleh pengelompokan terbaik dengan metode ensemble ROCK.

Berdasarkan uraian tersebut maka dilakukan penelitian untuk pengelompokan kabupaten/kota di Jawa Timur

menggunakan metode Ensembel ROCK untuk mengetahui karakteristik pada masing-masing daerah untuk memenuhi kebutuhan air bersih. Serta diharapkan mendapatkan hasil pengelompokan yang terbaik berdasarkan konsumsi air bersih pada kabupaten/kota di Jawa Timur.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dijelaskan, maka permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana karakteristik kabupaten/kota di Jawa Timur berdasarkan kebutuhan konsumsi air bersih. Setelah diketahui karakteristiknya maka selanjutnya akan dibahas bagaimana pengelompokan kabupaten/kota di Jawa Timur dengan menggunakan metode Ensembel ROCK.

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah tersebut maka tujuan yang ingin dicapai dari penelitian adalah sebagai berikut.

1. Mendeskripsikan karakteristik kabupaten/kota di Jawa Timur berdasarkan kebutuhan konsumsi air bersih
2. Menentukan kelompok kabupaten/kota di Jawa Timur dengan menggunakan metode Ensembel ROCK

## **1.4 Manfaat Penelitian**

Manfaat yang ingin diperoleh dari penelitian ini adalah mampu memberikan hasil pengelompokan kabupaten/kota di Jawa Timur dalam mencukupi kebutuhan air bersih atau air minum. Selain itu dapat memberikan informasi dan evaluasi bagi pihak pemerintah Indonesia mengenai pemenuhan kebutuhan air bersih bagi masyarakat Jawa Timur sehingga dapat dijadikan kebijakan atau perbaikan dalam meningkatkan kualitas air bersih di provinsi Jawa Timur. Manfaat bagi peneliti yaitu dapat mengaplikasikan ilmu statistika dalam melakukan analisis pengelompokan.

## **1.5 Batasan Masalah**

Dalam penelitian ini dilakukan pengelompokan kabupaten/kota berdasarkan variabel-variabel yang ada dalam kuisioner SUSENAS tahun 2016. Metode yang digunakan untuk

pengelompokan pada data numerik adalah metode pengelompokan hirarki *agglomerative* dengan jarak *euclidean*.

## **BAB II**

### **TINJAUAN PUSTAKA**

Dalam penelitian ini digunakan literatur untuk membantu peneliti dalam menyelesaikan permasalahan yang ada. Metode statistik yang digunakan menggunakan yaitu metode Ensemble ROCK karena merupakan salah satu metode pengelompokan untuk data campuran. Terdapat pula faktor-faktor yang mempengaruhi rumah tangga untuk konsumsi air bersih. Berikut dijelaskan beberapa teori yang terkait.

#### **2.1 Statistika Deskriptif**

Statistika terdiri dari dua macam yaitu statistika deskriptif dan statistika inferensia. Statistika deskriptif merupakan metode yang digunakan untuk menyajikan data agar lebih informatif. Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna (Walpole, 1995). Karakteristik pada data dengan skala numerik dapat dijelaskan dengan ukuran pemusatan data yaitu rata-rata dan median. Pada data berskala kategorik dapat dijelaskan dengan ukuran pemusatan data yaitu modus dengan melihat nilai yang sering muncul atau frekuensi yang paling banyak.

Statistika deskriptif dapat dijelaskan secara visual dalam bentuk gambar, tabel serta grafik untuk mempermudah menyampaikan informasi dari data. Pada data numerik dapat disajikan dengan *boxplot* melalui lima ukuran yaitu nilai observasi terkecil, kuartil terendah atau kuartil pertama (Q1), nilai pertengahan atau median (Q2), kuartil tertinggi atau kuartil ketiga (Q3), dan nilai observasi terbesar. Selain itu dalam *boxplot* juga ditunjukkan nilai *outlier* dari observasi (Junaidi, 2014).

#### **2.2 Data Numerik dan Data Kategorik**

Data adalah sumber informasi yang bentuknya masih mentah. Menurut Indrajani (2015), data adalah fakta-fakta mentah kemudian dikelola sehingga menghasilkan informasi yang penting bagi sebuah perusahaan atau organisasi. Data dikelompokkan

menjadi dua berdasarkan jenis variabelnya yaitu data numerik dan data kategorik. Data numerik terdiri dari skala data interval dan rasio, sedangkan data kategorik terdiri dari skala data nominal dan ordinal.

Data numerik digolongkan menjadi dua yaitu data diskrit, misalnya jumlah penduduk suatu daerah dan data kontinu yaitu data yang didapat dari hasil pengukuran, misalnya berat badan. Data kategorik adalah data yang sifatnya hanya dapat dibedakan dan dapat diurutkan serta merupakan data dimana variabel-variabelnya dapat dikelompokkan menjadi beberapa kelompok atau kategori seperti jenis kelamin, agama yang dianut, atau jenjang pendidikan.

### **2.3 Analisis Klaster**

Analisis klaster merupakan salah satu teknik dalam analisis multivariat metode independen yang mereduksi (mengelompokkan) data. 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.

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). Tahap pengelompokan dalam analisis kelompok dibedakan menurut jenis data yang dimiliki. Analisis kelompok umumnya terfokus pada data numerik, akan tetapi terdapat kasus dengan data kategorik yang penyelesaiannya berbeda dengan data numerik.

### 2.3.1. Pengelompokkan Data Numerik

Pengelompokkan pada data numerik didasarkan pada ukuran ketakmiripan atau jarak. Pengamatan yang memiliki jenis data numerik memiliki banyak metode untuk menghitung ukuran jarak namun yang biasa digunakan adalah jarak *euclidean*. Jarak *euclidean* relatif mudah dimengerti serta dapat digunakan pada data dengan variabel lebih dari dua. Rumus untuk menghitung jarak *Euclidean* dengan jumlah variabel sebanyak  $m$  dapat ditunjukkan pada persamaan (2.1).

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2}, i = (1, 2, \dots, m) \quad (2.1)$$

dan  $j = (1, 2, \dots, m)$

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, Anderson, & Tatham, 2009).

Data numerik dapat dikelompokkan berdasarkan ukuran ketidakmiripan atau jarak untuk data numerik. Hasil dari pengelompokkan disajikan dalam bentuk dendrogram atau diagram pohon sehingga dapat mengamati objek dengan lebih mudah serta informatif. Terdapat metode hirarki dan non hirarki yang digunakan dalam pengelompokkan.

Pengelompokkan hirarki dimulai dengan dua atau lebih objek yang mempunyai kesamaan paling dekat lalu diteruskan ke objek lain yang mempunyai kedekatan kedua. Analisis dilakukan hingga kelompok membentuk hirarki (tingkatan) yang jelas antar

objek dari yang mirip sampai paling tidak mirip. Terdapat dua teknik pengelompokan dalam analisis kelompok hirarki yaitu teknik pembagian (*divide*) dan teknik penggabungan (*agglomerative*).

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

Andaikan  $d_{uv}$  merupakan ukuran ketidakmiripan antara kelompok ke- $u$  dengan kelompok ke- $v$  serta  $d_{w(w)}$  merupakan ukuran ketidakmiripan antara kelompok ke- $w$  dengan kelompok ( $u, v$ ) yang merupakan penggabungan antara kelompok ke- $u$  dan kelompok ke- $v$ , maka beberapa teknik pengelompokan antar kelompok dinyatakan sebagai berikut (Johnson & Winchern, 2007).

- a. Pautan Tunggal (*Single Linkage/Nearest Neighbor*)  
Teknik ini didasarkan pada jarak terkecil atau jarak terdekat. Jika dua objek terpisah oleh jarak yang pendek maka kedua objek tersebut digabung menjadi satu kelompok dan demikian seterusnya. Ukuran jarak yang digunakan adalah

$$d_{w(w)} = \min(d_{wu}, d_{wv}) \quad (2.2)$$

- b. Pautan Lengkap (*Complete Linkage/Farthest Neighbor*)  
Teknik ini berlawanan dengan *single linkage* karena dalam prosedur ini didasarkan pada jarak terbesar atau jarak terjauh. Ukuran jarak yang digunakan adalah

$$d_{w(w)} = \max(d_{wu}, d_{wv}) \quad (2.3)$$

c. Pautan Rataan (*Average Linkage Between Method / BAVERAGE*)

Teknik dalam pautan rata-rata hampir sama dengan *single linkage* maupun *complete linkage*, namun kriteria yang digunakan adalah rata-rata jarak seluruh individu dalam suatu kelompok dengan jarak seluruh individu dalam kelompok yang lain dengan  $n_u$  dan  $n_v$  merupakan jumlah pengamatan dalam kelompok ke- $u$  dan ke- $v$ , ukuran jarak yang digunakan adalah

$$d_{w(uv)} = \frac{n_u}{n_u + n_v} d_{wu} + \frac{n_v}{n_u + n_v} d_{wv} \quad (2.4)$$

d. *Ward's Method*

*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 meminimumkan variansi dalam kelompok. Nilai ESS ditunjukkan pada persamaan (2.5)

$$ESS = \sum_{j=1}^N (x_j - \bar{x})'(x_j - \bar{x}) \quad (2.5)$$

Tahapan penting setelah proses pengelompokan adalah menentukan jumlah kelompok optimum. 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 adalah sebagai berikut,

*Sum of Square Total (SST)*

$$SST = \sum_{l=1}^{m_{numerik}} \sum_{i=1}^n (x_{il} - \bar{x}_l)^2 \quad (2.6)$$

*Sum of Square Within Group (SSW)*

$$SSW = \sum_{c=1}^C \sum_{l=1}^{m_{numerik}} \sum_{i=1}^{n_c} (x_{ilc} - \bar{x}_{lc})^2 \quad (2.7)$$

*Sum of Square Between Group (SSB)*

$$SSB = SST - SSW \quad (2.8)$$

dengan,

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

$C$  : Jumlah kelompok yang dibentuk dalam pengamatan,

$n$  : Total jumlah objek pengamatan,

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

$\bar{x}_l$  : Rata-rata keseluruhan objek pada variabel ke- $l$  untuk

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

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

$R^2$  kelompok baru adalah rasio dari  $SSB$  dan  $SST$ .  $R^2$  dapat didefinisikan sebagai ukuran perbedaan antar kelompok, dengan nilai berkisar dari 0 sampai 1. Nilai  $R^2 = 0$  artinya tidak terdapat perbedaan antara kelompok, sedangkan nilai  $R^2 = 1$  menunjukkan adanya perbedaan yang signifikan antar kelompok yang terbentuk. Nilai  $R^2$  dirumuskan pada persamaan (2.9).

$$R^2 = \frac{SSB}{SST} = \frac{[SST - SSW]}{SST} \quad (2.9)$$

Penentuan jumlah kelompok yang terbentuk dapat dilihat berdasarkan nilai maksimum dari *PseudoF* (Orpin & Kostylev, V. E., 2006). Rumus yang digunakan untuk menghitung nilai *PseudoF* adalah sebagai berikut.

$$PseudoF = \frac{\left( \frac{R^2}{c-1} \right)}{\left( \frac{1-R^2}{n-c} \right)} \quad (2.10)$$

dimana,

$n$  : banyak objek

$c$  : banyak klaster.

### 2.3.2. Pengelompokan Data Kategorik

Metode tradisional yang menggunakan konsep jarak antar titik untuk mengelompokkan variabel bertipe kategorik dinilai tidak tepat digunakan pada data kategorik. Oleh karena itu, dikembangkan metode pengelompokan hirarki *agglomerative* yang digunakan untuk data kategorik yaitu metode ROCK atau *RObust Clustering using linKs* (Guha, Rastogi, & Shim, 2000).

Pada metode ROCK dibentuk konsep baru yaitu *link* yang merupakan kesamaan/kedekatan antara sepasang titik data. Metode ROCK menggunakan informasi tentang *link* sebagai ukuran kemiripan antar objek. Jika terdapat objek pengamatan  $X_i$ ,  $X_j$ , dan  $X_k$  dimana  $X_i$  tetangga dari  $X_j$ , dan  $X_j$  tetangga dari  $X_k$ , maka dikatakan  $X_i$  memiliki *link* dengan  $X_k$  walaupun  $X_i$  bukan tetangga dari  $X_k$ . Pengamatan yang memiliki tingkat hubungan (*link*) tinggi digabungkan dalam satu kelompok, sedangkan yang rendah dipisahkan dari data yang dikelompokkan (Guha, Rastogi, & Shim, 2000).

Pengelompokan dengan metode ROCK terdiri atas beberapa langkah. Berikut adalah langkah pengelompokan dengan metode ROCK (Dutta, Mahanta, & Pujari, 2005).

1. Menentukan inisialisasi untuk masing-masing objek sebagai *cluster* pada awalnya.
2. Menghitung Kemiripan (*Similarity*)

Kemiripan antara dua objek dapat didefinisikan oleh beberapa dugaan yang tepat dari jarak antar objek data. Ukuran kemiripan antara pasangan pengamatan ke- $i$  dan ke- $j$  dihitung dengan rumus pada persamaan (2.11).

$$\text{sim}(X_i, X_j) = \frac{|X_i \cap X_j|}{|X_i \cup X_j|}, i \neq j \quad (2.11)$$

dengan

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

$X_i$ : Himpunan pengamatan ke- $i$  dengan  $X_i = \{x_{i1}, x_{i2}, \dots, x_{m_{\text{kategorik } i}}\}$

$X_j$ : Himpunan pengamatan ke- $j$  dengan  $X_j = \{x_{1j}, x_{2j}, \dots, x_{m_{\text{kategori } j}}\}$

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

### 3. Menentukan Tetangga (*Neighbors*)

Pengamatan  $X_i$  dan  $X_j$  dinyatakan sebagai tetangga jika nilai  $\text{sim}(X_i, X_j) \geq \theta$ , dimana nilai *threshold*  $\theta$  berkisar antara 0 hingga 1. Nilai *threshold* ( $\theta$ ) merupakan parameter yang ditentukan oleh pengguna yang dapat digunakan untuk mengontrol seberapa dekat hubungan antara objek.

### 4. Menghitung *Link*

*Link* adalah kesamaan atau kedekatan antara sepasang titik data. Cara untuk menghitung *link* untuk semua kemungkinan pasangan dari  $n$  objek dapat menggunakan matriks  $\mathbf{A}$  yang merupakan matriks berukuran  $n \times n$  yang bernilai 1 jika  $X_i$  dan  $X_j$  dinyatakan mirip (tetangga) dan bernilai 0 jika  $X_i$  dan  $X_j$  tidak mirip (bukan tetangga). Jumlah *link* antar pasangan  $X_i$  dan  $X_j$  diperoleh dari hasil kali antara baris ke  $X_i$  dan kolom ke  $X_j$  dari matriks  $\mathbf{A}$ . Jika *link* antara  $X_i$  dan  $X_j$  semakin besar maka semakin besar kemungkinan  $X_i$  dan  $X_j$  berada dalam satu kelompok yang sama (Dutta, Mahanta, & Arun, 2005).

### 5. *Goodness Measure*

Pada algoritma ROCK, penggabungan kelompok didasarkan pada ukuran kebaikan atau *goodness measure* antar kelompok. *Goodness measure* adalah persamaan yang menghitung jumlah *link* dibagi dengan kemungkinan *link* yang terbentuk berdasarkan ukuran kelompoknya (Tyagi & Sharma, 2012). *Goodness measure* dapat dihitung dengan rumus pada persamaan (2.12).

$$g(C_i, C_j) = \frac{\text{link}[C_i, C_j]}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}}, \quad (2.12)$$

dimana  $g(C_i, C_j)$  merupakan gabungan dua klaster  $C_i, C_j$ .

*Goodness measure* digunakan untuk mengidentifikasi pasangan terbaik untuk digabungkan pada setiap langkah

pada metode ROCK dengan  $link[C_i, C_j] = \sum_{X_i \in C_i, X_j \in C_j} link(X_i, X_j)$

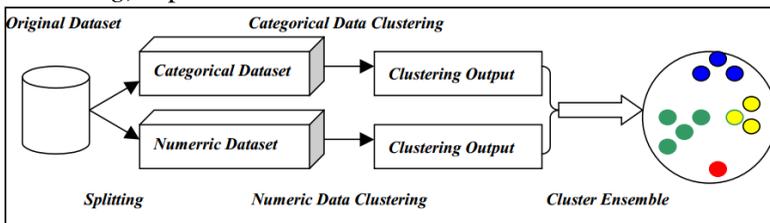
yang menyatakan jumlah semua kemungkinan pasangan objek yang ada di dalam  $C_i$  dan  $C_j$ , serta  $n_i$  dan  $n_j$  masing-masing menyatakan jumlah anggota dalam kelompok ke- $i$  dan ke- $j$ . Adapun  $f(\theta)$  dinyatakan sebagai  $\frac{1-\theta}{1+\theta}$ .

Pengelompokkan pada algoritma ROCK akan berhenti ketika dalam keadaan berikut.

1. Jumlah dari kelompok yang diharapkan sudah terpenuhi, atau
2. Tidak ada *link* antar kelompok.

## 2.4 Pengelompokan Ensemble

Pengelompokan ensemble bertujuan untuk menggabungkan hasil pengelompokan dari beberapa algoritma pengelompokan untuk mendapatkan hasil pengelompokan yang lebih baik dan *robust* (Yoon, Ahn, Lee, Cho, & Kim, 2006). Pengelompokan *ensemble* terdiri dari dua tahap algoritma. Tahap pertama adalah melakukan pengelompokan dengan beberapa algoritma dan menyimpan hasil pengelompokan tersebut. Tahap kedua adalah menggunakan fungsi consensus untuk menentukan *final cluster* dari kelompok-kelompok yang sudah diperoleh pada tahap pertama. Pengelompokan *ensemble* dapat menggunakan algoritma CEBMDC (*Cluster Ensemble Based Mixed Data Clustering*) seperti dalam Gambar 2.1.



