



TUGAS AKHIR - SS141501

**PEMETAAN KABUPATEN/KOTA DI JAWA TIMUR
BERDASARKAN INDIKATOR DAERAH TERTINGGAL
DENGAN METODE DATA CAMPURAN *ENSEMBLE*
ROCK DAN SWFM**

DWI HARID SETIADI
NRP 062114 4000 0066

**Dosen Pembimbing
Dr. Santi Puteri Rahayu, M.Si**

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



TUGAS AKHIR - SS141501

**PEMETAAN KABUPATEN/KOTA DI JAWA TIMUR
BERDASARKAN INDIKATOR DAERAH TERTINGGAL
DENGAN METODE DATA CAMPURAN *ENSEMBLE*
ROCK DAN SWFM**

**DWI HARID SETIADI
NRP. 062114 4000 0066**

**Dosen Pembimbing
Dr. Santi Puteri Rahayu, M.Si**

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



FINAL PROJECT - SS 141501

**MAPPING OF DISTRICTS IN EAST JAVA
BASED ON UNDERDEVELOPED INDICATORS
USING MIXED DATA METHODS ENSEMBLE
ROCK AND SWFM**

DWI HARID SETIADI
SN. 062114 4000 0066

Supervisor
Dr. Santi Puteri Rahayu, M.Si

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

LEMBAR PENGESAHAN

PEMETAAN KABUPATEN/KOTA DI JAWA TIMUR BERDASARKAN INDIKATOR DAERAH TERTINGGAL DENGAN METODE DATA CAMPURAN *ENSEMBLE ROCK DAN SWFM*

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 :
Dwi Harid Setiadi
NRP. 062114 4000 0066

Disetujui oleh Pembimbing:
Dr. Santi Puteri Rahayu, M.Si
NIP. 19750115 199903 2 003



Mengetahui,
Kepala Departemen



SURABAYA, JULI 2018

**PEMETAAN KABUPATEN/KOTA DI JAWA TIMUR
BERDASARKAN INDIKATOR DAERAH TERTINGGAL
DENGAN METODE DATA CAMPURAN ENSEMBLE
ROCK DAN SWFM**

Nama Mahasiswa : Dwi Harid Setiadi
NRP : 062114 4000 0066
Departemen : Statistika
Dosen Pembimbing : Dr. Santi Puteri Rahayu, M.Si

Abstrak

Provinsi Jawa Timur merupakan provinsi dengan jumlah Kabupaten/Kota yang termasuk daerah tertinggal terbanyak diantara provinsi di Pulau Jawa.. Penelitian yang dilakukan Kementerian DPDTT terkait penetapan daerah tertinggal hanya mengakomodasi data yang bersifat numerik. Oleh karena itu, dilakukan penelitian dengan metode yang berbeda yaitu dengan memetakan Kabupaten/Kota di Provinsi Jawa Timur yang mengakomodasi data numerik dan kategorik dengan mempertimbangkan aspek statistik. Pada penelitian ini dilakukan analisis pengelompokan menggunakan metode data campuran yaitu ensemble Robust Clustering using Links (ROCK) dan Similarity Weight and Filter Methods (SWFM). Hasil analisis menunjukkan metode ensemble ROCK lebih baik karena memiliki akurasi sebesar 0,724 yang lebih besar dari akurasi SWFM yaitu 0,618. Metode ensemble ROCK menghasilkan 4 klaster optimum. Dimana klaster 1 menunjukkan kondisi Kabupaten/Kota tertinggal karena memiliki 11 indikator yang rendah dibanding klaster lainnya. Hasil pengelompokan menunjukkan bahwa 32% Kabupaten/Kota termasuk klaster daerah tertinggal dari total 38 Kabupaten/Kota di Jawa Timur. Selain itu, 75% Kabupaten di Pulau Madura merupakan daerah tertinggal.

**Kata Kunci : Data Campuran, Indikator Daerah Tertinggal,
ROCK, SWFM**

(Halaman ini sengaja dikosongkan)

MAPPING OF DISTRICTS IN EAST JAVA BASED ON UNDERDEVELOPED INDICATORS USING MIXED DATA METHODS ENSEMBLE ROCK AND SWFM

Name : Dwi Harid Setiadi

Student Number : 062114 4000 0066

Department : Statistics

Supervisor : Dr. Santi Puteri Rahayu, M.Si

Abstract

East Java Province is a province that has the most underdeveloped districts among provinces in Java. Research conducted by the Ministry of DPDTT concerning the determination of underdeveloped areas only accommodates numerical data. Therefore, the research conducted with different methods that is by mapping the District in East Java Province that accommodates numerical and categorical data by considering the statistical aspects. In this research, grouping using mixed data method of Robust Clustering ensemble using Links (ROCK) and Similarity Weight and Filter Methods (SWFM). The results showed ensemble ROCK method is better because it has an accuracy of 0.724 greater than the accuracy of SWFM is 0.618. The Ensemble ROCK method produces 4 optimum clusters. Where cluster 1 shows cities left behind because it has 11 indicators that are lower than other clusters. The results of the clustering indicate that 32% of districts including clusters of underdeveloped district from a total 38 districts in East Java. In addition, 75% of districts in Madura Island are underdeveloped districts.

**Keywords : Ensemble ROCK, Underdeveloped Indicators,
Mixed Data, ROCK, SWFM**

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur yang kehadirat Allah SWT, Tuhan Yang Maha Esa. Berkat rahmat dan ridho-Nya penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul **“Pemetaan Kabupaten/Kota Di Jawa Timur Berdasarkan Indikator Daerah Tertinggal dengan Metode Data Campuran Ensemble ROCK dan SWFM”** dengan lancar.

Keberhasilan penyusunan Tugas Akhir ini tidak lepas dari banyaknya bantuan dan dukungan yang diberikan dari berbagai pihak. Penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Suhartono selaku Kepala Departemen Statistika dan Bapak Dr. Sutikno, M.Si selaku Ketua Program Studi Sarjana yang telah memberikan fasilitas untuk kelancaran penyelesaian Tugas Akhir.
2. Ibu Dr. Santi Puteri Rahayu, M.Si selaku dosen pembimbing yang dengan sabar telah memberikan bimbingan, ilmu, saran, dan dukungan selama penyusunan Tugas Akhir.
3. Bapak Imam Safawi Ahmad, M.Si dan Bapak M. Sjahid Akbar, M.Si selaku dosen penguji yang telah memberikan banyak bantuan dan saran untuk Tugas Akhir ini.
4. Bapak Dr. Ir. Setiawan, MS selaku dosen wali yang telah memberikan nasehat dan semangat selama masa perkuliahan.
5. Ibu dan Bapak tercinta, kakak tersayang serta keluarga yang selalu memberikan dukungan, kasih sayang, semangat, dan doa yang tidak pernah putus kepada penulis.
6. Kabinet dan Staf HIMASTA-ITS 2015/2016 dan 2016/2017 khususnya Departemen Kesma yang telah menemani dan membantu penulis untuk berkembang dalam berorganisasi.
7. Keluarga Statistika ITS angkatan 2014 (Respect) yang sudah menemani dan mensupport penulis selama masa kuliah. Khususnya kelompok 9 OKKBK dan teman-teman anggota pectpella.
8. Keluarga SEITO SMAN 3 Sidoarjo, khususnya yang tergabung dalam ‘Thats Right’ yang selalu support penulis.