**Gambar 2.1** Tahapan Analisis dengan Algoritma CEBMDC

Pengelompokan *ensemble* dengan data campuran dapat menggunakan algoritma CEBMDC (*Cluster Ensemble Based*

*Mixed Data Clustering*) dengan tahapan sebagai berikut (He, Xu, & Deng, 2005).

- a. Membagi data menjadi dua subdata, yaitu murni numerik dan murni kategorik.
- b. 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.
- c. Menggabungkan (*combining*) hasil pengelompokan dari variabel numerik dan kategorik yang disebut proses ensemble. Tahapan ini sama dengan melakukan pengelompokan data kategorik menggunakan metode ROCK, dimana *input* untuk tahapan ini adalah kelompok hasil metode hirarki *agglomerative* (*output 1*) dan kelompok hasil metode ROCK (*output 2*). *Output 1* dan *output 2* dinyatakan sebagai variabel kategorik yang digunakan untuk menyusun *final cluster* menggunakan metode ROCK. *Input* yang diberikan dalam analisis pengelompokan ensemble ROCK merupakan matriks **D** sebagai berikut.

$$\mathbf{D} = \left[ \begin{array}{cccc|cccc} x_{11} & x_{12} & \cdots & x_{1m_{\text{numerik}}} & y_{11} & y_{12} & \cdots & y_{1m_{\text{kategorik}}} \\ x_{21} & x_{22} & \cdots & x_{2m_{\text{numerik}}} & y_{21} & y_{22} & \cdots & y_{2m_{\text{kategorik}}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm_{\text{numerik}}} & y_{n1} & y_{n2} & \cdots & y_{nm_{\text{kategorik}}} \end{array} \right]$$

- d. Melakukan pengelompokan ensemble menggunakan algoritma pengelompokan data kategorik (metode ROCK) untuk mendapatkan kelompok akhir (*final cluster*) dengan menggunakan ukuran jarak pada persamaan (2.11). Jarak yang diperoleh dari  $n$  objek pengamatan dinyatakan dalam matriks **sim** berukuran  $n \times n$ .

$$\mathbf{sim} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

## 2.5 Kinerja Hasil Pengelompokan

Kinerja hasil pengelompokan dilakukan untuk mengetahui validitas suatu pengelompokan. Kelompok yang baik adalah kelompok yang memiliki kehomogenan tinggi antar anggota dalam kelompok dan keheterogenan yang tinggi antar kelompok (Hair, Black, Anderson, & Tatham, 2009). Kinerja hasil pengelompokan untuk variabel dengan skala data numerik dapat diketahui dengan melihat rasio nilai  $S_w$  dan  $S_B$  (Bunkers & James, 1996). Digunakan nilai rata-rata variabel, simpangan baku dalam kelompok serta simpangan baku antar kelompok seperti pada persamaan berikut

$$S_w = \frac{1}{C} \sum_{c=1}^C S_c, \quad (2.13)$$

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

$$S_B = \left[ \frac{1}{C-1} \sum_{c=1}^C (\bar{x}_c - \bar{x})^2 \right]^{1/2}, \quad (2.14)$$

dimana  $\bar{x}_c$  adalah rata-rata kelompok ke- $c$  dan  $\bar{x}$  adalah rata-rata keseluruhan kelompok. Apabila nilai rasio antara  $S_w$  dan  $S_B$  semakin kecil maka kinerja suatu metode pengelompokan semakin baik. Sehingga dapat diketahui bahwa terdapat homogenitas maksimum dalam kelompok dan heterogenitas maksimum antar kelompok (Bunkers & James, 1996).

Pengukuran kinerja menggunakan rasio  $S_w$  dan  $S_B$  hanya dapat digunakan untuk data numerik sedangkan untuk data kategorik adalah dengan menggunakan tabel kontingensi atau sama 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 sebuah variabel dengan data kategorik dapat dirumuskan pada persamaan (2.15). Total jumlah kuadrat dalam kelompok ( $SSW$ ) dan jumlah kuadrat antar kelompok ( $SSB$ ) dapat dirumuskan seperti pada persamaan (2.16) (Dewi, 2012).

$$SST = \frac{n}{2} - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (2.15)$$

$$SSW = \frac{n}{2} - \frac{1}{2} \sum_{c=1}^C \frac{1}{n_{.c}} \sum_{k=1}^K n_{kc}^2 \quad \text{dan} \quad SSB = \frac{1}{2} \left( \sum_{c=1}^C \frac{1}{n_{.c}} \sum_{k=1}^K n_{kc}^2 \right) - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (2.16)$$

Simpangan baku dalam kelompok ( $S'_w$ ) dan simpangan baku antar kelompok ( $S'_B$ ) untuk data kategorik dapat dirumuskan pada persamaan (2.17).

$$S'_w = \left[ \frac{SSW}{(n-C)} \right]^{1/2} \quad \text{dan} \quad S'_B = \left[ \frac{SSB}{C-1} \right]^{1/2}, \quad (2.17)$$

dimana,

$n$  : banyak pengamatan,

$C$  : jumlah kelompok yang terbentuk,  $c = 1,2,3,\dots,C$ .

Apabila nilai rasio antara  $S'_w$  dan  $S'_b$  semakin kecil maka menunjukkan kinerja metode yang semakin baik karena homogenitas maksimum dalam kelompok dan heterogenitas maksimum antar kelompok.

## 2.6 Uji *Kruskal-Wallis*

Salah satu teknik nonparametrik yang digunakan untuk mengetahui apakah sampel berasal dari populasi identik atau tidak yaitu uji *Kruskal-Wallis*. Hipotesis yang digunakan untuk pengujian *Kruskal-Wallis* adalah sebagai berikut.

$H_0$  : Kedua populasi identik

$H_1$  : Tidak semua populasi identik

Dengan statistik uji sesuai persamaan (2.18)

$$H = \frac{12}{N(N+1)} \sum_{i=1}^g \frac{R_i}{n_i} - 3(N+1) \quad (2.18)$$

dimana

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

$n_i$  : banyaknya pengamatan tiap kelompok

$N$  : banyaknya seluruh pengamatan

$K$  : banyaknya kelompok

Diperoleh keputusan Tolak  $H_0$  apabila nilai  $H$  lebih besar dari  $\chi_{(0,05,K-1)}$  atau nilai  $\chi_{hitung}$  lebih besar daripada  $\chi_{(0,05,K-1)}$  (Daniel, 1989).

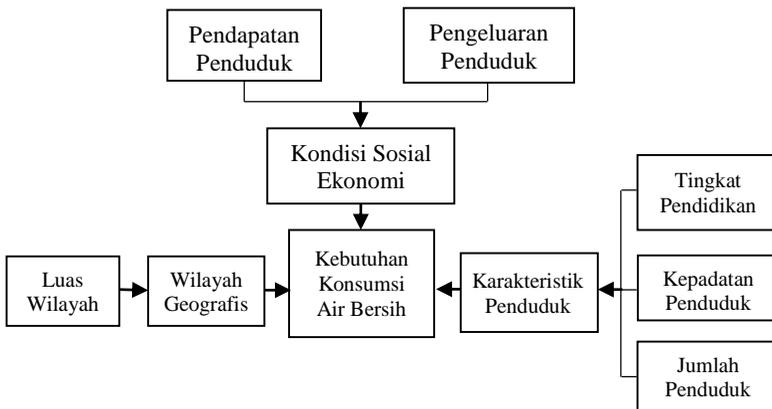
## 2.7 Faktor yang Mempengaruhi Pemenuhan Konsumsi Air Bersih

Upaya dalam meningkatkan jumlah konsumsi air bersih di suatu wilayah perlu diperhatikan dengan mengidentifikasi faktor yang mempengaruhi tingkat konsumsi air bersih. Menurut Linsley (1995), faktor kebutuhan air bersih salah satunya yaitu mengenai karakteristik kependudukan seperti kepadatan penduduk dan jumlah penduduk pada kawasan daerah maupun perkotaan, baik jumlah penduduk yang telah terlayani maupun yang membutuhkan air bersih.

Menurut Undang-Undang No. 20 Tahun 2003 tentang penentuan tingkat pendidikan berdasarkan kategori jenjang pendidikan adalah sebagai berikut.

1. Pendidikan dasar (SD-SMP)
2. Pendidikan menengah (SMA/SMK/ sederajat)
3. Pendidikan tinggi (Perguruan tinggi)

Penelitian yang dilakukan oleh Winarna (2003) menyebutkan bahwa faktor-faktor yang mempengaruhi konsumsi air bersih pelanggan rumahtangga diantaranya adalah pendapatan keluarga, pengeluaran total rumahtangga, jumlah anggota rumahtangga, pendidikan kepala rumahtangga dan ada atau tidaknya sumber air PDAM. Pengeluaran untuk konsumsi air tidak berbeda jauh dengan pengeluaran lain dalam rumah tangga.



**Gambar 2.2** Faktor-Faktor yang Mempengaruhi Kebutuhan Konsumsi Air Rumah Tangga

Menurut Darr dan Kamen (1976) beberapa faktor yang menjadi pertimbangan dalam mempengaruhi tingkat konsumsi air rumah tangga antara lain ukuran rumah tangga, pendapatan perkapita, perkembangan wilayah, tipe meteran, pendidikan responden, kepadatan ruang, cakupan pelayanan. Sehingga beberapa faktor yang mempengaruhi konsumsi air bersih diantaranya yaitu

berdasarkan wilayah geografis, beban tanggungan keluarga, kondisi sosial ekonomi, dan karakteristik rumah tangga.

*(Halaman ini sengaja dikosongkan)*

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Sumber Data**

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari hasil SUSENAS (Survei Sosial Ekonomi Nasional) 2016 di Jawa Timur dan BPS (Badan Pusat Statistik) Provinsi Jawa Timur dengan unit pengamatan kabupaten/kota yang terdiri dari 29 kabupaten dan 9 kota.

#### **3.2 Variabel Penelitian**

Variabel penelitian yang digunakan terdiri atas 3 variabel dengan tipe data numerik dan 4 variabel dengan tipe data kategorik dengan total sebanyak 7 variabel. Daftar variabel penelitian disajikan dalam Tabel 3.1.

**Tabel 3.1** Variabel Penelitian

<b>No.</b>	<b>Variabel</b>	<b>Keterangan</b>	<b>Skala Data</b>
1	X <sub>1</sub>	Kepadatan Penduduk	Numerik
2	X <sub>2</sub>	Persentase Tingkat Pendidikan Terakhir Kepala Rumah Tangga dibawah SMA	Numerik
3	X <sub>3</sub>	Rata-rata Pengeluaran yang Dikeluarkan Oleh Suatu Rumah Tangga	Numerik
4	X <sub>4</sub>	Sumber Air Minum Utama	Kategorik
5	X <sub>5</sub>	Sumber Air Masak	Kategorik
6	X <sub>6</sub>	Sumber Air Mandi/Cuci	Kategorik
7	X <sub>7</sub>	Penggunaan Air PAM	Kategorik

Definisi operasional dari variabel-variabel yang digunakan dalam penelitian sesuai dengan nomor urut dalam Tabel 3.1 adalah sebagai berikut.

1. Kepadatan penduduk adalah perbandingan jumlah penduduk dengan luas wilayahnya.
2. Tingkat pendidikan terakhir kepala Rumah Tangga dibawah SMA merupakan jenjang pendidikan terakhir yang ditempuh

oleh kepala Rumah Tangga dengan pendidikan terakhir yaitu pendidikan dasar (SD dan SMP) maupun tidak bersekolah

3. Rata-rata pengeluaran yang dikeluarkan oleh suatu rumahtangga merupakan rata-rata pengeluaran yang dikeluarkan oleh suatu rumah tangga pada wilayah tertentu untuk konsumsi sehari-hari seperti makanan, pendidikan, dan kebutuhan sehari-hari pada sebulan terakhir.
4. Sumber air minum utama merupakan sumber asal air yang sering dikonsumsi untuk diminum sehari-hari serta mendominasi atau paling banyak dikonsumsi pada suatu wilayah.

Kategori berdasarkan SUSENAS:

- 1) Kode 0 (nol) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu air kemasan bermerek
- 2) Kode 1 (satu) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu air isi ulang
- 3) Kode 2 (dua) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu leding meteran
- 4) Kode 3 (tiga) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu leding eceran
- 5) Kode 4 (empat) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu sumur bor atau pompa
- 6) Kode 5 (lima) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu sumur terlindungi
- 7) Kode 6 (enam) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggu-

- nakan sumber air untuk minum yaitu sumur tak terlindungi
- 8) Kode 7 (tujuh) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu mata air terlindungi
  - 9) Kode 8 (delapan) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu mata air tak terlindungi
  - 10) Kode 9 (sembilan) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu air permukaan (sungai, danau/waduk, kolam, irigasi)
  - 11) Kode 10 (sepuluh) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu air hujan
  - 12) Kode 11 (sebelas) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk minum yaitu air lainnya
5. Sumber air masak utama merupakan sumber asal air yang sering dikonsumsi untuk memasak sehari-hari serta mendominasi atau paling banyak dikonsumsi pada suatu wilayah.

Kategori berdasarkan SUSENAS:

- 1) Kode 0 (nol) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu air kemasan bermerek
- 2) Kode 1 (satu) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu air isi ulang
- 3) Kode 2 (dua) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu leding meteran

- 4) Kode 3 (tiga) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu leding eceran
  - 5) Kode 4 (empat) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu sumur bor atau pompa
  - 6) Kode 5 (lima) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu sumur terlindungi
  - 7) Kode 6 (enam) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu sumur tak terlindungi
  - 8) Kode 7 (tujuh) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu mata air terlindungi
  - 9) Kode 8 (delapan) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu mata air tak terlindungi
  - 10) Kode 9 (sembilan) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu air permukaan (sungai, danau/waduk, kolam, irigasi)
  - 11) Kode 10 (sepuluh) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu air hujan
  - 12) Kode 11 (sebelas) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk masak yaitu air lainnya
6. Sumber air mandi atau cuci merupakan sumber asal air yang sering dikonsumsi untuk mandi serta mencuci sehari-hari

serta mendominasi atau paling banyak dikonsumsi pada suatu wilayah.

Kategori berdasarkan SUSENAS:

- 1) Kode 0 (nol) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu air kemasan bermerek
- 2) Kode 1 (satu) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu air isi ulang
- 3) Kode 2 (dua) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu leding meteran
- 4) Kode 3 (tiga) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu leding eceran
- 5) Kode 4 (empat) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu sumur bor atau pompa
- 6) Kode 5 (lima) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu sumur terlindungi
- 7) Kode 6 (enam) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu sumur tak terlindungi
- 8) Kode 7 (tujuh) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu mata air terlindungi
- 9) Kode 8 (delapan) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggu-

nakan sumber air untuk mandi/cuci yaitu mata air tak terlindungi

- 10) Kode 9 (sembilan) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu air permukaan (sungai, danau/waduk, kolam, irigasi)
  - 11) Kode 10 (sepuluh) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu air hujan
  - 12) Kode 11 (sebelas) merupakan kabupaten/kota yang memiliki persentase sebagian besar penduduk menggunakan sumber air untuk mandi/cuci yaitu air lainnya
7. Penggunaan air PAM merupakan kabupaten/kota dengan sebagian besar penduduk menggunakan atau tidak menggunakan air PAM serta mendominasi atau paling banyak digunakan pada suatu wilayah.

Kategori:

- 1) Kode 0 (nol) merupakan kabupaten/kota yang tidak menggunakan air PAM
- 2) Kode 1 (satu) merupakan kabupaten/kota yang menggunakan air PAM

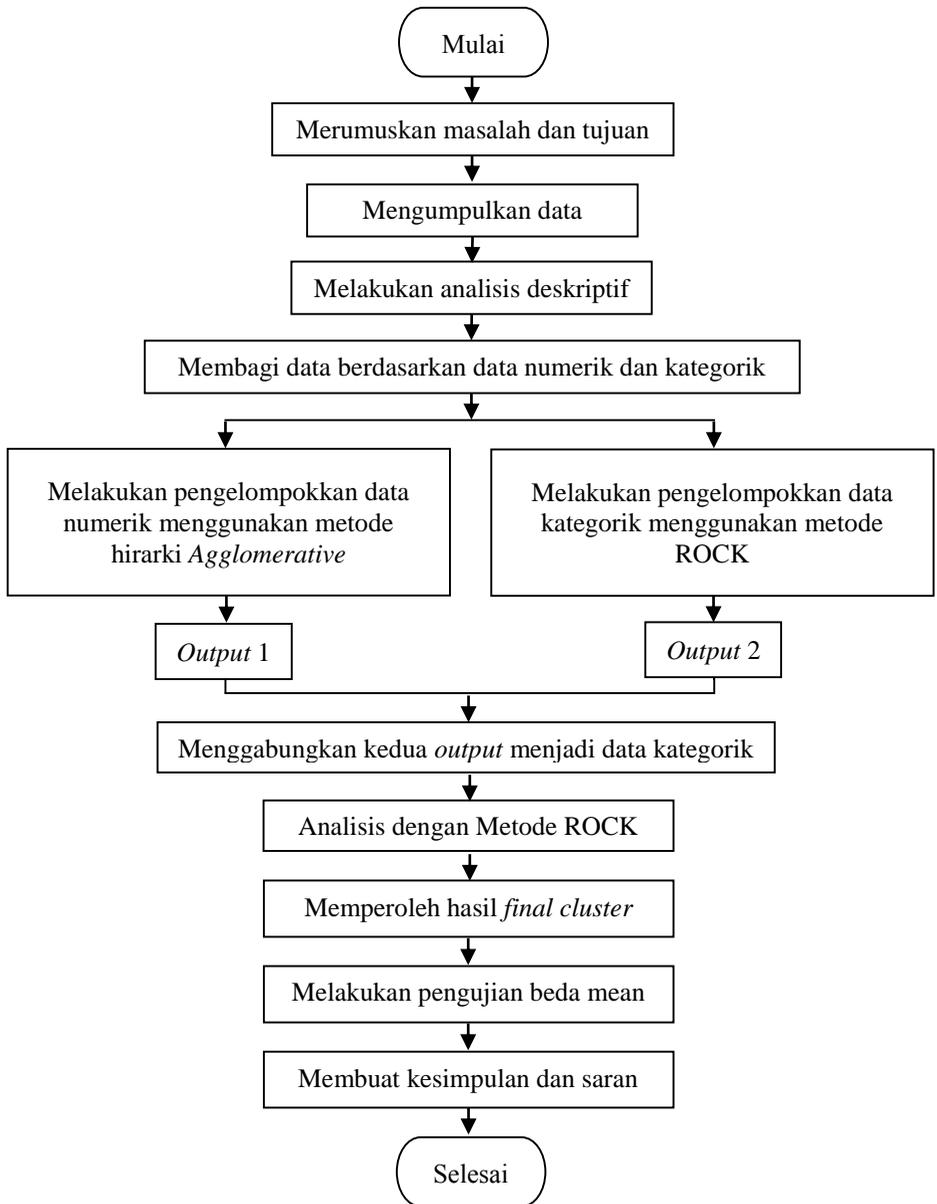
### 3.3 Langkah Analisis

Langkah-langkah yang dikerjakan dalam pengelompokan kabupaten/kota di Jawa Timur menggunakan metode *ensemble ROCK* dengan tahapan pada masing-masing jenis data sebagai berikut.

1. Membuat analisis statistika deskriptif untuk variabel-variabel numerik dan kategorik
2. Membagi variabel penelitian menjadi data kategorik dan data numerik
3. Pengelompokkan variabel numerik menggunakan metode Hirarki *Agglomerative*
  - a. Menentukan ukuran kemiripan dengan jarak *Euclidean* dengan rumus pada persamaan (2.1) dan membuat matriks jarak berukuran  $n \times n$

- b. Menggabungkan kelompok yang memiliki jarak terdekat
  - c. Memperbarui matriks jarak dengan metode *single linkage*, *average linkage*, *complete linkage*, dan *ward* seperti pada persamaan (2.2), (2.3), (2.4), dan (2.5)
  - d. Menentukan jumlah kelompok yang optimum menggunakan nilai *pseudoF* seperti persamaan (2.10)
  - e. Menghitung dan memilih rasio  $S_W$  dan  $S_B$  terkecil untuk mendapatkan kinerja pengelompokkan terbaik untuk *single linkage*, *complete linkage*, *average linkage*, dan *ward* dengan rumusan pada persamaan (2.13) dan (2.14)
4. Pengelompokkan variabel kategorik menggunakan metode ROCK
- a. Melakukan inisialisasi objek sebagai kelompok dengan anggota tunggal
  - b. Membentuk *similarity* antara objek dengan kriteria menggunakan persamaan (2.11)
  - c. Menentukan *threshold* ( $\theta$ ) dalam matriks *adjacency* dengan nilai *threshold* yang digunakan antara 0.05 hingga 0.5
  - d. Menghitung nilai *link* antara pengamatan
  - e. Menghitung nilai *goodness measure* menggunakan persamaan (2.12)
  - f. Menentukan nilai maksimum *goodness measure* antara kolom pada baris ke- $i$
  - g. Ulangi langkah (e) dan (f), sehingga diperoleh nilai maksimum *goodness measure* dan nilai *goodness measure*
  - h. Selama ukuran data  $> k$ , dengan  $k$  adalah jumlah kelompok yang ditentukan maka lakukan penggabungan kelompok yang memiliki nilai *goodness measure* terbesar dengan maksimum *goodness measure* terbesar menjadi satu kelompok, kemudian tambahkan *link* antar kelompok yang digabungkan, selanjutnya hapus kelompok yang digabungkan dari nilai *goodness measure*

- measure* dan perbaruan nilai maksimum *goodness measure* dengan hasil penggabungan
- i. Lakukan langkah (h), sehingga diperoleh jumlah kelompok yang diharapkan atau tidak terdapatkan *link* antar kelompok
  - j. Mengulangi langkah (a) sampai dengan (i) dengan nilai  $\theta$  berbeda
  - k. Menghitung rasio  $S'_w$  dan  $S'_B$  untuk masing-masing nilai  $\theta$  dengan rumus seperti pada persamaan (2.17)
  - l. Membandingkan hasil langkah (k) untuk masing-masing nilai  $\theta$  dan menentukan jumlah kelompok yang optimum dengan kriteria rasio  $S'_w$  dan  $S'_B$  terkecil.
5. Penggabungan hasil pengelompokkan (tahapan ensemble)
  6. Melakukan klaster pada data hasil pengelompokkan menggunakan metode ROCK untuk mendapatkan kelompok akhir (*final cluster*) dengan menggunakan ukuran jarak pada persamaan (2.11). Jarak yang diperoleh dari  $n$  objek pengamatan dinyatakan dalam matriks sim berukuran  $n \times n$ .
  7. Menghitung rasio  $S'_w$  dan  $S'_B$  untuk masing-masing nilai  $\theta$  dengan rumus seperti pada persamaan (2.17)
  8. Membuat analisis statistika deskriptif berdasarkan hasil pengelompokkan yang telah diperoleh
  9. Melakukan pengujian beda mean menggunakan uji *Kruskal Wallis*
  10. Membuat kesimpulan dan saran  
Berdasarkan angka analisis tersebut maka dibuatlah diagram alir yang menggambarkan alur perjalanan pembuatan laporan ini seperti ditunjukkan pada Gambar 3.1.



**Gambar 3.1** Diagram Alir Penelitian

*(Halaman ini sengaja dikosongkan)*

## BAB IV ANALISIS DAN PEMBAHASAN

Pada bab ini dilakukan analisis terkait karakteristik data dan pengelompokkan kabupaten atau kota di Jawa Timur pada tahun 2016. Analisis yang digunakan adalah statistika deskriptif dan metode pengelompokkan *ensemble* ROCK.

### 4.1 Karakteristik Kabupaten atau Kota di Jawa Timur

Kabupaten atau kota di Jawa Timur terdiri dari 29 kabupaten dan 9 kota. Masing-masing kabupaten atau kota memiliki karakteristik berbeda-beda yang memiliki tipe data numerik dan tipe data kategorik. Oleh karena itu pada penelitian ini dilakukan analisis terkait karakteristik kabupaten atau kota berdasarkan tipe data untuk selanjutnya diperoleh hasil pengelompokan kabupaten atau kota di Jawa Timur.

Karakteristik variabel penelitian berskala numerik terdiri dari beberapa variabel diantaranya kepadatan penduduk ( $X_1$ ), persentase tingkat pendidikan terakhir kepala Rumah Tangga dibawah SMA ( $X_2$ ), dan rata-rata pengeluaran sebulan terakhir ( $X_3$ ). Karakteristik kabupaten atau kota di Jawa Timur berdasarkan data numerik ditunjukkan pada Tabel 4.1.

**Tabel 4.1** Deskripsi Variabel Penelitian Berskala Numerik

<b>Variabel</b>	<b>Rata-rata</b>	<b>StDev</b>	<b>Min</b>	<b>Nilai Tengah</b>	<b>Maks</b>
$X_1$ (ribu)	1,89	2,14	0,28	0,87	8,17
$X_2$	75,17	11,39	43,23	79,63	90,77
$X_3$ (juta)	3,08	1,10	1,9	2,56	6,74

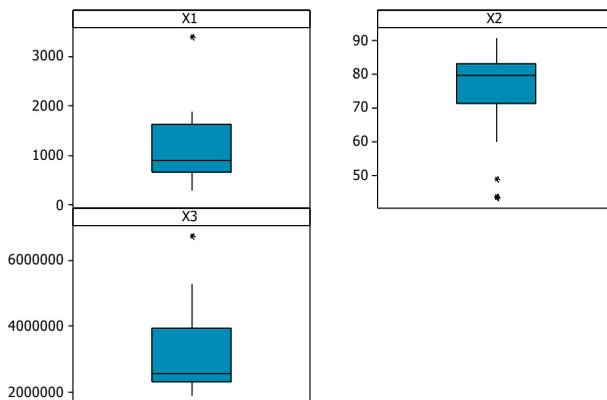
Tabel 4.1 menunjukkan bahwa variabel kepadatan penduduk pada kabupaten/kota di Jawa Timur memiliki rata-rata sebanyak 1890 penduduk per km<sup>2</sup> dengan kepadatan penduduk tertinggi yaitu 8170 penduduk per km<sup>2</sup> serta kepadatan penduduk terendah yaitu 277 penduduk per km<sup>2</sup>. Variabel kepadatan penduduk memiliki persebaran sebesar 2141 penduduk per km<sup>2</sup>.

Variabel persentase tingkat pendidikan terakhir kepala RT dibawah SMA pada kabupaten/ kota di Jawa Timur memiliki rata-rata persentase tingkat pendidikan sebesar 75,17%. Hal ini berarti sekitar 75 orang dari 100 orang kepala RT di Jawa Timur berpendidikan terakhir di bawah SMA. Persentase tingkat pendidikan yang tertinggi yaitu 90,77% serta persentase tingkat pendidikan yang terendah yaitu 43,23%. Variabel persentase tingkat pendidikan terakhir kepala RT dibawah SMA memiliki ragam sebesar 11,39% serta median sebesar 79,63% dimana tidak terpaut jauh dengan nilai *mean*.

Variabel pengeluaran dalam sebulan terakhir pada kabupaten/kota di Jawa Timur memiliki rata-rata pengeluaran yaitu Rp 3.080.460,- dengan pengeluaran tertinggi yaitu Rp 6.743.623,- serta pengeluaran terendah yaitu Rp 1.900.431,-. Variabel pengeluaran sebulan terakhir memiliki ragam sebesar Rp 1.109.867,- serta nilai median sebesar Rp 2.564.365,-.

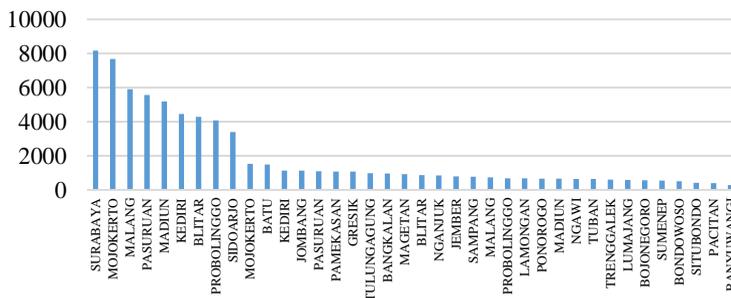
Untuk melihat penyebaran data dapat dilihat melalui sajian *boxplot* namun pada variabel dengan *outlier* yang mendekati 10% harus diatasi terlebih dahulu. Pada variabel kepadatan penduduk ( $X_1$ ) terdapat *outlier* yang lebih dari 10%, sehingga untuk variabel tersebut perlu diatasi terlebih dahulu menggunakan *imputasi mean* sehingga untuk selanjutnya dapat dianalisis dengan pengelompokan data.

Setelah dilakukan *imputasi mean* maka analisis dapat dilanjutkan. Gambar 4.1 menunjukkan *boxplot* pada variabel untuk data numerik setelah *outlier* diatasi. Dapat dilihat bahwa terdapat *outlier* pada setiap variabel numerik. Pada variabel kepadatan penduduk ( $X_1$ ) dan pengeluaran sebulan terakhir ( $X_3$ ) terlihat bahwa terdapat *outlier* di sisi atas yang berarti terdapat kepadatan penduduk dan pengeluaran sebulan terakhir diatas rata-rata. Pada variabel persentase tingkat pendidikan terakhir kepala RT dibawah SMA ( $X_2$ ) terdapat *outlier* di sisi bawah yang berarti terdapat penduduk yang memiliki persentase pendidikan dibawah rata-rata.



**Gambar 4.1** Karakteristik Variabel Data Numerik

Karakteristik variabel numerik yang pertama yaitu kepadatan penduduk ( $X_1$ ). Berikut tampilan grafik kepadatan penduduk kabupaten/kota di Jawa Timur.

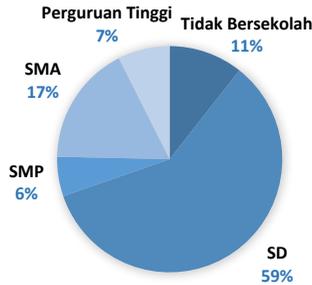


**Gambar 4.2** Kepadatan Penduduk

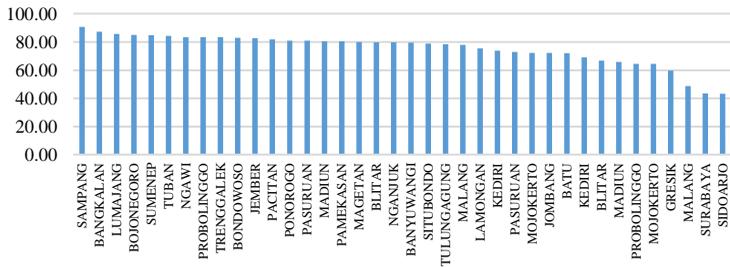
Gambar 4.2 menunjukkan kepadatan penduduk kabupaten atau kota di Jawa Timur. Jumlah kepadatan penduduk tertinggi berada pada kota Surabaya sedangkan jumlah kepadatan penduduk yang terendah yaitu pada kabupaten Banyuwangi.

Berikut merupakan persentase tingkat pendidikan kepala RT pada kabupaten atau kota di Jawa Timur menurut jenjang

pendidikan yang ditempuh. Sebagian besar kepala RT di Jawa Timur menempuh jenjang pendidikan Sekolah Dasar (SD) yaitu sebesar 59 persen dimana lebih dari setengah dari total kepala rumahtangga pada kabupaten atau kota di Jawa Timur. Kepala RT yang menempuh jenjang pendidikan SMA yaitu sebesar 17%, tidak bersekolah sebesar 11%, perguruan tinggi sebesar 7%, dan 6% sisanya menempuh jenjang pendidikan SMP. Hal ini dapat ditunjukkan melalui Gambar 4.3a.



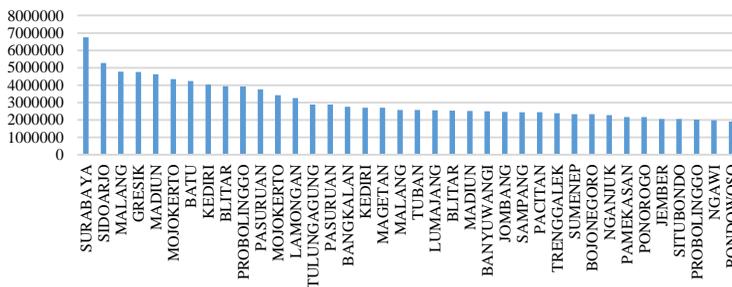
a)



b)

**Gambar 4.3** Persentase Tingkat Pendidikan Terakhir Kepala Rumah Tangga a) Menurut Jenjang Pendidikan b) Karakteristik Persentase Pendidikan dibawah SMA

Gambar 4.3b menunjukkan persentase tingkat pendidikan terakhir kepala RT dibawah SMA pada kabupaten atau kota di Jawa Timur. Persentase penduduk terbanyak terdapat pada kabupaten Sampang sedangkan persentase penduduk yang paling sedikit yaitu pada kabupaten Sidoarjo.

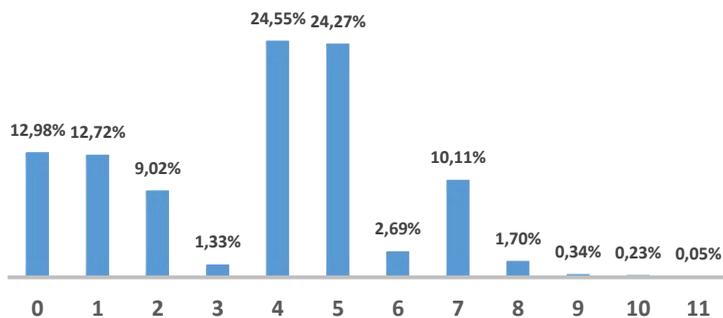


**Gambar 4.4** Rata-rata Pengeluaran Sebulan Terakhir

Gambar 4.4 menunjukkan rata-rata pengeluaran sebulan terakhir ( $X_3$ ) kabupaten atau kota di Jawa Timur. Pengeluaran terbanyak terdapat pada kota Surabaya sedangkan jumlah pengeluaran yang paling sedikit yaitu pada kabupaten Bondowoso.

Karakteristik kabupaten atau kota berdasarkan data kategorik terdiri dari beberapa variabel diantaranya sumber air minum utama ( $X_4$ ), sumber air masak ( $X_5$ ), sumber air mandi atau cuci ( $X_6$ ), dan penggunaan air non PAM atau PAM ( $X_7$ ). Karakteristik pada kabupaten/kota di Jawa Timur berdasarkan data kategorik adalah sebagai berikut.

Pada variabel sumber air minum utama ( $X_4$ ) terdapat 12 kategori. Dapat dilihat bahwa bahwa sebanyak 24,55% penduduk di Provinsi Jawa Timur banyak menggunakan sumur bor/pompa sebagai sumber utama untuk minum ( $X_4$ ). Selanjutnya terbanyak kedua yaitu sumur terlindungi dengan persentase 24,27%, diikuti oleh air kemasan bermerek (12,98%), air isi ulang (12,72%), mata air terlindungi (10,11%), leding meteran (9,02%), sumur tak terlindungi (2,69%), mata air tak terlindungi (1,7%), dan leding eceran (1,33%). Sumber air minum yang sedikit dikonsumsi oleh penduduk di Prinsi Jawa Timur yaitu air permukaan (0,34%), air hujan (0,23%), dan jenis air lainnya (0,05%). Kategori pada variabel sumber air minum utama dapat dilihat pada Gambar 4.5.