Penulis berharap hasil Tugas Akhir ini dapat bermanfaat bagi kita semua. Semoga kebaikan dan bantuan yang telah diberikan kepada penulis dibalas dengan kebaikan yang lebih besar lagi oleh Tuhan Yang Maha Esa. Aamiin.

Surabaya, Juli 2018

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
COVER PAGE	iii
LEMBAR PENGESAHAN	v
ABSTRAK.....	vii
ABSTRACT	ix
KATA PENGANTAR.....	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Statistika Deskriptif.....	7
2.2 Analisis Faktor	8
2.3 Analisis Klaster (<i>Cluster Analysis</i>).....	14
2.3.1 Pengelompokan Data Numerik	15
2.3.2 Pengelompokan Data Kategorik	17
2.4 Pengelompokan Data Numerik dan Kategorik	21
2.5 <i>Similarity Weight and Filter Method</i> (SWFM)....	23
2.6 Kinerja Pengelompokan	26
2.6.1 Kinerja Pengelompokan Data Numerik	26
2.6.2 Kinerja Pengelompokan Data Kategorik....	29
2.7 Uji Signifikansi Antar Klaster	30
2.7.1 Uji MANOVA	30

2.7.3 Uji Kruskal-Wallis	34
2.8 Indikator Daerah Tertinggal.....	35
BAB III METODOLOGI PENELITIAN.....	39
3.1 Sumber Data	39
3.2 Variabel Penelitian.....	39
3.3 Langkah Analisis	41
BAB IV ANALISIS DAN PEMBAHASAN.....	47
4.1 Karakteristik Kabupaten/Kota di Jawa Timur	47
4.2 Analisis Faktor Indikator Daerah Tertinggal Kabupaten/Kota di Jawa Timur	51
4.2 Pengelompokan Data Numerik Indikator Daerah Tertinggal	56
4.3 Pengelompokan Data Kategorik Indikator Daerah Tertinggal	57
4.4 Pengelompokan Data Campuran Indikator Daerah Tertinggal di Jawa Timur Menggunakan Metode Ensemble ROCK dan SWFM	58
4.5 Karakteristik Klaster Hasil Pengelompokan Data Campuran Menggunakan Metode Terbaik ..	60
4.5.1 One-way MANOVA dan ANOVA Data Numerik Hasil Pengelompokan Metode Terbaik	60
4.5.2 Uji Kruskal-Wallis Data Kategorik Hasil Pengelompokan dengan Metode Terbaik ...	62
BAB V KESIMPULAN DAN SARAN.....	71
5.1 Kesimpulan.....	71
5.2 Saran	73
DAFTAR PUSTAKA	75
LAMPIRAN.....	79
BIODATA PENULIS	121

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Kerangka Proses Analisis Cluster Ensemble	22
Gambar 2.2 Kerangka Proses Pengelompokan SWFM	23
Gambar 3.1 Diagram Alir Penelitian.....	43
Gambar 4.1 Boxplot (a)Presentase Jalan Aspal (b)Rasio SMA per 1000 penduduk (c)Rasio SMA per Penduduk Usia SMA (d)Presentase Pengguna Listrik PLN (e)Prsentase Lahan Kritis (f)Presentase Penduduk Miskin	48
Gambar 4.2 <i>Pie Chart</i> Data Kategorik.....	50
Gambar 4.3 <i>Qq Plot</i> Data Numerik.....	53
Gambar 4.4 Perbandingan (a) Rasio SMA per Penduduk Usia SMA (b) Presentase Luas Lahan Kritis (c) Pengeluaran Penduduk per Kapita	63
Gambar 4.5 Perbandingan Karakteristik (a) Presentase Jalan Aspal (b) Rasio rumah sakit per 1000 penduduk (c) Presentase Pengguna Listrik PLN (d) Presentase Penduduk Miskin	64
Gambar 4.6 Perbandingan Karakteristik (a) Angka Harapn Hidup (b) Angka Melek Huruf (c) Presentase Pengangguran.....	65
Gambar 4.7 Hasil Pemetaan Menggunakan Ensemble ROCK .	69

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1 Perhitungan KMO	10
Tabel 2.2 Ilustrasi Pemeriksaan Distribusi Normal Multivariat.	13
Tabel 2.3 Data Ilustrasi Metode Ward.....	16
Tabel 2.4 Data Ilustrasi Metode ROCK	20
Tabel 2.5 Data Ilustrasi Metode SWFM.....	25
Tabel 2.6 Perhitungan SST.....	27
Tabel 2.7 Perhitungan SSW	28
Tabel 2.8 Data Ilustrasi Perhitungan Akurasi	29
Tabel 2.9 Hasil Ilustrasi Akurasi	30
Tabel 2.10 Analysis of Variance	34
Tabel 3.1 Variabel Penelitian	39
Tabel 3.2 Struktur Data	41
Tabel 4.1 Karakteristik Data Numerik Kabupaten/Kota di Jawa Timur	47
Tabel 4.2 Karakteristik Data Kategorik.....	50
Tabel 4.3 Proporsi Keragaman	53
Tabel 4.4 Loading Factor	54
Tabel 4.5 Hasil Pengelompokan Data Numerik	56
Tabel 4.6 Perbandingan Analisis Klaster Data Numerik	57
Tabel 4.7 Hasil Pengelompokan Metode ROCK.....	58
Tabel 4.8 Struktur Data Analisis Ensemble ROCK dan SWFM	58
Tabel 4.9 Hasil Pengelompokan Ensemble ROCK	59
Tabel 4.10 Hasil Pengelompokan Ensemble SWFM	59
Tabel 4.11 Perbandingan Metode Ensemble ROCK dan SWFM	60
Tabel 4.12 Hasil Uji Box'M.....	61
Tabel 4.13 Hasil One-way ANOVA.....	61
Tabel 4.14 Uji Kruskal-Wallis	62
Tabel 4.15 Urutan Klaster Tiap Variabel	66
Tabel 4.16 Variabel Terendah Setiap Klaster	68
Tabel 4.17 Anggota Klaster.....	69