Keterangan:

0 : Air kemasan bermerek

1 : Air isi ulang

2 : Leding meteran

3 : Leding eceran

4 : Sumur Bor/Pompa

5 : Sumur Terlindungi

6 : Sumur Tak Terlindungi

7 : Mata Air Terlindungi

8 : Mata Air Tak Terlindungi

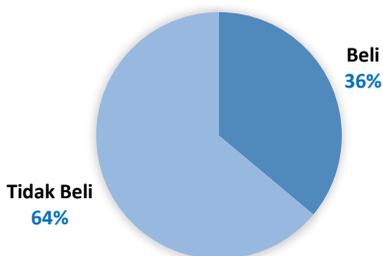
9 : Air Permukaan

10: Air Hujan

11: Lainnya

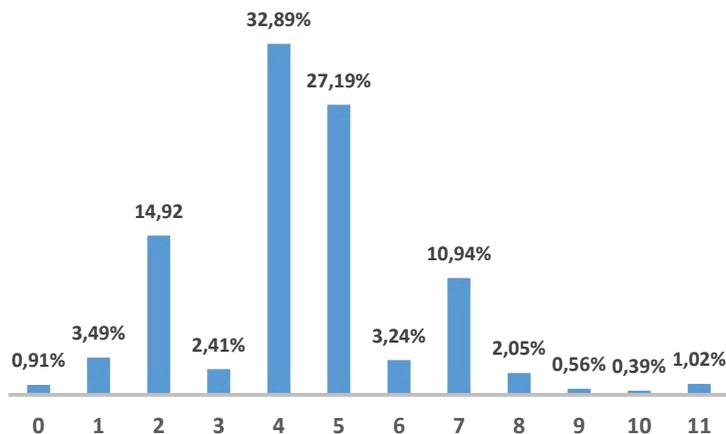
**Gambar 4.5** Sumber Air Minum Utama

Dari karakteristik sumber air minum yang telah dijelaskan sebelumnya maka kategori untuk sumber air minum dapat di *recoding* sehingga diperoleh informasi yang lebih informatif serta pengelompokan yang lebih efektif. Kategori baru setelah di *recoding* adalah sebagai berikut.



**Gambar 4.6** Kategori Baru Sumber Air Minum Utama

Gambar 4.6 menunjukkan bahwa terdapat dua kategori baru pada sumber air minum utama. Kategori pertama yaitu beli yang terdiri dari beberapa sumber air minum yaitu air kemasan bermerek, air isi ulang, leding meteran, dan leding eceran. Kategori kedua yang diperoleh yaitu tidak beli yang terdiri dari sumur bor/pompa, sumur terlindungi, sumur tak terlindungi, mata air terlindungi, mata air tak terlindungi, air permukaan, air hujan, dan air lainnya. Dari kategori baru tersebut dapat diketahui bahwa sebanyak 64% penduduk pada kabupaten/kota di Jawa Timur memperoleh sumber air minum dengan tidak membeli, sedangkan untuk 36% sisanya diperoleh dengan cara membeli.



Keterangan:

0 : Air kemasan bermerek

1 : Air isi ulang

2 : Leding meteran

3 : Leding eceran

4 : Sumur Bor/Pompa

5 : Sumur Terlindungi

6 : Sumur Tak Terlindungi

7 : Mata Air Terlindungi

8 : Mata Air Tak Terlindungi

9 : Air Permukaan

10 : Air Hujan

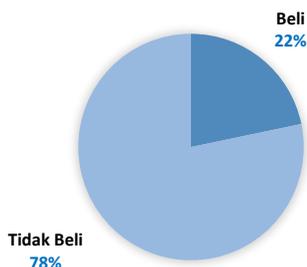
11 : Lainnya

**Gambar 4.7** Sumber Air Masak Utama

Gambar 4.7 menunjukkan bahwa terdapat 12 kategori pada variabel sumber air masak utama ( $X_5$ ). Sebanyak 32,89% penduduk di Provinsi Jawa Timur banyak menggunakan sumur

bor/pompa sebagai sumber utama untuk masak. Selanjutnya terbanyak kedua yaitu sumur terlindungi dengan persentase 27,19%, diikuti oleh air leding meteran (14,92%), mata air terlindungi (10,94%), air isi ulang (3,49%), sumur tak terlindungi (3,24%), leding eceran (2,41%), mata air tak terlindungi (2,05%), dan air lainnya (1,02%). Sumber air yang sedikit dikonsumsi oleh penduduk di Prinsi Jawa Timur yaitu air kemasan bermerek (0,91%), air permukaan (0,56%), dan air hujan (0,39%).

Dari karakteristik sumber air masak utama yang telah dijelaskan sebelumnya maka kategori untuk sumber air masak utama dapat di *recoding* sehingga diperoleh informasi yang lebih informatif serta pengelompokkan yang lebih efektif. Kategori baru setelah di *recoding* adalah sebagai berikut.

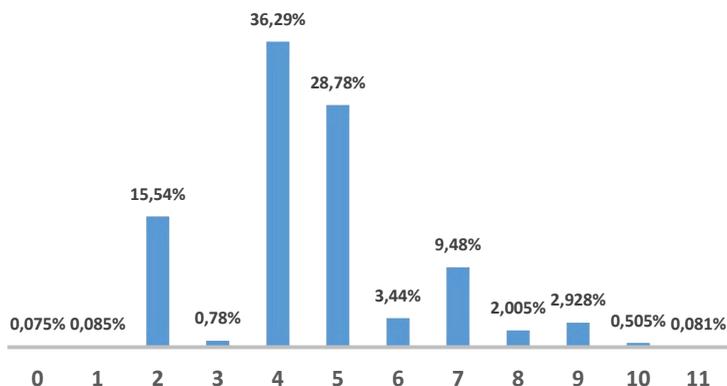


**Gambar 4.8** Kategori Baru Sumber Air Masak Utama

Gambar 4.8 menunjukkan bahwa terdapat dua kategori baru pada sumber air masak utama. Kategori pertama yaitu beli yang terdiri dari beberapa sumber air minum yaitu air kemasan bermerk, air isi ulang, leding meteran, dan leding eceran. Kategori kedua yang diperoleh yaitu tidak beli yang terdiri kategori sisanya. Dari kategori baru tersebut dapat diketahui bahwa sebanyak 78% penduduk pada kabupaten/kota di Jawa Timur memperoleh sumber air masak dengan tidak membeli, sedangkan untuk 22% sisanya diperoleh dengan cara beli.

Pada variabel sumber air mandi atau cuci utama ( $X_6$ ) terdapat 12 kategori. Dapat dilihat bahwa sebanyak 36,3%

penduduk di Provinsi Jawa Timur banyak menggunakan sumur bor/pompa sebagai sumber utama untuk mandi atau cuci. Selanjutnya terbanyak kedua yaitu sumur terlindungi dengan persentase 28,78%, diikuti oleh air leding meteran (15,54%), mata air terlindungi (9,48%), sumur tak terlindungi (3,44%), air permukaan (2,93%), mata air tak terlindungi (2,01%), leding eceran (0,78%), dan air hujan (0,51%). Sumber air yang sedikit dikonsumsi oleh penduduk di Prinsi Jawa Timur yaitu air isi ulang (0,09%), air lainnya (0,08%), dan air kemasan bermerek (0,075%). Kategori pada variabel sumber air mandi atau cuci utama dapat dilihat pada Gambar 4.9.



Keterangan:

0 : Air kemasan bermerek

1 : Air isi ulang

2 : Leding meteran

3 : Leding eceran

4 : Sumur Bor/Pompa

5 : Sumur Terlindungi

6 : Sumur Tak Terlindungi

7 : Mata Air Terlindungi

8 : Mata Air Tak Terlindungi

9 : Air Permukaan

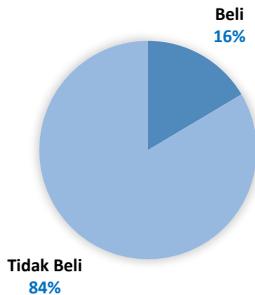
10: Air Hujan

11: Lainnya

**Gambar 4.9** Sumber Air Mandi/Cuci Utama

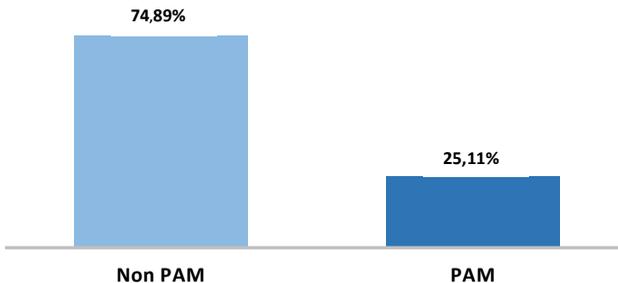
Dari karakteristik sumber air mandi atau cuci utama yang telah dijelaskan sebelumnya diperoleh persentase dari tiap kategori dan diperoleh hasil bahwa terdapat kategori yang sangat mendominasi ditunjukkan dengan persentase yang tinggi apabila

dibandingkan dengan persentase dengan kategori lainnya. Maka kategori untuk sumber air mandi atau cuci utama dapat di *recoding* sehingga diharapkan diperoleh informasi yang lebih informatif serta pengelompokkan yang lebih efektif. Kategori baru setelah di *recoding* adalah sebagai berikut.



**Gambar 4.10** Kategori Baru Sumber Air Mandi/Cuci Utama

Gambar 4.10 menunjukkan bahwa terdapat dua kategori baru pada sumber air mandi atau cuci utama. Kategori pertama yaitu beli yang terdiri dari beberapa sumber air mandi atau cuci yaitu air kemasan bermerk, air isi ulang, leding meteran, dan leding eceran. Kategori kedua yang diperoleh yaitu tidak beli yang terdiri dari kategori sisanya. Dari kategori baru tersebut dapat diketahui bahwa sebanyak 84% penduduk pada kabupaten/kota di Jawa Timur memperoleh sumber air mandi atau cuci dengan tidak beli, sedangkan untuk 16% sisanya diperoleh dengan cara beli.



**Gambar 4.11** Penggunaan Air PAM

Gambar 4.11 menunjukkan bahwa terdapat dua kategori untuk variabel penggunaan air PAM ( $X_7$ ) yaitu sebanyak 74,89% penduduk atau lebih dari setengah penduduk di Jawa Timur tidak menggunakan air PAM. Sedangkan untuk penggunaan air PAM pada kabupaten/kota di Jawa Timur yaitu sebanyak 25,11%.

## 4.2 Pengelompokkan Kabupaten atau Kota di Jawa Timur Menggunakan Metode *Ensemble ROCK*

Pengelompokkan kabupaten atau kota di Jawa Timur untuk data numerik dilakukan menggunakan metode hirarki *agglomerative* yang terdiri dari beberapa pautan yaitu *single linkage*, *complete linkage*, *average linkage*, dan *ward*. Pengelompokkan untuk data kategorik dilakukan menggunakan metode ROCK. Selanjutnya hasil pengelompokkan dari data numerik dan data kategorik tersebut digabung kemudian dikelompokkan kembali menggunakan metode *ensemble ROCK*.

### 4.2.1. Pengelompokkan Variabel Penelitian Berskala Numerik

Pengelompokkan kabupaten atau kota untuk data numerik menggunakan metode *single linkage*, *complete linkage*, *average linkage*, dan *ward* dengan ukuran jarak yang digunakan adalah jarak *euclidean*. Kabupaten atau kota dikelompokkan menjadi dua sampai lima klaster untuk masing-masing pautan. Dalam setiap pautan dipilih jumlah klaster optimum berdasarkan berdasarkan nilai *PseudoF*. *PseudoF* yang tinggi menunjukkan jumlah kelompok yang optimal. Nilai *PseudoF* untuk masing-masing metode dapat ditunjukkan pada Tabel 4.2.

**Tabel 4.2** Nilai *PseudoF* Hasil Pengelompokkan Data Numerik

Pautan	Jumlah Cluster	<i>PseudoF</i>
<i>Single Linkage</i>	2	15,60
	3	12,70
	4	105,19
	5	<b>194,48</b>
<i>Complete Linkage</i>	2	23,68
	3	121,76
	4	194,77

**Tabel 4.2** Nilai *PseudoF* Hasil Pengelompokan Data Numerik (Lanjutan)

<b>Pautan</b>	<b>Jumlah Cluster</b>	<b><i>PseudoF</i></b>	
<i>Complete Linkage</i>	5	<b>256,30</b>	
	2	15,60	
	<i>Average Linkage</i>	3	122,48
		4	<b>156,17</b>
		5	152,42
<i>Ward</i>	2	117,41	
	3	71,49	
	4	123,89	
	5	<b>160,44</b>	

Dari keempat metode tersebut dapat diketahui nilai *PseudoF* pada masing-masing metode. Jumlah kelompok optimum untuk metode *single*, *complete*, dan *ward* adalah sebanyak 5 kelompok, sedangkan untuk metode *average* adalah sebanyak 4 kelompok.

Setelah mendapatkan jumlah kelompok optimum maka selanjutnya adalah memilih metode pengelompokan yang terbaik berdasarkan nilai rasio  $S_w$  dan  $S_B$  terkecil dari masing-masing metode. Nilai rasio untuk masing-masing metode adalah sebagai berikut.

**Tabel 4.3** Nilai Rasio Pengelompokan Variabel Penelitian Berskala Numerik

<b>Metode</b>	<b>Jumlah Kelompok</b>	$S_w$	$S_B$	<b>Rasio</b>
<i>Single Linkage</i>	5	35406,45	561305,8	<b>0,063078</b>
<i>Complete Linkage</i>	5	42048,29	562596,8	0,074739
<i>Average Linkage</i>	4	54589,77	606397,3	0,090023
<i>Ward</i>	5	41329,18	541830,0	0,076277

Tabel 4.3 menunjukkan bahwa nilai rasio  $S_w$  dan  $S_B$  terkecil yaitu pada metode *single linkage* dengan nilai sebesar 0,063078. Dengan demikian didapatkan pengelompokan optimum kabupaten atau kota di Jawa Timur menggunakan pautan *single linkage* dengan jumlah cluster sebanyak 5 kelompok. Hasil pengelompokan kabupaten/kota di Jawa Timur adalah sebagai berikut.

**Tabel 4.4** Hasil Pengelompokkan Variabel Penelitian Berskala Numerik

<b>Kelompok</b>	<b>Anggota</b>
Kelompok 1	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Sumenep
Kelompok 2	Sidoarjo
Kelompok 3	Mojokerto, Lamongan
Kelompok 4	Kabupaten Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, dan Kota Batu
Kelompok 5	Kota Surabaya

Tabel 4.4 menunjukkan bahwa kelompok pertama terdiri dari 25 kabupaten, 1 kabupaten dalam kelompok kedua, 2 kabupaten dalam kelompok ketiga, 9 kabupaten atau kota dalam kelompok keempat, dan 1 kota masuk dalam kelompok kelima. Karakter hasil pengelompokkan pada variabel penelitian berskala numerik untuk masing-masing kelompok adalah sebagai berikut.

**Tabel 4.5** Karakteristik Hasil Pengelompokkan Variabel Berskala Numerik

<b>Variabel</b>	<b>Kel 1</b>	<b>Kel 2</b>	<b>Kel 3</b>	<b>Kel 4</b>	<b>Kel 5</b>
X <sub>1</sub> (ribu)	0,734	3,39	1,09	4,4	8,16
X <sub>2</sub>	81,53	43,23	73,84	64,86	43,42
X <sub>3</sub> (juta)	2,39	5,27	3,34	4,26	6,74

Tabel 4.5 menunjukkan karakteristik hasil pengelompokan pada masing-masing kelompok. Rata-rata terbesar pada variabel kepadatan penduduk (X<sub>1</sub>) berada pada kelompok 5 sedangkan rata-rata kepadatan terendah berada pada kelompok 1. Pada variabel persentase tingkat pendidikan terakhir kepala RT dibawah SMA (X<sub>2</sub>) terlihat bahwa rata-rata tertinggi berada pada kelompok 1 sedangkan rata-rata terendah berada pada kelompok 2. Dapat

diketahui bahwa kabupaten/kota yang berada di kelompok 5 memiliki tingkat perekonomian yang sangat tinggi dilihat dari pengeluaran sebulan terakhir ( $X_3$ ) dibandingkan dengan kelompok lainnya. Pada persentase tingkat pendidikan terakhir kepala RT yang rendah menunjukkan bahwa sebagian besar kepala RT telah menempuh pendidikan diatas SMA.

Kelompok 2 memiliki tingkat perekonomian yang tinggi dapat dilihat dari pengeluaran sebulan, memiliki persentase tingkat pendidikan terakhir kepala RT dibawah SMA yang rendah serta kepadatan penduduk yang rendah dibandingkan kelompok 5. Pada kelompok 4 memiliki tingkat perekonomian yang sedang dibandingkan dengan kelompok lainnya. Kelompok 3 termasuk dalam tingkat perekonomian yang rendah, serta pada kelompok 1 memiliki tingkat perekonomian yang sangat rendah dibandingkan dengan kelompok lainnya yang dapat dilihat dari pengeluaran sebulan terakhir yang rendah, memiliki kepadatan penduduk yang rendah, serta persentase tingkat pendidikan terakhir kepala RT dibawah SMA yang cukup tinggi yang menunjukkan bahwa sebaian besar kepala RT hanya menempuh pendidikan dibawah SMA.

#### **4.2.2. Pengelompokkan Variabel Penelitian Berskala Kategorik**

Data kategorik pada penelitian ini terdiri dari beberapa variabel yaitu sumber air minum utama, sumber air masak, sumber air mandi atau cuci, dan penggunaan PAM. Berdasarkan variabel tersebut akan dilakukan pengelompokkan kabupaten atau kota di Jawa Timur menggunakan metode ROCK. Dalam penelitian ini digunakan nilai *threshold* sebesar 0,05; 0,1; 0,15; 0,2; 0,25; 0,3; 0,35; 0,4; 0,45; dan 0,5. Setelah dilakukan pengelompokkan selanjutnya memilih jumlah kelompok optimum berdasarkan nilai rasio antara simpangan baku dalam kelompok ( $S'_w$ ) dan simpangan baku antar kelompok ( $S'_B$ ). Semakin kecil nilai rasio yang didapatkan menunjukkan bahwa terdapat homogenitas maksimum dalam kelompok dan heterogenitas maksimum antar kelompok. Nilai rasio pada pengelompokkan data numerik ditunjukkan oleh Tabel 4.6.

**Tabel 4.6** Nilai Rasio Pengelompokan Data Kategorik

<b>Threshold (<math>\theta</math>)</b>	<b>Klaster</b>	<b>Rasio</b>
0,05	2	$4,85 \times 10^{-2}$
0,1	2	0,000
0,15	3	$2,21 \times 10^{-16}$
0,2	3	$2,21 \times 10^{-16}$
<b>0,25</b>	<b>2</b>	<b><math>6,83 \times 10^{-17}</math></b>
<b>0,3</b>	<b>2</b>	<b><math>6,83 \times 10^{-17}</math></b>
0,35	4	$7,26 \times 10^{-2}$
0,4	4	$0,19 \times 10$
0,45	4	$0,19 \times 10$
0,5	4	$0,19 \times 10$

Nilai rasio terkecil yang ditunjukkan oleh Tabel 4.6 adalah pengelompokan nilai  $\theta$  sebesar 0,1. Namun hasil tersebut tidak dapat digunakan, karena tidak sesuai dengan hasil yang diharapkan dimana hanya terbentuk satu kelompok yang sama untuk seluruh objek pengamatan. Maka dari itu dipilih rasio terkecil kedua yaitu dengan nilai  $\theta$  sebesar 0,25 dan 0,30 dengan nilai rasio yang sama sebesar  $6,83 \times 10^{-17}$ . Hasil pengelompokan menggunakan metode ROCK dengan  $\theta$  sebesar 0,25 maupun 0,30 menghasilkan kelompok yang sama yaitu 2 kelompok. Daftar anggota dari setiap cluster hasil pengelompokan kategorik dapat dilihat pada Tabel 4.7 berikut.

**Tabel 4.7** Hasil Pengelompokan Variabel Penelitian Berskala Kategorik

<b>Kelompok</b>	<b>Anggota</b>
Kelompok 1	Kota Malang, Kota Madiun, Kota Surabaya
Kelompok 2	Pacitan, Trenggalek, Tulungagung, Blitar, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Magetan, Ngawi, Bangkalan, Sampang, Pamekasan, Sumenep, Blitar, Ponorogo, Kediri, Situbondo,
	Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Bojonegoro, Tuban, Lamongan, Gresik, Kota Kediri, Kota Blitar, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Batu

Tabel 4.7 menunjukkan bahwa sebanyak 3 kota masuk dalam cluster pertama dan 35 kabupaten/kota masuk dalam cluster kedua. Karakter hasil pengelompokan pada variabel penelitian berskala kategorik untuk masing-masing kelompok adalah sebagai berikut.

**Tabel 4.8** Karakteristik Hasil Pengelompokan Variabel Penelitian Berskala Kategorik

Variabel		Kel 1	Kel 2
Sumber Air Minum Utama	Beli	<b>86,32%</b>	13,68%
	Tidak Beli	32,99%	<b>67,01%</b>
Sumber Air Masak Utama	Beli	<b>71,46%</b>	17,95%
	Tidak Beli	28,54%	<b>82,04%</b>
Sumber Air Mandi atau Cuci	Beli	<b>64,4%</b>	12,72%
	Tidak Beli	35,6%	<b>87,28%</b>
Penggunaan Air PAM	Non PAM	40,96%	<b>77,50%</b>
	PAM	<b>59,04%</b>	22,49%

Pada Tabel 4.8 menyajikan persentase dari kategori pada setiap variabel-variabel yang ada dalam kelompok. Pada kelompok 1 terlihat bahwa sebagian besar penduduk untuk sumber air utama ( $X_4$ ), sumber air masak ( $X_5$ ), dan sumber air mandi atau cuci ( $X_6$ ) berasal dari air beli yang dapat dilihat dari masing-masing persentase berturut-turut sebesar 86,32%, 71,46%, dan 64,4%. Pada kelompok 1 untuk variabel penggunaan air PAM ( $X_7$ ) sebagian besar penduduk pada kabupaten/kota yang terbentuk menggunakan air PAM yang ditunjukkan oleh nilai sebesar 59,04%. Pada kelompok 2 sebagian besar penduduk untuk sumber air utama, sumber air masak utama, dan sumber air minum berasal dari air yang tidak beli. Selanjutnya pada variabel penggunaan air PAM yaitu pada kelompok 2 sebanyak 77,5% penduduk pada kabupaten atau kota yang terbentuk tidak menggunakan air PAM.

#### **4.2.3. Pengelompokan Variabel Penelitian Berskala Numerik dan Kategorik Menggunakan Metode *Ensemble* ROCK**

Hasil pengelompokan pada data numerik dan kategorik dilakukan dengan menggunakan pengelompokan metode

*ensemble* ROCK. Hasil pengelompokan data campuran yang telah didapatkan dari pembahasan sebelumnya dianggap sebagai variabel baru yang memiliki skala data kategorik. Hasil pengelompokan data numerik menghasilkan empat kategori dan hasil pengelompokan data kategorik menghasilkan dua kategori. Struktur data baru dapat dilihat pada Tabel 4.9.

Tabel 4.9 Struktur Data dengan Variabel Baru

<b>Kab/kota</b>	<b>Output cluster Numerik</b>	<b>Output cluster Kategorik</b>	<b>Kab/kota</b>	<b>Output cluster Numerik</b>	<b>Output cluster Kategorik</b>
Pacitan	1	2	Pamekasan	1	2
Ponorogo	1	2	Sumenep	1	2
Trenggalek	1	2	Kediri	4	2
Tulungagung	1	2	Blitar	4	2
Blitar	1	2	Malang	4	1
Kediri	1	2	Probolinggo	4	2
Malang	1	2	Pasuruan	4	2
Lumajang	1	2	Mojokerto	4	2
Jember	1	2	Madiun	4	1
Banyuwangi	1	2	Surabaya	5	1
Bondowoso	1	2	Batu	4	2
Situbondo	1	2			
Probolinggo	1	2			
Pasuruan	1	2			
Sidoarjo	2	2			
Mojokerto	3	2			
Jombang	1	2			
Nganjuk	1	2			
Madiun	1	2			
Magetan	1	2			
Ngawi	1	2			
Bojonegoro	1	2			
Tuban	1	2			
Lamongan	3	2			
Gresik	4	2			
Bangkalan	1	2			
Sampang	1	2			

Setelah mendapatkan hasil pengelompokan maka selanjutnya kembali melakukan pengelompokan kabupaten atau kota di Jawa Timur dengan menggunakan metode ROCK. Nilai *threshold* yang digunakan tetap sebesar 0,05; 0,1; 0,15; 0,2; 0,25; 0,3; 0,35; 0,4; 0,45; dan 0,5 namun pada nilai 0,35; 0,4; 0,45; dan 0,5 tidak dapat mengelompokkan pengamatan sehingga nilai *threshold* yang dipakai yaitu 0,05 sampai dengan 0,3. Dari keenam nilai *threshold* yang digunakan akan dipilih jumlah cluster optimum berdasarkan rasio dari  $S'_w$  dengan  $S'_B$  terkecil.

**Tabel 4.10** Rasio  $S_w$  dan  $S_B$  Pengelompokan Metode Ensemble ROCK

<i>Threshold</i> ( $\theta$ )	Rasio $S'_w$ dan $S'_B$
0,05	$4,82 \times 10^{-11}$
0,1	$4,82 \times 10^{-11}$
0,15	$6,92 \times 10^{-11}$
0,2	$6,92 \times 10^{-11}$
0,25	$6,92 \times 10^{-11}$
0,3	$6,92 \times 10^{-11}$

Berdasarkan Tabel 4.10 menunjukkan bahwa nilai rasio terkecil pada pengelompokan *ensemble* ROCK adalah pengelompokan dengan  $\theta$  sebesar 0,05 dan 0,1 yang menghasilkan nilai rasio yang sama bernilai  $4,82 \times 10^{-11}$ . Jumlah kelompok yang terbentuk yaitu sebanyak 2 kelompok. Daftar anggota dari setiap cluster hasil pengelompokan kategorik dapat dilihat pada Tabel 4.11.

**Tabel 4.11** Hasil Pengelompokan dengan Metode *Ensemble* ROCK

Kelompok	Anggota
Kelompok 1	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Bangkalan, Sampang, Pamekasan, Sumenep

**Tabel 4.12** Hasil Pengelompokkan dengan Metode *Ensemble ROCK* (Lanjutan)

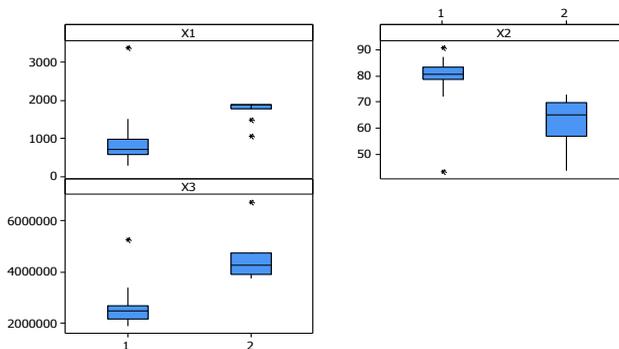
Kelompok	Anggota
Kelompok 2	Kabupaten Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu

Tabel 4.11 menunjukkan bahwa sebanyak 28 kabupaten masuk dalam cluster pertama dan 10 kabupaten atau kota masuk dalam cluster kedua. Karakter hasil pengelompokkan pada data numerik untuk masing-masing kelompok adalah sebagai berikut.

**Tabel 4.13** Karakteristik Hasil Pengelompokkan Data Numerik pada Metode *Ensemble ROCK*

Variabel	Kel 1	Kel 2
$X_1$ (ribu)	0,85	4,78
$X_2$	79,61	62,71
$X_3$ (juta)	2,6	4,5

Tabel 4.12 menunjukkan karakteristik hasil pengelompokkan pada masing-masing kelompok. Kelompok 1 memiliki rata-rata tertinggi pada variabel persentase tingkat pendidikan terakhir kepala RT dibawah SMA ( $X_2$ ) sedangkan kelompok 2 memiliki rata-rata tertinggi pada variabel kepadatan penduduk ( $X_1$ ) dan rata-rata pengeluaran sebulan terakhir ( $X_3$ ).

**Gambar 4.12** Karakteristik Hasil Pengelompokkan Data Numerik pada Hasil Pengelompokkan *Ensemble ROCK*

Gambar 4.12 menunjukkan *boxplot* dari anggota kelompok pada kelompok 1 dan 2. Terlihat bahwa pada tiap variabel untuk masing-masing kelompok terdapat *outlier* kecuali pada variabel persentase pendidikan terakhir kepala RT di bawah SMA ( $X_2$ ) untuk kelompok 2. Karakter hasil pengelompokan pada data kategorik untuk masing-masing kelompok adalah sebagai berikut.

**Tabel 4.14** Karakteristik Hasil Pengelompokan Data Kategorik pada Metode *Ensemble ROCK*

Variabel		Kel 1	Kel 2
Sumber Air Minum Utama	Beli	27,38%	<b>64,70%</b>
	Tidak Beli	<b>72,62%</b>	35,30%
Sumber Air Masak Utama	Beli	14,96%	42,41%
	Tidak Beli	<b>85,04%</b>	<b>57,59%</b>
Sumber Air Mandi atau Cuci	Beli	11,20%	32,50%
	Tidak Beli	<b>88,80%</b>	<b>67,50%</b>
Penggunaan Air PAM	Non PAM	<b>80,38%</b>	<b>58,49%</b>
	PAM	19,62%	41,50%

Pada Tabel 4.13 menyajikan persentase dari kategori pada setiap variabel-variabel yang ada dalam kelompok. Pada kelompok 1 terlihat sebagian besar sumber air minum ( $X_4$ ), sumber air masak ( $X_5$ ), dan sumber mandi/cuci ( $X_6$ ) yang dipakai berasal dari air tidak beli, serta pada variabel penggunaan air PAM ( $X_7$ ) sebagian besar penduduk pada kabupaten/kota yang terbentuk tidak menggunakan PAM air yang ditunjukkan dengan persentase sebesar 80,38%. Pada kelompok 2 terlihat bahwa pada variabel sumber air minum, sumber air masak, dan sumber air mandi atau cuci berasal dari air beli serta sebagian besar penduduk tidak menggunakan air PAM yang ditunjukkan dengan persentase sebesar 58,49%.

Setelah diperoleh kelompok yang optimum maka selanjutnya adalah melakukan uji beda rata-rata antar kelompok untuk mengetahui apakah kelompok yang dihasilkan berbeda secara signifikan atau tidak. Oleh karena itu dilakukan pengujian

beda rata-rata antar kelompok menggunakan uji *Kruskal Wallis* pada 7 variabel terhadap 2 kelompok yang terbentuk.

**Tabel 4.15** Hasil Pengujian *Kruskal Wallis*

Variabel	Chi-Square	P-value
Kepadatan Penduduk	17,053	0,000
Persentase Tingkat Pendidikan Terakhir Kepala RT dibawah SMA	18,004	0,000
Rata-rata Pengeluaran dalam sebulan	18,858	0,000
Sumber Air Minum Utama	15,683	0,000
Sumber Air Masak Utama	15,697	0,000
Sumber Air Mandi / Cuci Utama	8,880	0,003
Penggunaan Air PAM	5,321	0,021

Tabel 4.14 menunjukkan hasil pengujian perbedaan pengelompokan terhadap 7 variabel. Dapat diketahui bahwa seluruh variabel yang digunakan dalam analisis memiliki nilai *p-value* yang lebih kecil dari  $\alpha$  (0,05) serta memiliki nilai  $\chi_{hitung}$  lebih besar daripada  $\chi_{(0,05,2-1)}$  yaitu 3,841 yang artinya tujuh variabel tersebut signifikan berbeda antara kabupaten/kota pada kelompok pertama dan kelompok kedua.

Hasil pengelompokan data campuran menggunakan metode *ensemble* ROCK menghasilkan 2 kelompok dengan karakteristik setiap kelompok sebagai berikut,

a. Kelompok Perekonomian Rendah

Kelompok 1 beranggotakan 28 kabupaten (dapat dilihat pada Tabel 4.11) dapat dikatakan bahwa memiliki perekonomian rendah yang dapat dilihat dari pengeluaran sebulan terakhir yang lebih rendah dibandingkan kelompok 2. Sebagian besar penduduk menggunakan sumber air minum, sumber air masak, serta sumber air mandi atau cuci diperoleh dengan cara tidak membeli. Sebagian besar penduduk juga tidak menggunakan air PAM. Dapat dikatakan bahwa sebagian besar wilayah tersebut memiliki tingkat konsumsi air bersih yang rendah sehingga belum tercukupi dalam mendapatkan akses air bersih. Kepadatan

penduduk pada kelompok ini sangat rendah namun persentase tingkat pendidikan terakhir kepala RT dibawah SMA sangat tinggi yang artinya sebagian besar penduduk hanya menempuh pendidikan dibawah SMA.

b. **Kelompok Perekonomian Tinggi**

Kelompok 2 beranggotakan 9 kota dan 1 kabupaten (dapat dilihat pada Tabel 4.11) dapat dikatakan bahwa memiliki perekonomian tinggi yang dapat dilihat dari pengeluaran sebulan terakhir yang tinggi dibandingkan kelompok 1. Sebagian besar penduduk menggunakan sumber air minum, sumber air masak, serta sumber air mandi atau cuci yang diperoleh dengan cara membeli. Sebagian besar penduduk juga menggunakan air PAM. Dapat dikatakan bahwa sebagian besar wilayah tersebut memiliki tingkat konsumsi air bersih yang tinggi sehingga sudah tercukupi dalam mendapatkan akses air bersih. Kepadatan penduduk pada kelompok ini sangat tinggi serta persentase tingkat pendidikan terakhir kepala RT dibawah SMA yaitu rendah yang artinya sebagian besar penduduk telah menempuh pendidikan diatas SMA.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1. Kesimpulan**

1. Kabupaten atau kota di Jawa Timur terdiri dari 29 kabupaten dan 9 kota. Masing-masing kabupaten atau kota memiliki karakteristik berbeda-beda yang memiliki tipe data numerik dan tipe data kategorik. Pada variabel numerik terdiri dari 3 variabel yang terdiri dari kepadatan penduduk ( $X_1$ ), persentase tingkat pendidikan terakhir kepala RT dibawah SMA ( $X_2$ ), dan rata-rata pengeluaran sebulan terakhir ( $X_3$ ). Pada data kategorik terdapat 4 variabel yaitu sumber air minum ( $X_4$ ), sumber air masak ( $X_5$ ), sumber air mandi/cuci ( $X_6$ ), dan penggunaan air PAM ( $X_7$ ). Variabel kepadatan penduduk pada kabupaten/ kota di Jawa Timur memiliki rata-rata kepadatan penduduk sebanyak 1890 penduduk per km<sup>2</sup>. Variabel persentase tingkat pendidikan terakhir kepala RT dibawah SMA pada kabupaten/kota di Jawa Timur memiliki rata-rata persentase tingkat pendidikan sebesar 75,17% serta variabel pengeluaran dalam sebulan terakhir pada kabupaten/kota di Jawa Timur memiliki rata-rata pengeluaran yaitu Rp 3.080.460,-. Pada variabel sumber air minum ( $X_4$ ) dapat diketahui bahwa sebanyak 64% penduduk pada kabupaten/kota di Jawa Timur memperoleh sumber air minum dengan tidak membeli, sedangkan untuk 36% sisanya diperoleh dengan cara membeli. Pada variabel sumber air masak ( $X_5$ ) diketahui bahwa sebanyak 78% penduduk pada kabupaten/kota di Jawa Timur memperoleh sumber air masak dengan tidak membeli, sedangkan untuk 22% sisanya diperoleh dengan cara beli. Variabel sumber air mandi/cuci ( $X_6$ ) diketahui sebanyak 84% penduduk pada kabupaten/kota di Jawa Timur memperoleh sumber air mandi atau cuci dengan tidak beli, sedangkan untuk 16% sisanya diperoleh dengan cara beli. Lalu untuk variabel penggunaan air PAM ( $X_7$ ) yaitu sebanyak 74,89% penduduk di Jawa Timur tidak menggunakan air PAM dimana terdapat lebih dari setengah

penduduk pada kabupaten/kota di Jawa Timur. Sedangkan untuk penggunaan air PAM pada kabupaten/kota di Jawa Timur yaitu sebanyak 25,11%.