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data Penelitian Skala Numerik	79
Lampiran 2 Data Penelitian Skala Kategorik	81
Lampiran 3 Syntax Pengelompokan Data Numerik	82
Lampiran 4 Syntax Pengelompokan Data Kategorik.....	88
Lampiran 5 Syntax Pengelompokan Ensemble ROCK	93
Lampiran 6 Syntax Metode SWFM	98
Lampiran 7 Output Statistika Deskriptif	102
Lampiran 8 Output Analisis Faktor.....	103
Lampiran 9 Hasil Pengelompokan Metode Ward.....	105
Lampiran 10 Anggota Kelompok Optimum Metode Ward	106
Lampiran 11 Anggota Kelompok Optimum K-means.....	107
Lampiran 12 Output Pengelompokan Data Kategorik.....	108
Lampiran 13 Output Pengelompokan Ensemble ROCK	111
Lampiran 14 Output Pengelompokan SWFM	115
Lampiran 15 Output One-way MANOVA	117
Lampiran 16 Output ANOVA dan Uji Kruskal-Wallis	118
Lampiran 17 Surat Keterangan Data Penelitian	120

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pembangunan dalam suatu negara merupakan hal yang sangat penting karena tujuan utama dari pelaksanaan pembangunan sendiri untuk meningkatkan kesejahteraan masyarakat dengan memenuhi kebutuhan dan aspirasi mereka (Badan Pusat Statistik, 2015). Pembangunan merupakan rangkaian usaha pertumbuhan dan perubahan yang berencana dan dilakukan secara sadar oleh suatu bangsa, negara dan pemerintah menuju modernitas dalam rangka pembinaan bangsa (Siagian, 2008). Dalam pembangunan nasional mencakup beberapa aspek antara lain ekonomi, politik, sosial budaya dan pertahanan keamanan. Pembangunan nasional dapat diwujudkan ketika terjadi pembangunan yang merata dari setiap wilayah. Pembangunan wilayah bertujuan untuk meningkatkan daya saing wilayah, pertumbuhan ekonomi, mengurangi ketimpangan antar wilayah serta memajukan kehidupan masyarakat. Selain itu, pembangunan wilayah juga mengupayakan keseimbangan pembangunan antar daerah sesuai dengan potensinya masing-masing (Badan Pusat Statistik, 2015). Pembangunan wilayah yang strategis dan berkualitas menjadi harapan setiap daerah di Indonesia.

Salah satu masalah besar dalam pembangunan wilayah adalah tidak meratanya pembangunan antar daerah. Tidak meratanya pembangunan antar daerah menyebabkan ketimpangan serta adanya daerah yang tertinggal jika dibandingkan dengan daerah lainnya. Berdasarkan Peraturan Pemerintah No.78 Tahun 2014, daerah tertinggal merupakan daerah yang masyarakat dan wilayahnya kurang berkembang dibandingkan daerah lain dalam skala nasional. Keberadaan daerah tertinggal merupakan indikator adanya kesenjangan dalam pembangunan. Berdasarkan Peraturan Presiden Nomor 131 Tahun 2015, jumlah Kabupaten/Kota yang dinyatakan sebagai daerah tertinggal mencapai 122 Kabupaten/Kota. Beberapa penyebab ketimpangan pembangunan

antar daerah dalam hal ini Kabupaten/Kota yaitu adanya regulasi yang tidak memihak terhadap percepatan pembangunan, lemahnya kordinasi antar pelaku pembangunan, belum optimalnya kebijakan yang afirmatif, rendahnya kualitas sumber daya manusia dan tingkat kesejahteraan masyarakat, terbatasnya ketersediaan sarana dan prasarana publik dasar, rendahnya produktivitas masyarakat, belum optimalnya pengelolaan potensi sumber daya lokal dalam pengembangan ekonomi, kurangnya aksesibilitas daerah terhadap pusat pertumbuhan dan belum adanya insentif terhadap sektor swasta dan pelaku usaha untuk berinvestasi. Proses identifikasi daerah tertinggal dapat diukur berdasarkan enam kriteria yaitu ekonomi, sumber daya manusia, infrastruktur, kapasitas keuangan daerah, aksesibilitas dan karakteristik daerah (Kementerian DPDTT, 2016). Beberapa fokus upaya pemerintah dalam mengentaskan daerah tertinggal yaitu promosi potensi daerah tertinggal untuk mempercepat pembangunan, upaya pemenuhan kebutuhan dasar dan kebutuhan pelayanan dasar publik, pengembangan perekonomian masyarakat yang didukung oleh sumber daya manusia yang berkualitas dan infrastruktur penunjang konektivitas antara daerah tertinggal dan kawasan strategis.

Salah satu masalah di Jawa Timur adalah pemerataan pembangunan antar daerah. Dibandingkan dengan seluruh Provinsi yang ada di pulau Jawa, Provinsi Jawa Timur merupakan peringkat pertama dengan jumlah Kabupaten/Kota yang dinyatakan sebagai daerah tertinggal terbanyak. Meski mengalami penurunan jumlah Kabupaten/Kota berstatus daerah tertinggal pada periode 2015-2019, hal tersebut belum sesuai dengan target pemerintah dalam mengatasi permasalahan daerah tertinggal (Kementerian DPDTT, 2013). Penelitian yang dilakukan Kementerian DPDTT menggunakan indeks komposit dalam penetapan status daerah tertinggal. Dimana pada setiap indikator memiliki bobot yang berbeda. Selain itu, data yang digunakan hanya bersifat numerik. Oleh karena itu, dilakukan penelitian dengan metode yang berbeda dari penelitian Kementerian DPDTT yaitu dengan memetakan Kabupaten/Kota di Provinsi Jawa Timur

berdasarkan indikator daerah tertinggal yang mengakomodasi data numerik dan kategorik dengan mempertimbangkan aspek statistik. Dimana hasil pemetaan akan menunjukkan karakteristik yang berbeda antar kelompok. Dengan membandingkan setiap kelompok hasil pemetaan, dapat diketahui indikator apa saja yang masih rendah dalam kelompok yang akan diatasi. Selain itu, Pemerintah Kabupaten/Kota dapat saling bertukar informasi mengenai kebijakan yang tepat untuk memperbaiki indikator tersebut sehingga pemerataan pembangunan dapat terwujud. Dewangan, Sharma, dan Akasapu (2010) melakukan transformasi variabel kategorik ke dalam bentuk numerik, kemudian pengelompokan objek dilakukan dengan metode pengelompokan data numerik. Akan tetapi, metode tersebut memiliki kelemahan dalam menentukan transformasi yang tepat agar tidak kehilangan banyak informasi dari original datanya. Sehingga diperlukan metode yang mengakomodasi data campuran numerik dan kategorik. Metode yang digunakan untuk pemetaan dengan analisis pengelompokan pada data campuran yaitu metode *ensemble*. Pengelompokan *ensemble* adalah teknik pengelompokan untuk menggabungkan hasil pengelompokan beberapa algoritma untuk mendapatkan kelompok yang lebih baik (He, Xu, & Deng, 2005). Pada penelitian ini metode *ensemble* yang digunakan yaitu *ensemble* ROCK dan SWFM.