2. Hasil pengelompokkan kabupaten/kota di Jawa Timur menggunakan metode Ensembl ROCK menghasilkan pengelompokkan sebagai berikut,
  - a. Pada pengelompokkan data numerik diperoleh nilai *pseudo-f* pada masing-masing pautan lalu dihasilkan nilai rasio  $S_w$  dan  $S_b$  terkecil yaitu pada metode *single linkage* dengan nilai sebesar 0,063078 dengan kelompok yang terbentuk sejumlah 5 kelompok dengan masing-masing anggota yaitu kelompok pertama sebanyak 25 kabupaten/kota, kelompok 2 dengan 1 anggota, kelompok 3 sejumlah 2 anggota, kelompok 4 dengan 1 anggota, dan kelompok 5 sejumlah 9 anggota kabupaten/kota
  - b. Hasil pengelompokkan pada data kategorik menggunakan metode ROCK diperoleh nilai rasio terkecil dengan *threshold* ( $\theta$ ) 0,25 dan 0,30 menghasilkan 2 kelompok yang sama dengan nilai rasio yang sama sebesar  $6,83 \times 10^{-17}$  dimana sebanyak 3 kabupaten atau kota masuk dalam kelompok pertama dan 35 kabupaten atau kota masuk dalam kelompok kedua.
  - c. Nilai rasio terkecil pada pengelompokkan *ensemble* ROCK adalah sebesar  $6,79 \times 10^{-17}$  dengan nilai  $\theta$  sebesar 0,05 dan 0,1 yang menghasilkan nilai rasio yang sama. Jumlah kelompok yang terbentuk sebanyak 2 kelompok dimana sebanyak 28 kabupaten masuk dalam cluster pertama dan 10 kabupaten atau kota masuk dalam cluster kedua.

## 5.2. Saran

Saran yang dapat diberikan berdasarkan hasil penelitian ini adalah

1. Dalam penelitian ini menggunakan 3 variabel numerik dan 4 variabel kategorik namun sebenarnya masih banyak variabel yang dapat digunakan, tetapi karena keterbatasan data sekunder maka tidak dapat menggunakan banyak variabel. Sehingga untuk

penelitian selanjutnya dalam melakukan pengelompokan agar menambahkan variabel pengamatan.

2. Untuk mencapai target pada tahun 2030 dimana setiap penduduk dapat mengakses air bersih, kabupaten/kota yang memiliki karakteristik perekonomian rendah yang tercermin dari sumber air yang diperoleh dengan tidak membeli, perlu mendapat perhatian khusus dari pemerintah Jawa Timur supaya akses air bersih pada wilayah tersebut dapat terpenuhi.

*(Halaman ini sengaja dikosongkan)*

## DAFTAR PUSTAKA

- Alvionita. (2017). *Metode Ensemble ROCK dan SWFM Untuk Pengelompokkan Data Campuran Numerik dan Kategorik pada Kasus Akses Jeruk*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Bunkers, M. J., & James, R. M. (1996). Definition of Climate Region in The Northern Plains Using an Objective Cluster Modification Technique. *Journal of Climate*, 130-146.
- Daniel, W. (1989). *Statistika Nonparametrik Terapan*. Jakarta: PT. Gramedia.
- Darr, P., S.L. Feldman, & C. Kamen. (1976). *The Demand for Urban Water*. Martinus Nijhoff Social Division. Leiden.
- Dewi, A. (2012). *Metode Cluster Ensemble untuk Pengelompokkan Desa Pedesaan di Provinsi Riau*. Surabaya: Statistika: Institut Teknologi Sepuluh Nopember.
- Dutta, M., Mahanta, A. K., & Pujari, A. K. (2005). QROCK: A Quick of the ROCK Algorithm for Clustering of Categorical Data. *Proceedings of the 15 IEEE International Conference on Data Engineering*.
- Gabriel, J. F. (2001). *Fisika Lingkungan*. Jakarta: Hipokrates.
- Greene, W. H. (2008). *Econometric Analysis*. New Jersey: Prentice Hall.
- Guha, S., Rastogi, R., & Shim, K. (2000). ROCK: A Robust Clustering Algorithm for Categorical Attributes. *Proceedings of the 15th International Conference on Data Engineering*.
- Hair, J. F., Black, W. C., Anderson, R. E., & Tatham, R. E. (2009). *Multivariate Data Analysis 6th Edition*. New Jersey: Prentice-Hall International, Inc.
- Halkidi, Batistakis, & Vizirgiannis. (2001). On Clustering Validation Techniques. *Journal of Intelligent Systems*, 17, 107-145.

- He, Z., Xu, X. I., & Deng, S. (2005a). A Cluster Ensemble Method for Clustering Categorical Data. *Information Fusion*, 143-151.
- He, Z., Xu, X. I., & Deng, S. (2005b). Clustering Mixed Numeric and Categorical Data: A Cluster Ensemble Approach. *Department of Computer Science and Engineering Harbin Institute of Technology*.
- Indrajani. (2015). *Database Design*. Jakarta: PT. Elex Media Komputindo.
- Johnson, R. A., & Winchern, D. W. (2007). *Applied Multivariate Statistical Analysis* (7th ed.). New Jersey: Pearson Education Inc.
- Junaidi, J. (2014). *Deskripsi Data Melalui Box-Plot*. Fakultas Ekonomi dan Bisnis: Universitas Jambi.
- Keynes, J. M. (1936). *The General Theory of Employment Interest and Money*. Harcourt: Brace and World.
- Linsley, K. R. (1995). *Teknik Sumber Daya Air* (2nd ed.). Jakarta: Erlangga.
- Orpin, A. R., & Kostylev, V. E. (2006). Towards a statistically valid method of textural sea floor characterization of benthic habitats. *Marine Geology*, 225, 209:222.
- Ramdhany, R. (2017). *Pengelompokkan Desa di Kabupaten Bondowoso Berdasarkan Data Campuran Numerik dan Kategorik Menggunakan Metode Ensemble ROCK*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Sharma, S. (1996). *Applied Multivariate Technique*. New York: John Wiley and Sons, Inc.
- Tyagi, A., & Sharma, S. (2012). Implementation of ROCK CLustering Algorithm for the Optimization of Query Searching Time. *International Journal on Computer Science and Engineering*, 4, 809-815.
- 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.
- Yoon, H. S., Ahn, S. Y., Lee, S. H., Cho, S. B., & Kim, J. H. (2006). Heterogeneous Clustering Ensemble Method For Combining Different Cluster Result. *BioDM 2006*, 82-91.

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN

### Lampiran 1 Data Penelitian

No.	Kab/Kota	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
1	PACITAN	397	81.72	2433000	1	1	1	0
2	PONOROGO	665	80.85	2155049	1	1	1	0
3	TRENGGALEK	603	83.31	2376891	1	1	1	0
4	TULUNGAGUNG	972	78.47	2875572	1	1	1	0
5	BLITAR	860	79.78	2522493	1	1	1	0
6	KEDIRI	1121	73.80	2690661	1	1	1	0
7	MALANG	725	78.00	2565434	1	1	1	0
8	LUMAJANG	577	85.70	2541569	1	1	1	0
9	JEMBER	782	82.63	2052029	1	1	1	0
10	BANYUWANGI	277	79.61	2499721	1	1	1	0
11	BONDOWOSO	501	83.02	1900431	1	1	1	0
12	SITUBONDO	403	78.87	2047976	1	1	1	0
13	PROBOLINGGO	677	83.33	2005621	1	1	1	0
14	PASURUAN	1081	80.79	2873076	1	1	1	0
15	SIDOARJO	3390	43.23	5270614	0	1	1	1
16	MOJOKERTO	1519	72.22	3411942	1	1	1	0
17	JOMBANG	1119	72.20	2462072	1	1	1	0
18	NGANJUK	854	79.65	2264976	1	1	1	0
19	MADIUN	653	80.53	2507390	1	1	1	0
20	MAGETAN	912	79.97	2689796	1	1	1	0
21	NGAWI	640	83.44	1966729	1	1	1	0
22	BOJONEGORO	564	84.96	2324098	1	1	1	0
23	TUBAN	632	84.35	2563297	1	1	1	0
24	LAMONGAN	667	75.45	3259242	0	1	1	0
25	GRESIK	1067	59.64	4755713	0	0	1	0
26	BANGKALAN	961	87.27	2747079	1	1	1	0

**Lampiran 1** Data Penelitian (Lanjutan)

<b>No.</b>	<b>Kab/Kota</b>	<b>X<sub>1</sub></b>	<b>X<sub>2</sub></b>	<b>X<sub>3</sub></b>	<b>X<sub>4</sub></b>	<b>X<sub>5</sub></b>	<b>X<sub>6</sub></b>	<b>X<sub>7</sub></b>
27	SAMPANG	768	90.77	2440898	1	1	1	0
28	PAMEKASAN	1078	80.45	2156465	1	1	1	0
29	SUMENEP	539	84.81	2327545	1	1	1	0
30	KEDIRI	4448	69.00	4023833	1	1	1	0
31	BLITAR	4271	66.74	3942890	1	1	1	0
32	MALANG	5895	48.77	4773217	0	0	0	0
33	PROBOLINGGO	4078	64.46	3923567	0	1	1	0
34	PASURUAN	5560	72.84	3758081	0	0	1	0
35	MOJOKERTO	7675	64.38	4351524	0	1	1	0
36	MADIUN	5177	65.92	4615399	0	0	0	1
37	SURABAYA	8166	43.42	6743623	0	0	0	1
38	BATU	1480	71.98	4237965	1	1	1	1

## Lampiran 2 Syntax Analisis Pengelompokan Data Numerik

```

Numerik #Pautan single linkage
#Pautan single linkage
kluster.single = function(data,k){
datanumerik = data.frame(data$X1, data$X2, data$X3)
d = dist(datanumerik, method = "euclidean")
fit.sin = hclust(d, method = "single")
single = cutree(fit.sin, k=k)
hasil.kelompok.single = data.frame(data$kab/kota, single)
print(fit.sin)
print(single)
print(hasil.kelompok.single)
}
#Pautan complete linkage
kluster.complete = function(data,k){
datanumerik = data.frame(data$X1, data$X2, data$X3)
d = dist(datanumerik, method = "euclidean")
fit.sin = hclust(d, method = "complete")
complete = cutree(fit.sin, k=k)
hasil.kelompok.complete = data.frame(data$Desa, complete)
print(fit.sin)
print(complete)
print(hasil.kelompok.complete)
}

# pautan average linkage
kluster.average = function(data,k){
datanumerik = data.frame(data$X1, data$X2, data$X3)
d = dist(datanumerik, method =
"euclidean") fit.sin = hclust(d, method =
"average") average = cutree(fit.sin, k=k)

```

**Lampiran 2** *Syntax* Analisis Pengelompokan Data Numerik  
(Lanjutan)

```
hasil.kelompok.average = data.frame(data$Desa, average)
print(fit.sin)
print(average)
print(hasil.kelompok.average)
}
```

### Lampiran 3 Syntax Nilai Rasio Analisis Kelompok Hirarki *Agglomerative*

```

datanumerik = data.frame(data$X1, data$X2, data$X3)
d = dist(datanumerik, method = "euclidean")
#Analisis Kluster Hirarki
fit.sin = hclust(d, method = "single")
fit.com = hclust(d, method = "complete")
fit.ave = hclust(d, method = "average")
fit.ward=hclust(d, method = "ward.D")

#Memotong Dendogram untuk k kluster
single=cutree(fit.sin, k=5)
complete=cutree(fit.com, k=5)
average=cutree(fit.ave, k=4)
ward=cutree(fit.ward, k=5)
hasilkelompoknumerik=data.frame(data$Kabupaten,single, complete,
average,ward)
datanumerik2=cbind(data$X1, data$X2, data$X3)
data1 = mean(datanumerik2[1,])
data2 = mean(datanumerik2[2,])
data3 = mean(datanumerik2[3,])
data4 = mean(datanumerik2[4,])
data5 = mean(datanumerik2[5,])
data6 = mean(datanumerik2[6,])
data7 = mean(datanumerik2[7,])
data8 = mean(datanumerik2[8,])
data9 = mean(datanumerik2[9,])
data10 = mean(datanumerik2[10,])
data11 = mean(datanumerik2[11,])
data12 = mean(datanumerik2[12,])
data13 = mean(datanumerik2[13,])
data14 = mean(datanumerik2[14,])
data15 = mean(datanumerik2[15,])
data16 = mean(datanumerik2[16,])
data17 = mean(datanumerik2[17,])
data18 = mean(datanumerik2[18,])
data19 = mean(datanumerik2[19,])
data20 = mean(datanumerik2[20,])

```

```

data21 = mean(datanumerik2[21,])
data22 = mean(datanumerik2[22,])
data23 = mean(datanumerik2[23,])
data24 = mean(datanumerik2[24,])
data25 = mean(datanumerik2[25,])
data26 = mean(datanumerik2[26,])
data27 = mean(datanumerik2[27,])
data28 = mean(datanumerik2[28,])
data29 = mean(datanumerik2[29,])
data30 = mean(datanumerik2[30,])
data31 = mean(datanumerik2[31,])
data32 = mean(datanumerik2[32,])
data33 = mean(datanumerik2[33,])
data34 = mean(datanumerik2[34,])
data35 = mean(datanumerik2[35,])
data36 = mean(datanumerik2[36,])
data37 = mean(datanumerik2[37,])
data38 = mean(datanumerik2[38,])
datarata=
rbind(data1,data2,data3,data4,data5,data6,data7,data8,data9,data10,data
11,data12,data13,data14,data15,data16,data17,data18,data19,data20,data
21,data22,data23,data24,data25,data26,data27,data28,data29,data30,data
31,data32,data33,data34,data35,data36,data37,data38)
#Analisis single linkage
kelompok.single=hasilkelompoknumerik$single
data.single=data.frame(single, datarata)
data.single.sort=data.single[order(data.single$single),]
data.single.sort
a=length(data.single.sort$single[data.single.sort$single==1])
b=length(data.single.sort$single[data.single.sort$single==2])
c=length(data.single.sort$single[data.single.sort$single==3])
d=length(data.single.sort$single[data.single.sort$single==4])
e=length(data.single.sort$single[data.single.sort$single==5])
data.c1.single=data.single.sort[1:a,]
data.c2.single=data.single.sort[(a+1):(a+b),]
data.c3.single=data.single.sort[(a+b+1):(a+b+c),]
data.c4.single=data.single.sort[(a+b+c+1):(a+b+c+d),]
data.c5.single=data.single.sort[(a+b+c+d+1):38,]
mean.c1.single=mean(data.c1.single$datarata)

```

```

mean.c2.single=mean(data.c2.single$datarata)
mean.c3.single=mean(data.c3.single$datarata)
mean.c4.single=mean(data.c4.single$datarata)
mean.c5.single=mean(data.c5.single$datarata)
s1.single=sqrt((sum((data.c1.single$datarata-mean.c1.single)^2))/(a))
s2.single=sqrt((sum((data.c2.single$datarata-mean.c2.single)^2))/(b))
s3.single=sqrt((sum((data.c3.single$datarata-mean.c3.single)^2))/(c))
s4.single=sqrt((sum((data.c4.single$datarata-mean.c4.single)^2))/(d))
s5.single=sqrt((sum((data.c5.single$datarata-mean.c5.single)^2))/(e))
jumlah.s.single=sum(s1.single, s2.single, s3.single, s4.single, s5.single)
Sw.single=jumlah.s.single/5
y1.single=(mean.c1.single-mean(datarata))^2
y2.single=(mean.c2.single-mean(datarata))^2
y3.single=(mean.c3.single-mean(datarata))^2
y4.single=(mean.c4.single-mean(datarata))^2
y5.single=(mean.c5.single-mean(datarata))^2
jumlah.single=sum(y1.single, y2.single, y3.single, y4.single, y5.single)
sb.single=sqrt(jumlah.single/(5-1))
rasio.single=Sw.single/sb.single
hasil.single=c(Sw.single, sb.single,rasio.single)
#Analisis complete linkage
kelompok.complete=hasilkelompoknumerik$complete
data.complete=data.frame(complete, datarata)
data.complete.sort=data.complete[order(data.complete$complete),]
data.complete.sort
a=length(data.complete.sort$complete[data.complete.sort$complete==1])
b=length(data.complete.sort$complete[data.complete.sort$complete==2])
c=length(data.complete.sort$complete[data.complete.sort$complete==3])
d=length(data.complete.sort$complete[data.complete.sort$complete==4])
e=length(data.complete.sort$complete[data.complete.sort$complete==5])
data.c1.complete=data.complete.sort[1:a,]
data.c2.complete=data.complete.sort[(a+1):(a+b),]
data.c3.complete=data.complete.sort[(a+b+1):(a+b+c),]
data.c4.complete=data.complete.sort[(a+b+c+1):(a+b+c+d),]
data.c5.complete=data.complete.sort[(a+b+c+d+1):38,]
mean.c1.complete=mean(data.c1.complete$datarata)
mean.c2.complete=mean(data.c2.complete$datarata)
mean.c3.complete=mean(data.c3.complete$datarata)
mean.c4.complete=mean(data.c4.complete$datarata)

```

```

mean.c5.complete=mean(data.c5.complete$datarata)
s1.complete=sqrt((sum((data.c1.complete$datarata-
mean.c1.complete)^2))/(a))
s2.complete=sqrt((sum((data.c2.complete$datarata-
mean.c2.complete)^2))/(b))
s3.complete=sqrt((sum((data.c3.complete$datarata-
mean.c3.complete)^2))/(c))
s4.complete=sqrt((sum((data.c4.complete$datarata-
mean.c4.complete)^2))/(d))
s5.complete=sqrt((sum((data.c5.complete$datarata-
mean.c5.complete)^2))/(e))
jumlah.s.complete=sum(s1.complete,s2.complete,s3.complete,
s4.complete, s5.complete)
Sw.complete=jumlah.s.complete/5
y1.complete=(mean.c1.complete-mean(datarata))^2
y2.complete=(mean.c2.complete-mean(datarata))^2
y3.complete=(mean.c3.complete-mean(datarata))^2
y4.complete=(mean.c4.complete-mean(datarata))^2
y5.complete=(mean.c5.complete-mean(datarata))^2
jumlah.complete=sum(y1.complete, y2.complete, y3.complete,
y4.complete, y5.complete)
sb.complete=sqrt(jumlah.complete/(5-1))
rasio.complete=Sw.complete/sb.complete
hasil.complete=c(Sw.complete, sb.complete,rasio.complete)
#Analisis average linkage
kelompok.average=hasilkelompoknumerik$average
data.average=data.frame(average, datarata)
data.average.sort=data.average[order(data.average$average),]
data.average.sort
a=length(data.average.sort$average[data.average.sort$average==1])
b=length(data.average.sort$average[data.average.sort$average==2])
c=length(data.average.sort$average[data.average.sort$average==3])
d=length(data.average.sort$average[data.average.sort$average==4])
data.c1.average=data.average.sort[1:a,]
data.c2.average=data.average.sort[(a+1):(a+b),]
data.c3.average=data.average.sort[(a+b+1):(a+b+c),]
data.c4.average=data.average.sort[(a+b+c+1):38,]
mean.c1.average=mean(data.c1.average$datarata)
mean.c2.average=mean(data.c2.average$datarata)

```

```

mean.c3.average=mean(data.c3.average$datarata)
mean.c4.average=mean(data.c4.average$datarata)
s1.average=sqrt((sum((data.c1.average$datarata-
mean.c1.average)^2))/(a))
s2.average=sqrt((sum((data.c2.average$datarata-
mean.c2.average)^2))/(b))
s3.average=sqrt((sum((data.c3.average$datarata-
mean.c3.average)^2))/(c))
s4.average=sqrt((sum((data.c4.average$datarata-
mean.c4.average)^2))/(d))
jumlah.s.average=sum(s1.average, s2.average, s3.average, s4.average)
Sw.average=jumlah.s.average/4
y1.average=(mean.c1.average-mean(datarata))^2
y2.average=(mean.c2.average-mean(datarata))^2
y3.average=(mean.c3.average-mean(datarata))^2
y4.average=(mean.c4.average-mean(datarata))^2
jumlah.average=sum(y1.average, y2.average, y3.average, y4.average)
sb.average=sqrt(jumlah.average/(4-1))
rasio.average=Sw.average/sb.average
hasil.average=c(Sw.average, sb.average,rasio.average)
#Analisis ward linkage
kelompok.ward=hasilkelompoknumerik$ward
data.ward=data.frame(ward, datarata)
data.ward.sort=data.ward[order(data.ward$ward),]
data.ward.sort
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):38,]
mean.c1.ward=mean(data.c1.ward$datarata)
mean.c2.ward=mean(data.c2.ward$datarata)
mean.c3.ward=mean(data.c3.ward$datarata)
mean.c4.ward=mean(data.c4.ward$datarata)

```

```

mean.c5.ward=mean(data.c5.ward$datarata)
s1.ward=sqrt((sum((data.c1.ward$datarata-mean.c1.ward)^2))/(a))
s2.ward=sqrt((sum((data.c2.ward$datarata-mean.c2.ward)^2))/(b))
s3.ward=sqrt((sum((data.c3.ward$datarata-mean.c3.ward)^2))/(c))
s4.ward=sqrt((sum((data.c4.ward$datarata-mean.c4.ward)^2))/(d))
s5.ward=sqrt((sum((data.c5.ward$datarata-mean.c5.ward)^2))/(e))
jumlah.s.ward=sum(s1.ward, s2.ward, s3.ward, s4.ward, s5.ward)
Sw.ward=jumlah.s.ward/5
y1.ward=(mean.c1.ward-mean(datarata))^2
y2.ward=(mean.c2.ward-mean(datarata))^2
y3.ward=(mean.c3.ward-mean(datarata))^2
y4.ward=(mean.c4.ward-mean(datarata))^2
y5.ward=(mean.c5.ward-mean(datarata))^2
jumlah.ward=sum(y1.ward, y2.ward, y3.ward, y4.ward, y5.ward)
sb.ward=sqrt(jumlah.ward/(5-1))
rasio.ward=Sw.ward/sb.ward
hasil.ward=c(Sw.ward, sb.ward,rasio.ward)

hasil=rbind(hasil.single, hasil.complete, hasil.average, hasil.ward)
colnames(hasil)=c("Sw", "Sb", "Ratio")
print ("Nilai Rasio antara Sw dan Sb")
print(hasilkelompoknumerik)
print("-----")
print("Metode Sw Sb Ratio")
print(hasil)

```

#### Lampiran 4 Syntax Analisis Metode ROCK untuk Pengelompokan Data Kategorik

```
datakategorik1 = c(data$X4, data$X5, data$X6, data$X7)
datakategorik2 = matrix (datakategorik1,38,4)
x=as.dummy(datakategorik2)
rc.05 = rockCluster(x, n=5, theta=0.05, debug=FALSE)
rc.10 = rockCluster(x, n=5, theta=0.10, debug=FALSE)
rc.15 = rockCluster(x, n=5, theta=0.15, debug=FALSE)
rc.20 = rockCluster(x, n=5, theta=0.20, debug=FALSE)
rc.25 = rockCluster(x, n=5, theta=0.25, debug=FALSE)
rc.30 = rockCluster(x, n=5, theta=0.30, debug=FALSE)
rc.35 = rockCluster(x, n=5, theta=0.35, debug=FALSE)
rc.40 = rockCluster(x, n=5, theta=0.40, debug=FALSE)
rc.45 = rockCluster(x, n=5, theta=0.45, debug=FALSE)
rc.50 = rockCluster(x, n=5, theta=0.50, debug=FALSE)
rf.05 = fitted(rc.05)
rf.10 = fitted(rc.10)
rf.15 = fitted(rc.15)
rf.20 = fitted(rc.20)
rf.25 = fitted(rc.25)
rf.30 = fitted(rc.30)
rf.35 = fitted(rc.35)
rf.40 = fitted(rc.40)
rf.45 = fitted(rc.45)
rf.50 = fitted(rc.50)
theta.05 = rf.05$cl
theta.10 = rf.10$cl
theta.15 = rf.15$cl
theta.20 = rf.20$cl
theta.25 = rf.25$cl
theta.30 = rf.30$cl
theta.35 = rf.35$cl
theta.40 = rf.40$cl
theta.45 = rf.45$cl
theta.50 = rf.50$cl
Kelompok = data.frame(theta.05, theta.10, theta.15, theta.20,
                        theta.25, theta.30, theta.35, theta.40, theta.45, theta.50)
Hasil = data.frame(data$Kabupaten, Kelompok)
```

```

jumlah.05 = count(Hasil,"theta.05")
jumlah.10 = count(Hasil,"theta.10")
jumlah.15 = count(Hasil,"theta.15")
jumlah.20 = count(Hasil,"theta.20")
jumlah.25 = count(Hasil,"theta.25")
jumlah.30 = count(Hasil,"theta.30")
jumlah.35 = count(Hasil,"theta.35")
jumlah.40 = count(Hasil,"theta.40")
jumlah.45 = count(Hasil,"theta.45")
jumlah.50 = count(Hasil,"theta.50")
X4 = factor (data$X4)
X5 = factor (data$X5)
X6 = factor (data$X6)
X7 = factor (data$X7)
p.05 = as.numeric(as.character(theta.05))~X4+X5+X6+X7
p.10 = as.numeric(as.character(theta.10))~X4+X5+X6+X7
p.15 = as.numeric(as.character(theta.15))~X4+X5+X6+X7
p.20 = as.numeric(as.character(theta.20))~X4+X5+X6+X7
p.25 = as.numeric(as.character(theta.25))~X4+X5+X6+X7
p.30 = as.numeric(as.character(theta.30))~X4+X5+X6+X7
p.35 = as.numeric(as.character(theta.35))~X4+X5+X6+X7
p.40 = as.numeric(as.character(theta.40))~X4+X5+X6+X7
p.45 = as.numeric(as.character(theta.45))~X4+X5+X6+X7
p.50 = as.numeric(as.character(theta.50))~X4+X5+X6+X7
mylogit.05 = aov(p.05)
mylogit.10 = aov(p.10)
mylogit.15 = aov(p.15)
mylogit.20 = aov(p.20)
mylogit.25 = aov(p.25)
mylogit.30 = aov(p.30)
mylogit.35 = aov(p.35)
mylogit.40 = aov(p.40)
mylogit.45 = aov(p.45)
mylogit.50 = aov(p.50)
SSW.05 = sum((mylogit.05$residuals)^2)
SSW.10 = sum((mylogit.10$residuals)^2)
SSW.15 = sum((mylogit.15$residuals)^2)
SSW.20 = sum((mylogit.20$residuals)^2)
SSW.25 = sum((mylogit.25$residuals)^2)

```

```
SSW.30 = sum((mylogit.30$residuals)^2)
SSW.35 = sum((mylogit.35$residuals)^2)
SSW.40 = sum((mylogit.40$residuals)^2)
SSW.45 = sum((mylogit.45$residuals)^2)
SSW.50 = sum((mylogit.50$residuals)^2)
s.05 = summary(mylogit.05)
s.10 = summary(mylogit.10)
s.15 = summary(mylogit.15)
s.20 = summary(mylogit.20)
s.25 = summary(mylogit.25)
s.30 = summary(mylogit.30)
s.35 = summary(mylogit.35)
s.40 = summary(mylogit.40)
s.45 = summary(mylogit.45)
s.50 = summary(mylogit.50)
summary.05 = matrix(unlist(s.05),5,5)
summary.10 = matrix(unlist(s.10),5,5)
summary.15 = matrix(unlist(s.15),5,5)
summary.20 = matrix(unlist(s.20),5,5)
summary.25 = matrix(unlist(s.25),5,5)
summary.30 = matrix(unlist(s.30),5,5)
summary.35 = matrix(unlist(s.35),5,5)
summary.40 = matrix(unlist(s.40),5,5)
summary.45 = matrix(unlist(s.45),5,5)
summary.50 = matrix(unlist(s.50),5,5)
SSB.05 = sum(summary.05[1:4,2])
SSB.10 = sum(summary.10[1:4,2])
SSB.15 = sum(summary.15[1:4,2])
SSB.20 = sum(summary.20[1:4,2])
SSB.25 = sum(summary.25[1:4,2])
SSB.30 = sum(summary.30[1:4,2])
SSB.35 = sum(summary.35[1:4,2])
SSB.40 = sum(summary.40[1:4,2])
SSB.45 = sum(summary.45[1:4,2])
SSB.50 = sum(summary.50[1:4,2])
c.05 = sum(rf.05$size>0)
c.10 = sum(rf.10$size>0)
c.15 = sum(rf.15$size>0)
c.20 = sum(rf.20$size>0)
```

```

c.25 = sum(rf.25$size>0)
c.30 = sum(rf.30$size>0)
c.35 = sum(rf.35$size>0)
c.40 = sum(rf.40$size>0)
c.45 = sum(rf.45$size>0)
c.50 = sum(rf.50$size>0)
SW.05 = sqrt(SSW.05/(38-c.05))
SW.10 = sqrt(SSW.10/(38-c.10))
SW.15 = sqrt(SSW.15/(38-c.15))
SW.20 = sqrt(SSW.20/(38-c.20))
SW.25 = sqrt(SSW.25/(38-c.25))
SW.30 = sqrt(SSW.30/(38-c.30))
SW.35 = sqrt(SSW.35/(38-c.35))
SW.40 = sqrt(SSW.40/(38-c.40))
SW.45 = sqrt(SSW.45/(38-c.45))
SW.50 = sqrt(SSW.50/(38-c.50))
SB.05 = sqrt(SSB.05/(c.05-1))
SB.10 = sqrt(SSB.10/(c.10-1))
SB.15 = sqrt(SSB.15/(c.15-1))
SB.20 = sqrt(SSB.20/(c.20-1))
SB.25 = sqrt(SSB.25/(c.25-1))
SB.30 = sqrt(SSB.30/(c.30-1))
SB.35 = sqrt(SSB.35/(c.35-1))
SB.40 = sqrt(SSB.40/(c.40-1))
SB.45 = sqrt(SSB.45/(c.45-1))
SB.50 = sqrt(SSB.50/(c.50-1))
Ratio.05 = (SW.05/SB.05)
Ratio.10 = (SW.10/SB.10)
Ratio.15 = (SW.15/SB.15)
Ratio.20 = (SW.20/SB.20)
Ratio.25 = (SW.25/SB.25)
Ratio.30 = (SW.30/SB.30)
Ratio.35 = (SW.35/SB.35)
Ratio.40 = (SW.40/SB.40)
Ratio.45 = (SW.45/SB.45)
Ratio.50 = (SW.50/SB.50)
Ratio.kategorik = rbind(Ratio.05, Ratio.10, Ratio.15, Ratio.20,
Ratio.25, Ratio.30, Ratio.35, Ratio.40, Ratio.45,
Ratio.50)

```

```
print("Hasil Pengelompokan ROCK")
print(Hasil)
print("-----")
print(jumlah.05)
print(jumlah.10)
print(jumlah.15)
print(jumlah.20)
print(jumlah.25)
print(jumlah.30)
print(jumlah.35)
print(jumlah.40)
print(jumlah.45)
print(jumlah.50)
print("-----")
print("Nilai Ratio yang Terbentuk")
print(Ratio.kategorik)
u=c(0.05,0.10,0.15,0.20,0.25,0.30,0.35,0.40,0.45,0.50)
z = c(Ratio.05, Ratio.10, Ratio.15, Ratio.20, Ratio.25, Ratio.30,
      Ratio.35, Ratio.40, Ratio.45, Ratio.50)
plot(u, z, main = " Plot Nilai Ratio Metode ROCK",
     xlab = "Nilai Theta", ylab = "Nilai Ratio")
lines(u,z,col = "Red")
```

**Lampiran 5 Syntax Analisis Pengelompokan Ensemble ROCK**

```

datakategorik1=c(data$numerik,data$kategorik)
datakategorik2=matrix(datakategorik1,38,2)
x = as.dummy(datakategorik2)
rc.05= rockCluster(x, n=5, theta=0.05, debug=FALSE)
rc.10= rockCluster(x, n=5, theta=0.10, debug=FALSE)
rc.15= rockCluster(x, n=5, theta=0.15, debug=FALSE)
rc.20= rockCluster(x, n=5, theta=0.20, debug=FALSE)
rc.25= rockCluster(x, n=5, theta=0.25, debug=FALSE)
rc.30= rockCluster(x, n=5, theta=0.30, debug=FALSE)
rf.05 = fitted(rc.05)
rf.10 = fitted(rc.10)
rf.15 = fitted(rc.15)
rf.20 = fitted(rc.20)
rf.25 = fitted(rc.25)
rf.30 = fitted(rc.30)
theta.05 = rf.05$cl
theta.10 = rf.10$cl
theta.15 = rf.15$cl
theta.20 = rf.20$cl
theta.25 = rf.25$cl
theta.30 = rf.30$cl
Kelompok = data.frame( theta.05, theta.10, theta.15, theta.20,
                        theta.25, theta.30)
Hasil = data.frame(data$Kabupaten, Kelompok)
jumlah.05 = count(Hasil,"theta.05")
jumlah.10 = count(Hasil,"theta.10")
jumlah.15 = count(Hasil,"theta.15")
jumlah.20 = count(Hasil,"theta.20")
jumlah.25 = count(Hasil,"theta.25")
jumlah.30 = count(Hasil,"theta.30")
numerik = factor(data$numerik)
kategorik = factor(data$kategorik)
numtheta.05=as.numeric(as.character(theta.05))
numtheta.10=as.numeric(as.character(theta.10))
numtheta.15=as.numeric(as.character(theta.15))
numtheta.20=as.numeric(as.character(theta.20))
numtheta.25=as.numeric(as.character(theta.25))
numtheta.30=as.numeric(as.character(theta.30))

```

```

p.05= numtheta.05~numerik+kategorik
p.10= numtheta.10~numerik+kategorik
p.15= numtheta.15~numerik+kategorik
p.20= numtheta.20~numerik+kategorik
p.25= numtheta.25~numerik+kategorik
p.30= numtheta.30~numerik+kategorik
mylogit.05 = aov( p.05)
mylogit.10 = aov( p.10)
mylogit.15 = aov( p.15)
mylogit.20 = aov( p.20)
mylogit.25 = aov( p.25)
mylogit.30 = aov( p.30)
SSW.05 = sum((mylogit.05$residuals)^2)
SSW.10 = sum((mylogit.10$residuals)^2)
SSW.15 = sum((mylogit.15$residuals)^2)
SSW.20 = sum((mylogit.20$residuals)^2)
SSW.25 = sum((mylogit.25$residuals)^2)
SSW.30 = sum((mylogit.30$residuals)^2)
h.05=unlist(summary(mylogit.05))
h.10=unlist(summary(mylogit.10))
h.15=unlist(summary(mylogit.15))
h.20=unlist(summary(mylogit.20))
h.25=unlist(summary(mylogit.25))
h.30=unlist(summary(mylogit.30))
summary.05 = matrix(h.05, 3,5 )
summary.10 = matrix(h.10, 3,5 )
summary.15 = matrix(h.15, 3,5 )
summary.20 = matrix(h.20, 3,5 )
summary.25 = matrix(h.25, 3,5 )
summary.30 = matrix(h.30, 3,5 )
a=sum(rf.05$size>0, na.rm=TRUE)
b=sum(rf.10$size>0, na.rm=TRUE)
c=sum(rf.15$size>0, na.rm=TRUE)
d=sum(rf.20$size>0, na.rm=TRUE)
e=sum(rf.25$size>0, na.rm=TRUE)
f=sum(rf.30$size>0, na.rm=TRUE)
SSB.05 = sum(summary.05[1:2,2])
SSB.10 = sum(summary.10[1:2,2])
SSB.15 = sum(summary.15[1:2,2])