Terdapat beberapa penelitian terkait penggunaan metode *ensemble* ROCK dan SWFM. Alvionita (2017) melakukan penelitian dengan menggunakan metode *ensemble* ROCK dan SWFM pada pengelompokan aksesi jeruk berdasarkan karakteristik data bersifat numerik dan kategorik. Analisis untuk data numerik dengan menggunakan metode hierarki yang menghasilkan *single linkage* sebagai metode terbaik. Sedangkan untuk data kategorik menggunakan metode ROCK. Penggabungan hasil pengelompokan data numerik dan kategorik menggunakan *ensemble* ROCK dan SWFM masing-masing menghasilkan rasio S_w dan S_B yang lebih kecil dibanding metode *single linkage* dan ROCK. Selanjutnya Prakoso (2017) melakukan penelitian untuk

pengelompokan SMA di Sidoarjo menggunakan SWFM. Untuk data numerik menggunakan metode hierarki. Sedangkan untuk data kategorik menggunakan metode *k-modes*. Berdasarkan hasil analisis diperoleh kesimpulan bahwa jumlah kelompok optimum untuk pengelompokan data numerik sebanyak 4 kelompok menggunakan metode *single linkage* dengan nilai rasio S_w dan S_B 0,0201. Sedangkan untuk data kategorik menghasilkan kelompok optimum sebanyak 2 kelompok. Pada pengelompokan *ensemble* SWFM sebanyak 4 kelompok dengan nilai rasio S_w dan S_B sebesar $6,745 \times 10^{-17}$. Penelitian terkait pemetaan daerah tertinggal pernah dilakukan oleh Ramdhany (2017) dengan unit pengamatan desa di Kabupaten Bondowoso, Jawa Timur berdasarkan indeks pembangunan desa. Penelitian tersebut menggunakan metode *ensemble* ROCK karena mengandung data campuran. Hasil *final cluster* dengan menggunakan metode *ensemble* ROCK diperoleh dua kelompok optimum yang terbentuk dengan nilai *threshold* 0,1 dan rasio antara S_w dan S_B sebesar 0,091. Reddy dan Kavitha melakukan penelitian untuk membandingkan metode *k-prototype* dengan SWFM pada *dynamic dataset*. Kesimpulan yang diperoleh metode SWFM lebih baik daripada *k-prototype* karena menghasilkan *error clustering* yang lebih kecil yaitu 0,181 dibanding 0,311.

Penelitian terkait daerah tertinggal pernah dilakukan oleh Kementerian Pembangunan Daerah Tertinggal dengan menggunakan indeks ketertinggalan dimana hasilnya terbagi atas lima status daerah yaitu maju, agak tertinggal, tertinggal, sangat tertinggal dan sangat parah. Namun, penelitian terdahulu terkait daerah tertinggal dengan unit pengamatan Kabupaten hanya mengakomodasi data numerik. Sedangkan pada penelitian ini mengakomodasi data campuran. Dimana data numerik akan dianalisis menggunakan metode hierarki *ward*, yang memiliki keunggulan yaitu meminimumkan varians dalam kelompok dan hasil pengelompokan tidak ada yang memiliki satu anggota (Gong & Richman, 1995). Metode klaster *ward* akan dibandingkan dengan metode hierarki *complete linkage* dan non hierarki *k-means*.

Data kategorik akan dianalisis dengan menggunakan metode ROCK. Metode ini merupakan pengembangan dari metode hierarki untuk data kategorik. Kelebihan metode ROCK adalah memiliki akurasi yang lebih baik dibanding metode hierarki agglomerative dengan sifat skalabilitas yang baik. Penggabungan hasil pengelompokan data campuran menggunakan metode *ensemble* ROCK dan SWFM. Keunggulan dari SWFM adalah efisien untuk data dengan segala dimensi, mengurangi kompleksitas waktu, dan pencarian jarak yang digunakan lebih efisien (Reddy & Kavitha, 2012). Variabel yang digunakan dalam penelitian ini adalah indikator yang mewakili 6 kriteria dalam penetapan daerah tertinggal yaitu ekonomi, sumber daya manusia, infrastruktur, kapasitas keuangan daerah, aksesibilitas dan karakteristik daerah dan data potensi wilayah. Diharapkan hasil pemetaan dapat membantu pemerintah dalam menentukan kebijakan yang sesuai dengan karakteristik masing-masing daerah sehingga dapat mencapai target yang telah ditetapkan.

1.2 Rumusan Masalah

Salah satu masalah dalam pembangunan wilayah adalah pemerataan pembangunan antar daerah. Dibandingkan dengan seluruh Provinsi yang ada di pulau Jawa, Provinsi Jawa Timur merupakan Provinsi dengan jumlah daerah yang termasuk kategori daerah tertinggal terbanyak yaitu sebanyak 4 Kabupaten/Kota. Penelitian sebelumnya yang dilakukan Kementerian DPDTT terkait daerah tertinggal dengan unit pengamatan Kabupaten hanya mengakomodasi data numerik. Oleh karena itu, dilakukan penelitian dengan metode yang berbeda dari penelitian sebelumnya yaitu dengan memetakan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator daerah tertinggal yang mengakomodasi data numerik dan kategorik dengan mempertimbangkan aspek statistik. Pada penelitian ini diterapkan metode analisis pengelompokan ensemble data campuran numerik dan kategorik yaitu ensemble ROCK dan SWFM. Proses identifikasi daerah tertinggal diukur berdasarkan enam kriteria

yaitu ekonomi, sumber daya manusia, infrastruktur, kapasitas keuangan daerah, aksesibilitas dan karakteristik daerah.

1.3 Tujuan Penelitian

Berdasarkan rumusan permasalahan yang telah disusun, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mendeskripsikan karakteristik Kabupaten/Kota di Jawa Timur berdasarkan indikator daerah tertinggal.
2. Mengelompokkan Kabupaten/Kota di Jawa Timur berdasarkan indikator daerah tertinggal menggunakan metode *ensemble* ROCK.
3. Mengelompokkan Kabupaten/Kota di Jawa Timur berdasarkan indikator daerah tertinggal menggunakan metode *ensemble* SWFM.
4. Menentukan metode pengelompokan terbaik dan hasil pengelompokan beserta karakteristiknya.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut:

1. Hasil dari penelitian diharapkan bisa digunakan oleh Pemerintah Provinsi Jawa Timur sebagai bahan pertimbangan dalam menetapkan kebijakan dalam rangka mengatasi daerah tertinggal di Jawa Timur.
2. Memberikan pengetahuan dan informasi mengenai penerapan metode *ensemble* ROCK dan SWFM dalam pemetaan Kabupaten/Kota di Jawa Timur yang dapat dijadikan referensi untuk penelitian selanjutnya.

1.5 Batasan Masalah

Batasan Batasan masalah yang digunakan dalam penelitian ini adalah data yang digunakan merupakan data potensi desa dan susenas tahun 2014. Analisis faktor hanya dilakukan untuk data numerik. Data penelitian diasumsikan berdistribusi normal multivariat.

BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka pada penelitian ini membahas tentang statistika deskriptif, metode pengelompokan data campuran *ensemble* ROCK dan SWFM, serta kinerja pengelompokan untuk metode yang digunakan. Pembahasan akan dimulai pada statistika deskriptif.

2.1 Statistika Deskriptif

Statistika deskriptif ialah metode yang berhubungan dengan pengumpulan, penyusunan dan penyajian suatu data dengan tujuan memberikan informasi yang bermanfaat. Statistika deskriptif tidak bisa digunakan untuk menarik kesimpulan pada suatu data, melainkan hanya memberikan informasi terkait data tersebut. Statistika deskriptif dapat disajikan melalui perhitungan dan grafik. Berdasarkan perhitungan, statistika deskriptif terbagi atas ukuran pemusatan data dan penyebaran data. Ukuran pemusatan data antara lain rata-rata (*mean*), nilai tengah (*median*), dan nilai yang paling sering muncul (*mode*). Contoh ukuran penyebaran data adalah varians, simpangan baku, dan nilai jarak (Walpole, *et al*, 2007).

Ukuran pemusatan data yang umum digunakan adalah rata-rata. Rata-rata dari suatu data adalah penjumlahan dari setiap observasi yang dibagi dengan jumlah observasi itu sendiri (Weiss, 2008). Maka nilai dari rata-rata dapat dihitung dengan menggunakan persamaan 2.1,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.1)$$

dengan \bar{x} adalah nilai rata-rata, x_i merupakan nilai pengamatan ke- i dan n merupakan jumlah pengamatan yang dilakukan.

Statistika deskriptif dapat disajikan secara visual melalui grafik. Grafik yang dimaksud seperti diagram batang, diagram pencar, histogram dan lain-lain. Analisis statistika deskriptif untuk data numerik dapat disajikan dengan menggunakan

boxplot. Komponen yang terdapat pada *boxplot* antara lain kuartil pertama, median, kuartil ketiga, nilai minimum dan maksimum. *Boxplot* juga dapat digunakan untuk identifikasi adanya *outlier* (Johnson & Bhattacharyya, 2010).

2.2 Analisis Faktor

Analisis faktor digunakan untuk mengidentifikasi struktur hubungan antarvariabel (Hair, *et al*, 2010). Faktor-faktor yang dihasilkan dapat mewakili variabel. Vektor random teramati X dengan p komponen, memiliki rata-rata μ dan matrik kovarian.

Model analisis faktor adalah sebagai berikut :

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned} \quad (2.2)$$

dengan,

μ_i = rata-rata variabel i

ε_i = faktor spesifik ke $-i$

F_j = common faktor ke- j

ℓ_{ij} = loading dari variabel ke $-i$ pada faktor ke- j

Atau dapat dituliskan dalam notasi matriks sebagai berikut :

$$\mathbf{X}_{px1} - \boldsymbol{\mu}_{(px1)} = +\mathbf{L}_{(pxm)}\mathbf{F}_{(mx1)} + \boldsymbol{\varepsilon}_{px1} \quad (2.3)$$

dengan,

$$\mathbf{X} = (X_1, X_2, \dots, X_p)'$$

$$\mathbf{F} = (F_1, F_2, \dots, F_m)'$$

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)'$$

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)'$$

$$\mathbf{L} = \begin{pmatrix} L_{11} & L_{12} & \dots & L_{1m} \\ L_{21} & L_{22} & \dots & L_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ L_{p1} & L_{p2} & \dots & L_{pm} \end{pmatrix}$$

Bagian dari varian variabel ke $-i$ dari m common factor disebut komunalitas ke $-i$ yang merupakan jumlah kuadrat dari loading

variabel ke- i pada m common factor (Johnson & Wichern, 2007) seperti pada persamaan 2.4.

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 \quad (2.4)$$

dengan,

$$\ell_{ij} = \sqrt{\lambda_i} e_{ij} \quad (2.5)$$

Keterangan:

λ_i : eigenvalue variabel ke- i

e_{ij} : eigenvectore variabel ke- i faktor ke- j

Tujuan analisis faktor adalah menggunakan matriks korelasi/kovarians untuk mengidentifikasi jumlah terkecil dari faktor umum (yaitu model faktor yang paling parsimonii) yang mempunyai penjelasan terbaik atau menghubungkan variabel yang berkorelasi tinggi kedalam suatu faktor. Dimana antar faktor yang terbentuk tidak berkorelasi. Penggunaan matriks korelasi/kovarians merujuk pada data yang digunakan. Jika data tidak distandarisasi maka analisis faktor menggunakan matriks korelasi. Penentuan jumlah faktor yang terbentuk berdasarkan eigenvalue lebih dari 1 atau varians kumulatif lebih dari 60% (Hair, et al, 2010). Pada penelitian ini, analisis faktor digunakan untuk mengurangi dampak multikolinearitas pada analisis cluster. Dimana dalam setiap faktor yang terbentuk digunakan variabel yang memiliki kontribusi terbesar yang digunakan dalam pengelompokan. Dengan menghapus fitur yang memiliki korelasi rendah akan menghasilkan metrik jarak dan hasil pengelompokan yang robust (Ben-Hur & Guyon, 2003).

Sebelum dilakukan analisis faktor, diperlukan beberapa pemeriksaan untuk mengetahui kelayakan analisis faktor antara lain identifikasi korelasi antar variabel, uji korelasi antar variabel dan pemeriksaan asumsi distribusi normal multivariat.

a. Identifikasi kecukupan korelasi antar variabel

Pemeriksaan kecukupan korelasi antar variabel menggunakan nilai KaiserMeyer-Olkin (KMO) yang

merupakan nilai korelasi antar variabel. Nilai KMO sesuai dengan persamaan 2.6.

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (2.6)$$

dengan,

r_{ij} = koefisien korelasi antar variabel i dan j

a_{ij} = koefisien korelasi parsial antar variabel i dan j

perhitungan KMO diawali dengan menghitung nilai korelasi dan korelasi parsial antar variabel. Ilustrasi perhitungan KMO adalah sebagai berikut.