```

```

SSB.20 = sum(summary.20[1:2,2])
SSB.25 = sum(summary.25[1:2,2])
SSB.30 = sum(summary.30[1:2,2])
SW.05 = sqrt(SSW.05/(38-a))
SW.10 = sqrt(SSW.10/(38-b))
SW.15 = sqrt(SSW.15/(38-c))
SW.20 = sqrt(SSW.20/(38-d))
SW.25 = sqrt(SSW.25/(38-e))
SW.30 = sqrt(SSW.30/(38-f))
SB.05 = sqrt(SSB.05/(a-1))
SB.10 = sqrt(SSB.10/(b-1))
SB.15 = sqrt(SSB.15/(c-1))
SB.20 = sqrt(SSB.20/(d-1))
SB.25 = sqrt(SSB.25/(e-1))
SB.30 = sqrt(SSB.30/(f-1))
Ratio.05 = (SW.05/SB.05)
Ratio.10 = (SW.10/SB.10)
Ratio.15 = (SW.15/SB.15)
Ratio.20 = (SW.20/SB.20)
Ratio.25 = (SW.25/SB.25)
Ratio.30 = (SW.30/SB.30)
Ratio.EnsembleRock = rbind(Ratio.05, Ratio.10, Ratio.15, Ratio.20,
Ratio.25, Ratio.30)
print("Hasil Pengelompokan Ensemble ROCK")
print(Hasil)
print(jumlah.05)
print(jumlah.10)
print(jumlah.15)
print(jumlah.20)
print(jumlah.25)
print(jumlah.30)
print("-----")
print("Nilai Ratio yang Terbentuk")
print(Ratio.EnsembleRock)
u=c(0.05,0.10,0.15,0.20,0.25,0.30)
z = c(Ratio.05, Ratio.10, Ratio.15, Ratio.20, Ratio.25, Ratio.30)
plot(u, z, main = " Plot Nilai Ratio Metode Ensembl ROCK",
      xlab = "Nilai Theta", ylab = "Nilai Ratio")
lines(u,z,col = "Red")

```

**Lampiran 6** *Output* untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 2

```
> cluster.2(data)
[1] "Hasil Pengelompokan Hirarki"
      single complete average ward
1         1           1         1     1
2         1           1         1     1
3         1           1         1     1
4         1           1         1     1
5         1           1         1     1
6         1           1         1     1
7         1           1         1     1
8         1           1         1     1
9         1           1         1     1
10        1           1         1     1
11        1           1         1     1
12        1           1         1     1
13        1           1         1     1
14        1           1         1     1
15        1           2         1     2
16        1           1         1     1
17        1           1         1     1
18        1           1         1     1
19        1           1         1     1
20        1           1         1     1
21        1           1         1     1
22        1           1         1     1
23        1           1         1     1
24        1           1         1     1
25        1           1         1     2
26        1           1         1     1
27        1           1         1     1
28        1           1         1     1
29        1           1         1     1
30        1           1         1     2
31        1           1         1     2
32        1           1         1     2
33        1           1         1     2
34        1           1         1     2
```

```

35      1      1      1      2
36      1      1      1      2
37      2      2      2      2
38      1      1      1      2
[1] "-----"
-----"
[1] " Untuk 2 kelompok yang terbentuk "
[1] "Metode SST SSW R-square"
                SST          SSW  R-Squa
re  PseudoF
Hasil.single  4.557692e+13 3.179548e+13 0.30237
75 15.60384
Hasil.complete 4.557692e+13 2.749455e+13 0.39674
40 23.67616
Hasil.average  4.557692e+13 3.179548e+13 0.30237
75 15.60384
Hasil.ward     4.557692e+13 1.069556e+13 0.76532
94 117.40649

```

**Lampiran 7** *Output* untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 3

```

> cluster.3(data)
[1] "Hasil Pengelompokan Hirarki"
      single complete average ward
1         1           1         1     1
2         1           1         1     2
3         1           1         1     1
4         1           1         1     1
5         1           1         1     1
6         1           1         1     1
7         1           1         1     1
8         1           1         1     1
9         1           1         1     2
10        1           1         1     1
11        1           1         1     2
12        1           1         1     2
13        1           1         1     2
14        1           1         1     1
15        2           2         2     3
16        1           1         2     1
17        1           1         1     1
18        1           1         1     1
19        1           1         1     1
20        1           1         1     1
21        1           1         1     2
22        1           1         1     1
23        1           1         1     1
24        1           1         2     1
25        1           3         2     3
26        1           1         1     1
27        1           1         1     1
28        1           1         1     2
29        1           1         1     1
30        1           3         2     3
31        1           3         2     3
32        1           3         2     3
33        1           3         2     3
34        1           3         2     3

```

```

35      1      3      2      3
36      1      3      2      3
37      3      2      3      3
38      1      3      2      3
[1] "-----"
-----"
[1] " Untuk 3 kelompok yang terbentuk "
[1] "Metode SST SSW R-square"
                SST          SSW  R-Squa
re  PseudoF
Hasil.single  4.557692e+13 2.640967e+13 0.42054
73 12.70091
Hasil.complete 4.557692e+13 5.727456e+12 0.87433
43 121.75834
Hasil.average  4.557692e+13 5.697986e+12 0.87498
09 122.47860
Hasil.ward     4.557692e+13 8.961938e+12 0.80336
67 71.49817

```

**Lampiran 8** *Output* untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 4

```

> cluster.4(data)
[1] "Hasil Pengelompokan Hirarki"
      single complete average ward
1         1           1         1     1
2         1           1         1     2
3         1           1         1     1
4         1           2         1     1
5         1           1         1     1
6         1           2         1     1
7         1           1         1     1
8         1           1         1     1
9         1           1         1     2
10        1           1         1     1
11        1           1         1     2
12        1           1         1     2
13        1           1         1     2
14        1           2         1     1
15        2           3         2     3
16        3           2         3     1
17        1           1         1     1
18        1           1         1     1
19        1           1         1     1
20        1           2         1     1
21        1           1         1     2
22        1           1         1     1
23        1           1         1     1
24        3           2         3     1
25        3           4         2     3
26        1           2         1     1
27        1           1         1     1
28        1           1         1     2
29        1           1         1     1
30        3           4         3     3
31        3           4         3     3
32        3           4         2     3
33        3           4         3     3
34        3           4         3     3

```

```

35      3      4      3      3
36      3      4      2      3
37      4      3      4      4
38      3      4      3      3
-----
[1] "-----"
-----"
[1] " Untuk 4 kelompok yang terbentuk "
[1] "Metode SST SSW R-square"
                SST      SSW      R-Square
PseudoF
Hasil.single  4.557692e+13 4.432727e+12 0.90274
18 105.1950
Hasil.complete 4.557692e+13 2.506175e+12 0.94501
22 194.7730
Hasil.average  4.557692e+13 3.083829e+12 0.93233
79 156.1657
Hasil.ward    4.557692e+13 3.819652e+12 0.91619
33 123.8984

```

**Lampiran 9** *Output* untuk Pengelompokan Data Numerik dengan Jumlah Kelompok 5

```

> cluster.5(data)
[1] "Hasil Pengelompokan Hirarki"
      single complete average ward
1         1           1         1     1
2         1           1         1     2
3         1           1         1     1
4         1           2         1     1
5         1           1         1     1
6         1           2         1     1
7         1           1         1     1
8         1           1         1     1
9         1           1         1     2
10        1           1         1     1
11        1           1         1     2
12        1           1         1     2
13        1           1         1     2
14        1           2         1     1
15        2           3         2     3
16        3           2         3     1
17        1           1         1     1
18        1           1         1     1
19        1           1         1     1
20        1           2         1     1
21        1           1         1     2
22        1           1         1     1
23        1           1         1     1
24        3           2         3     1
25        4           4         2     3
26        1           2         1     1
27        1           1         1     1
28        1           1         1     2
29        1           1         1     1
30        4           4         4     4
31        4           4         4     4
32        4           4         2     3
33        4           4         4     4
34        4           4         4     4

```

```

35      4      4      4      4
36      4      4      2      3
37      5      5      5      5
38      4      4      4      4
[1] "-----"
-----"
[1] " Untuk 5 kelompok yang terbentuk "
[1] "Metode SST SSW R-square"
                SST          SSW    R-Square
PseudoF
Hasil.single   4.557692e+13 1.854646e+12 0.95930
73 194.4892
Hasil.complete 4.557692e+13 1.421296e+12 0.96881
54 256.3040
Hasil.average  4.557692e+13 2.340296e+12 0.94865
17 152.4175
Hasil.ward     4.557692e+13 2.229061e+12 0.95109
23 160.4352

```

**Lampiran 10** *Output* untuk Pengelompokan Data Kategorik Menggunakan Metode ROCK

Hasil Pengelompokan ROCK					
	data. Kabupa ten	theta. 05	theta. 10	theta. 15	theta. 20
1	PACITAN	4	3	1	1
2	PONOROGO	4	3	1	1
3	TRENGGALEK	4	3	1	1
4	TULUNGAGUNG	4	3	1	1
5	BLITAR	4	3	1	1
6	KEDIRI	4	3	1	1
7	MALANG	4	3	1	1
8	LUMAJANG	4	3	1	1
9	JEMBER	4	3	1	1
10	BANYUWANGI	4	3	1	1
11	BONDOWOSO	4	3	1	1
12	SITUBONDO	4	3	1	1
13	PROBOLINGGO	4	3	1	1
14	PASURUAN	4	3	1	1
15	SIDOARJO	5	3	1	1
16	MOJOKERTO	4	3	1	1
17	JOMBANG	4	3	1	1
18	NGANJUK	4	3	1	1
19	MADIUN	4	3	1	1
20	MAGETAN	4	3	1	1
21	NGAWI	4	3	1	1
22	BOJONEGORO	4	3	1	1
23	TUBAN	4	3	1	1
24	LAMONGAN	5	3	1	1
25	GRESIK	5	3	1	1



2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	2	1	1	1
2	2	2	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1
1	1	2	1	1	1
2	2	2	1	1	1
2	2	2	1	1	1
2	2	2	1	1	1
1	1	<NA>	<NA>	<NA>	<NA>
1	1	<NA>	<NA>	<NA>	<NA>
2	2	1	1	1	1
theta.05 freq					
1	4	28			
2	5	10			
theta.10 freq					
1	3	38			
theta.15 freq					
1	1	35			
2	2	3			

```
theta.20 freq
1      1   35
2      2    3
theta.25 freq
1      1    3
2      2   35
theta.30 freq
1      1    3
2      2   35
theta.35 freq
1      1   30
2      2    6
3     <NA>   2
theta.40 freq
1      1   36
2     <NA>   2
theta.45 freq
1      1   36
2     <NA>   2
theta.50 freq
1      1   36
2     <NA>   2
```

```
[1] "Nilai Ratio yang Terbentuk"
```

```
      [,1]
Ratio.05 4.845121e-02
Ratio.10 0.000000e+00
Ratio.15 2.213046e-16
Ratio.20 2.213046e-16
Ratio.25 6.838183e-17
Ratio.30 6.838183e-17
Ratio.35 7.258743e-02
Ratio.40 1.903324e+00
Ratio.45 1.903324e+00
Ratio.50 1.903324e+00
```

**Lampiran 11** *Output* untuk Pengelompokan Data Campuran  
Menggunakan Metode *Ensemble ROCK*

[1] "Hasil Pengelompokan Ensemble ROCK"					
	data.kabupaten	theta.05	theta.10	theta.15	theta.20
1	PACITAN	2	2	1	1
2	PONOROGO	2	2	1	1
3	TRENGGALEK	2	2	1	1
4	TULUNGAGUNG	2	2	1	1
5	BLITAR	2	2	1	1
6	KEDIRI	2	2	1	1
7	MALANG	2	2	1	1
8	LUMAJANG	2	2	1	1
9	JEMBER	2	2	1	1
10	BANYUWANGI	2	2	1	1
11	BONDOWOSO	2	2	1	1
12	SITUBONDO	2	2	1	1
13	PROBOLINGGO	2	2	1	1
14	PASURUAN	2	2	1	1
15	SIDOARJO	2	2	1	1
16	MOJOKERTO	2	2	1	1
17	JOMBANG	2	2	1	1
18	NGANJUK	2	2	1	1
19	MADIUN	2	2	1	1
20	MAGETAN	2	2	1	1
21	NGAWI	2	2	1	1
22	BOJONEGORO	2	2	1	1
23	TUBAN	2	2	1	1
24	LAMONGAN	2	2	1	1
25	GRESIK	4	4	1	1
26	BANGKALAN	2	2	1	1
27	SAMPANG	2	2	1	1
28	PAMEKASAN	2	2	1	1
29	SUMENEP	2	2	1	1
30	KEDIRI	4	4	1	1
31	BLITAR	4	4	1	1
32	MALANG	4	4	2	2
33	PROBOLINGGO	4	4	1	1
34	PASURUAN	4	4	1	1
35	MOJOKERTO	4	4	1	1

36	MADIUN	4	4	2	2
37	SURABAYA	4	4	<NA>	<NA>
38	BATU	4	4	1	1
	data.Kabupaten	theta.25	theta.30		
1	PACITAN	1	1		
2	PONOROGO	1	1		
3	TRENGGALEK	1	1		
4	TULUNGAGUNG	1	1		
5	BLITAR	1	1		
6	KEDIRI	1	1		
7	MALANG	1	1		
8	LUMAJANG	1	1		
9	JEMBER	1	1		
10	BANYUWANGI	1	1		
11	BONDOWOSO	1	1		
12	SITUBONDO	1	1		
13	PROBOLINGGO	1	1		
14	PASURUAN	1	1		
15	SIDOARJO	1	1		
16	MOJOKERTO	1	1		
17	JOMBANG	1	1		
18	NGANJUK	1	1		
19	MADIUN	1	1		
20	MAGETAN	1	1		
21	NGAWI	1	1		
22	BOJONEGORO	1	1		
23	TUBAN	1	1		
24	LAMONGAN	1	1		
25	GRESIK	1	1		
26	BANGKALAN	1	1		
27	SAMPANG	1	1		
28	PAMEKASAN	1	1		
29	SUMENEP	1	1		
30	KEDIRI	1	1		
31	BLITAR	1	1		
32	MALANG	2	2		
33	PROBOLINGGO	1	1		
34	PASURUAN	1	1		
35	MOJOKERTO	1	1		
36	MADIUN	2	2		

```

37      SURABAYA    <NA>    <NA>
38      BATU        1        1
theta.05 freq
1        2        28
2        4        10
  theta.10 freq
1        2        28
2        4        10
  theta.15 freq
1        1        35
2        2         2
3      <NA>        1
  theta.20 freq
1        1        35
2        2         2
3      <NA>        1
  theta.25 freq
1        1        35
2        2         2
3      <NA>        1
  theta.30 freq
1        1        35
2        2         2
3      <NA>        1

```

```

[1] "-----"
----"

```

```

[1] "Nilai Ratio yang Terbentuk"

```

```

[,1]

```

```

Ratio.05 4.821843e-17

```

```

Ratio.10 4.821843e-17

```

```

Ratio.15 6.919863e-17

```

```

Ratio.20 6.919863e-17

```

```

Ratio.25 6.919863e-17

```

```

Ratio.30 6.919863e-17

```

## Lampiran 12 Surat Keterangan Pengambilan Data

 <p><b>BADAN PUSAT STATISTIK PROVINSI JAWA TIMUR</b></p>	 <p><b>SENSUS EKONOMI</b></p>
<b><u>SURAT KETERANGAN</u></b>	
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	: Vinda Ferdhiani
Fakultas/Program Studi	: Fakultas Matematika, Komputasi Dan Sains Data / Statistika
N.R.P	: 06211440000069
Alamat Rumah	: Jl. Perumahan City Home E1/18 Gg. Makam, Keputih, Sukolilo, Surabaya
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 :	
<b><i>" Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Faktor-Faktor Terkait Kebutuhan Konsumsi Air Bersih Rumah Tangga Menggunakan Metode Ensemble ROCK"</i></b>	
Demikian surat keterangan ini dibuat dan agar dipergunakan sebagaimana mestinya	
Surabaya, 24 Mei 2018	
An. Kepala BPS Provinsi Jawa Timur Kepala Bidang IPDS	
Thomas Wunang Tjahjo, M.Sc, M.Eng.	
	
Jalan Raya Kendangsari Industri No. 43 - 44, Surabaya - 60292 Telp. 031 - 8439343 Fax. 031 - 8494007. Homepage: <a href="http://jatm.bps.go.id">http://jatm.bps.go.id</a> E-mail: <a href="mailto:bps3500@bps.go.id">bps3500@bps.go.id</a>	

## BIODATA PENULIS



Penulis memiliki nama lengkap Vinda Ferdhiani atau biasa dipanggil Vinda. Penulis merupakan anak kedua dari pasangan Bapak Puji dan Ibu Fatma Nurnaningsih yang lahir di Kediri pada tanggal 22 Agustus 1995. Penulis telah menempuh pendidikan formal di SD Muhammadiyah 1 Ngadiluwih (2002-2008), SMP Negeri 1 Kediri (2008-2011), dan SMA Negeri 2 Kediri (2011-2014). Setelah menempuh pendidikan SMA, penulis melanjutkan pendidikan di jenjang perguruan tinggi di departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data, Institut Teknologi Sepuluh Nopember (ITS) melalui jalur SBMPTN pada tahun 2014. Selama berkuliah di departemen Statistika ITS, penulis bergabung UKM dan komunitas yaitu UKM *Technopreneurship Development Center* (TDC). Selain itu penulis juga berpartisipasi dalam kepanitiaan yang diadakan oleh departemen Statistika seperti kegiatan STATION (*Statistics Competition*) tingkat Nasional sebagai anggota Sie Acara. Segala kritik dan saran serta diskusi mengenai Tugas Akhir ini dapat menghubungi email penulis yaitu [vindaferr@gmail.com](mailto:vindaferr@gmail.com).