Tabel 2.1 Perhitungan KMO

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_1X_3	X_2X_3
2.5	4	3	6.25	16	9	10	7.5	12
2	2	2.5	4	4	6.25	4	5	5
4	3	2	16	9	4	12	8	6
2	4	5	4	16	25	8	10	20
1	8	8	1	64	64	8	8	64
6	2	4	36	4	16	12	24	8
1.5	3	3	2.25	9	9	4.5	4.5	9
1.2	2.5	6	1.44	6.25	36	3	7.2	15
2.9	4	2	8.41	16	4	11.6	5.8	8
4	5	3	16	25	9	20	12	15

Selanjutnya dilakukan perhitungan untuk memperoleh korelasi dan korelasi parsial yang digunakan dalam perhitungan KMO. Perhitungan nilai korelasi dan korelasi parsial antara variabel X_1 dan X_3 seperti berikut.

$$\begin{aligned}
r_{x_1x_2} &= \frac{n \sum x_1x_3 - (\sum x_1)(\sum x_3)}{\sqrt{n \sum x_1^2 - (\sum x_1)^2} \times \sqrt{n \sum x_3^2 - (\sum x_3)^2}} \\
&= \frac{10(92) - (27.1)(38.5)}{\sqrt{10(95.35) - (27.1)^2} \times \sqrt{10(182.3) - (38.5)^2}} \\
&= \frac{-123}{273} = -0.45
\end{aligned}$$

Berdasarkan perhitungan tersebut, diperoleh $r_{x_1x_3}$ sebesar -0.45. perhitungan dilanjutkan sehingga diperoleh nilai $r_{x_1x_2}$ dan $r_{x_2x_3}$ berturut-turut sebesar -0.34 dan 0.565. selanjutnya dilakukan perhitungan korelasi parsial sebagai berikut.

$$\begin{aligned}
a_{x_1(x_2x_3)} &= \frac{r_{x_2x_3} - (r_{x_1x_2})(r_{x_1x_3})}{\sqrt{(1-r_{x_1x_2}^2)(1-r_{x_1x_3}^2)}} \\
&= \frac{0.565 - (-0.34)(-0.45)}{\sqrt{(1-(-0.34)^2)(1-(-0.45)^2)}} \\
&= \frac{0.411}{0.839} = 0.49
\end{aligned}$$

Berdasarkan perhitungan di atas, diperoleh nilai $a_{x_1(x_2x_3)}$ sebesar 0.49. Perhitungan dilanjutkan sehingga diperoleh nilai $a_{x_2(x_1x_3)}$ dan $a_{x_3(x_1x_2)}$ berturut-turut sebesar -0.334 dan -0.12. Selanjutnya dilakukan perhitungan berdasarkan persamaan 2.5, sehingga nilai KMO yang dihasilkan sebesar 0.636. Korelasi variabel dikatakan cukup ketika nilai KMO lebih besar dari 0,5 (Johnson & Wichern, 2007)

b. Uji dependensi antar variabel

Uji dependensi dilakukan untuk menguji hubungan antar variabel dimana X_1, X_2, \dots, X_p dikatakan saling bebas jika matriks korelasi antar variabel membentuk matriks identitas menggunakan uji *Bartlett* dengan hipotesis sebagai berikut.

$H_0 : \mathbf{p} = \mathbf{I}$ (matriks korelasi sama dengan matriks identitas)

$H_1 : \mathbf{p} \neq \mathbf{I}$ (matriks korelasi tidak sama dengan matriks identitas)

dengan statistik uji didapatkan dari perhitungan sesuai persamaan 2.7.

$$\chi^2_{hitung} = - \left\{ n - 1 - \frac{2p + 5}{6} \right\} \ln |\mathbf{R}| \quad (2.7)$$

dengan,

n : jumlah observasi

p : jumlah variabel

$|R|$: determinan dari matriks korelasi

Ilustrasi perhitungan uji *Bartlett* menggunakan nilai *chi-square* dengan data yang disajikan pada Tabel 2.1. Tahap awal adalah menghitung korelasi antar variabel yang sudah dijelaskan pada subbab 2.2 (a). sehingga nilai korelasi disajikan dalam bentuk matriks korelasi.

$$R = \begin{pmatrix} 1 & -0.34 & -0.45 \\ -0.34 & 1 & 0.565 \\ -0.45 & 0.565 & 1 \end{pmatrix}$$

Setelah matriks korelasi terbentuk. Dilakukan perhitungan berdasarkan persamaan 2.6. Sehingga diperoleh nilai *chi-square* sebesar 4.47. Dimana Daerah kritis dari pengujian ini adalah H_0 ditolak ketika $\chi_{\text{hitung}}^2 > \chi_{\alpha, \frac{1}{2}(p-1)}^2$ (Hair, et al,2010).

c. Pemeriksaan Asumsi Distribusi Normal Multivariat

Pemeriksaan asumsi distribusi normal multivariat menggunakan metode *Shapiro-wilk test*. Hipotesis yang digunakan dalam pengujian disitribusi normal multivariat adalah sebagai berikut

H_0 : Data mengikuti distribusi normal multivariat

H_1 : Data tidak mengikuti distribusi normal multivariat

Algoritma pemeriksaan distribusi normal multivariat menggunakan *Shapiro-wilk* sebagai berikut:

1. Mengurutkan data pengamatan dari yang terkecil sampai terbesar.
2. Menghitung nilai S^2

$$S^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.8)$$

3. (a) Jika n genap, $m = n/2$, maka

$$b = \sum_{i=1}^m a_{n-i+1} (x_{n-i+1} - x_i) \quad (2.9)$$

- (b) Jika n ganjil, $m = n - 1/2$

$$b = a_n (x_n - x_1) + \dots + a_{m+2} (x_{m+2} - x_m) \quad (2.10)$$

dimana nilai x_{m+1} adalah median dan nilai a merupakan koefisien tabel *normality test*

4. Menghitung nilai W_k

$$W_k = \frac{b^2}{S^2} \quad (2.11)$$

Tahap selanjutnya adalah menghitung statistik uji secara multivariat W^*

$$W^* = \frac{1}{p} \sum_{k=1}^p W_k \quad (2.12)$$

dengan p merupakan jumlah variabel.

Ilustrasi pemeriksaan distribusi normal multivariat menggunakan *Shapiro-wilk* menggunakan data pada Tabel 2.1. dengan langkah-langkah sebagai berikut:

1. Mengurutkan data pengamatan dari yang terkecil sampai terbesar kemudian menghitung nilai s^2 sesuai persamaan 2.7. Hasil perhitungan disajikan pada Tabel 2.2. Ilustrasi berikut

Tabel 2.2 Ilustrasi Pemeriksaan Distribusi Normal Multivariat

Obj	X ₁	X ₂	X ₃
1	1	2	2
2	1.2	2	2
3	1.5	2.5	2.3
4	2	3	3
5	2	3	3
6	2.5	4	3.3

Tabel 2.2 Ilustrasi Pemeriksaan Distribusi Normal Multivariat (Lanjutan)

Obj	X ₁	X ₂	X ₃
7	2.9	4	4
8	4	4	5
9	4	5	6
10	6	8	8
Total	27.1	37.5	38.6
s²	21.9	28.6	34.2

2. Menghitung nilai b . pada data ilustrasi banyaknya objek pengamatan adalah 10. Sehingga menggunakan persamaan 2.8. Dimana nilai b masing-masing variabel adalah 4.456, 4.914 dan 5.472.
 3. Menghitung W_k secara univariat.
- $$W_1 = 0.906 \qquad \qquad \qquad W_3 = 0.876$$
- $$W_2 = 0.844$$
4. Menghitung nilai W^* yaitu rata-rata dari W_1 , W_2 , dan W_3 yaitu 0.875.

Variabel dikatakan mengikuti distribusi normal multivariat jika nilai statistik uji W^* mendekati 1, namun jika nilai statistik uji W^* kecil atau jauh dari 1 maka dikatakan tidak mengikuti distribusi normal multivariat. Daerah penolakan H_0 yaitu saat $W^* > C_{(\alpha,n)}$ (Alva & Estrada, 2009).

2.3 Analisis Klaster (*Cluster Analysis*)

Analisis klaster adalah salah satu teknik yang primitif karena tidak membutuhkan asumsi dalam pengelompokan data. Pengelompokan yang dilakukan berdasarkan pada kemiripan dan ketidakmiripan (Johnson & Wichern, 2007). Analisis klaster digunakan untuk mengelompokkan observasi ke dalam suatu kelompok berdasarkan kemiripan atau homogenitas antar anggota kelompok. Sedangkan antar kelompok memiliki ketidakmiripan atau heterogenitas (Hair, et al, 2010). Hasil *clustering* dipengaruhi oleh objek yang akan dikelompokkan,

peubah yang diamati, ukuran kemiripan, skala ukuran yang digunakan serta metode klaster yang digunakan.

2.3.1 Pengelompokan Data Numerik

Analisis pengelompokan untuk data numerik didasarkan pada ukuran ketidakmiripan dari data numerik yang diamati. Dalam pengelompokan untuk data numerik terdapat beberapa metode yang terbagi atas metode hierarki dan nonhierarki. Metode hierarki dimulai dengan pengelompokan dua atau lebih objek yang memiliki kemiripan paling dekat. Begitu seterusnya sehingga klaster akan membentuk tingkatan yang jelas antar objek dari yang terdekat hingga terjauh (Hair, *et al*, 2010). Penyajian hasil pengelompokan biasanya dengan menggunakan dendogram yang membantu memberikan informasi hasil pengelompokan secara visual.

Metode hierarki klaster terbagi atas dua tipe, yaitu *agglomerative* (pemusatan) dan *divisive* (penyebaran). Pada penelitian ini akan digunakan metode klaster hierarki *agglomerative*. Pada metode *agglomerative*, setiap objek dianggap sebagai suatu klaster. Selanjutnya dua objek yang memiliki kemiripan dikelompokkan menjadi satu klaster baru. Penentuan kemiripan dilakukan dengan menghitung jarak antar klaster (Johnson & Wichern, 2007). Sedangkan metode non hierarki yang digunakan adalah metode *k-means*.

a) Metode hierarki

Metode pengelompokan hierarki *agglomerative* yang digunakan pada penelitian ini antara lain:

1. *Ward's method*

Metode *ward* merupakan pengelompokan objek-objek didasarkan dengan memaksimalkan kehomogenan dalam kelompok. Dalam metode ini jarak antara dua *cluster* adalah jumlah kuadrat antara dua *cluster* untuk seluruh variabel. Pengelompokan metode *ward* adalah meminimalkan peningkatan kriteria *error sum of square* (ESS). Dua *cluster* yang memiliki peningkatan ESS paling minimum akan berkelompok. Sehingga untuk menghitung jarak antara dua

cluster menggunakan metode *ward* diperoleh berdasarkan persamaan 2.13.

$$ESS = \sum_{j=1}^N (\mathbf{x}_j - \bar{\mathbf{x}})' (\mathbf{x}_j - \bar{\mathbf{x}}) \quad (2.13)$$

dimana x_j merupakan pengukuran multivariat terkait dengan item j dan \bar{x} adalah rata-rata dari semua item. Metode *ward's* digunakan karena mampu meminimalkan varians dalam setiap *cluster* (Gong & Richman, 1995).

Berikut ilustrasi pengelompokan menggunakan metode *ward* dengan data ilustrasi yang disajikan pada Tabel 2.3

Tabel 2.3 Data Ilustrasi Metode Ward

Objek	A	B	C	D	E
X1	2	2	1	1	3
X2	2	1	2	2	2

Langkah awal adalah menghitung *ESS* antar objek sesuai dengan persamaan 2.13. pada ilustrasi ini dilakukan perhitungan $ESS_{(A,B)}$ dan $ESS_{(A,E)}$ sebagai berikut:

$$ESS_{(A,B)} = (2-2)^2 + (2-2)^2 + (2-1.5)^2 + (1-1.5)^2 = 0.5$$

$$ESS_{(A,E)} = (2-2.5)^2 + (3-2.5)^2 + (2-2)^2 + (2-2)^2 = 0.5$$

Setelah diperoleh *ESS* antar seluruh objek, nilai *ESS* disajikan dalam bentuk matriks sebagai berikut:

$$\mathbf{ESS} = \begin{pmatrix} A & 1 \\ B & 0,5 & 1 \\ C & 0 & 0 & 1 \\ D & 0,2 & 0,2 & 0,2 & 1 \\ E & 0,2 & 0,5 & 0 & 0,5 & 1 \end{pmatrix}$$

Penggabungan objek menjadi satu kelompok didasarkan pada *ESS* terkecil. Pada ilustrasi ini, objek A dan C menjadi satu kelompok. Selanjutnya dilakukan perhitungan *ESS* yang baru sampai terbentuk kelompok yang diinginkan.

2. Complete linkage

Persamaan untuk metode *complete linkage* disajikan pada persamaan 2.14.

$$d_{w(UV)} = \max(d_{wu}, d_{wv}) \quad (2.14)$$

Prosedur yang dilakukan pada metode ini didasarkan pada jarak terjauh dari objek (Johnson & Wichern, 2007).

b) Metode non hierarki

Metode non hierarki yang digunakan yaitu metode *k-means*. *K-means* merupakan metode analisis klaster non hierarki yang berusaha mempartisi data kedalam bentuk kelompok. Berikut algoritma *k-means*:

1. Menentukan k sebagai jumlah klaster yang diinginkan
2. Mengalokasikan data kedalam klaster
3. Menentukan pusat klaster (*centroid*) sesuai persamaan 2.15

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \quad (2.15)$$

dengan:

C_{kj} : *centroid* ke- k pada variabel ke- j , dimana $j = 1, 2, \dots, p$

n : banyaknya data

4. Menentukan jarak setiap objek dengan setiap *centroid* menggunakan jarak *euclidean*.
5. Menghitung kembali *centroid* kelompok ketika menerima item baru maupun item keluar (Johnson & Wichern, 2007).

2.3.2 Pengelompokan Data Kategorik

Metode tradisional untuk pengelompokan data kategorik yang didasarkan pada ukuran kemiripan atau jarak untuk data kategorik dinilai tidak tepat digunakan untuk data kategorik. Sehingga dikembangkan metode pengelompokan hierarki *agglomerative* untuk data kategorik yaitu metode ROCK. Metode ROCK dibentuk dengan menggunakan konsep baru yaitu *link*. *Link* digunakan untuk mengukur kesamaan/kedekatan antara sepasang titik data. Apabila terdapat pengamatan yang memiliki tingkat hubungan (*link*) tinggi maka dapat digabungkan menjadi satu kelompok sedangkan yang

memiliki tingkat hubungan rendah dipisahkan dari kelompok dimana data tersebut dikelompokkan. Kelebihan metode ROCK adalah memiliki akurasi yang lebih baik dibanding metode hierarki agglomerative dengan sifat skalabilitas yang baik.

Pengelompokan dengan metode ROCK terdiri dari beberapa langkah. Berikut ini adalah langkah-langkah pengelompokan dengan metode ROCK (Dutta, Mahanta, & Arun, 2005)

1. Menentukan inisialisasi untuk masing-masing data poin sebagai *cluster* pada awalnya.
2. Menghitung Similaritas. Ukuran kemiripan antara pasangan pengamatan ke-*i* dan ke-*j* dapat dihitung dengan rumus pada persamaan 2.16,

$$sim(X_i, X_j) = \frac{X_i \cap X_j}{X_i \cup X_j}, i \neq j \quad (2.16)$$

dengan,

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

X_i : Himpunan pengamatan ke-*i* dengan

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

X_j : Himpunan pengamatan ke-*j* dengan

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

3. Menentukan *Neighbors*

Pengamatan X_i dan X_j dinyatakan sebagai tetangga (*neighbors*) apabila nilai $sim(X_i, X_j) \geq \theta$. Nilai *threshold* (θ) yang digunakan berkisar antara 0 sampai 1 dan dapat ditentukan dengan menyesuaikan data yang ada.

4. Menghitung *Link*

Link (X_i, X_j) antar objek diperoleh dari jumlah *common neighbor* antara X_i dan X_j . Besarnya *link* dipengaruhi oleh nilai *threshold* (θ) yang telah ditentukan. Jika nilai *link* (X_i, X_j) besar, maka kemungkinan X_i dan X_j berada pada kelompok yang sama. 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_i tetangga dari X_k maka dikatakan X_i memiliki *link* dengan X_k walaupun X_i bukan tetangga dari X_k . Cara untuk menghitung *link* untuk semua kemungkinan pasangan dari n objek dapat menggunakan matriks \mathbf{A} . Matriks \mathbf{A} 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* antara 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 suatu kelompok yang sama. Pengelompokan pada algoritma ROCK akan berhenti ketika keadaaan jumlah dari kelompok yang diharapkan sudah terpenuhi atau tidak ada *link* antar kelompok (Dutta, Mahanta, & Arun, 2005).

5. *Goodness Measure*

Penggabungan kelompok pada algoritma ROCK didasarkan pada ukuran kebaikan atau *goodness measure* antar kelompok. *Goodness measure* merupakan persamaan yang menghitung jumlah *link* dibagi dengan kemungkinan *link* yang terbentuk berdasarkan ukuran kelompoknya sesuai dengan persamaan 2.17,

$$g(C_i, C_j) = \frac{\text{link}(C_i, C_j)}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}} \quad (2.17)$$

dengan $\text{link}(C_i, C_j) = \sum_{X_i \in C_i, X_j \in C_j} \text{link}(X_i, X_j)$ merupakan

jumlah *link* dari semua kemungkinan pasangan objek yang ada dalam C_i dan C_j . Sedangkan n_i dan n_j adalah jumlah anggota dalam kelompok ke- i dan kelompok ke- j

dan $f(\theta) = \frac{1-\theta}{1+\theta}$, dimana θ merupakan nilai *threshold* yang digunakan.

Selain algoritma, akan disajikan ilustrasi pengelompokan secara manual menggunakan metode ROCK:

1. Data yang digunakan disajikan pada Tabel 2.4.

Tabel 2.4 Data Ilustrasi Metode ROCK

Objek	A	B	C	D	E
X1	1	2	3	3	2
X2	2	2	1	2	2
X3	3	3	2	1	1

2. Menghitung similaritas antar objek sehingga membentuk matriks similaritas:

$$\begin{aligned} sim(A, B) &= \frac{A \cap B}{A \cup B} = \frac{n(A \cap B)}{n(A) + n(B) - n(A \cap B)} \\ &= \frac{2}{3+3-2} = 0,5 \end{aligned}$$

Perhitungan tersebut dilakukan sampai membentuk matriks similaritas

$$\mathbf{sim} = C \begin{pmatrix} A & 1 \\ B & 0,5 & 1 \\ C & 0 & 0 & 1 \\ D & 0,2 & 0,2 & 0,2 & 1 \\ E & 0,2 & 0,5 & 0 & 0,5 & 1 \end{pmatrix}$$

3. Menentukan *neighbors*, dimana objek dikatakan bertetangga jika nilai $sim(X_i, X_j) \geq \theta$. Nilai threshold yang digunakan 0,01. Keseluruhan nilai tersebut dikonversi dalam bentuk matriks A.

$$\mathbf{A} = C \begin{pmatrix} A & 1 & 1 & 0 & 1 & 1 \\ B & 1 & 1 & 0 & 1 & 1 \\ C & 0 & 0 & 1 & 1 & 0 \\ D & 1 & 1 & 1 & 1 & 1 \\ E & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

4. Menghitung *link* antar objek dengan melakukan perkalian matriks $\mathbf{A}^* \mathbf{A}$. sehingga diperoleh matriks \mathbf{L} pada halaman selanjutnya.

$$\mathbf{L} = \mathbf{C} \begin{pmatrix} 4 & 4 & 1 & 4 & 4 \\ 4 & 4 & 1 & 4 & 4 \\ 1 & 1 & 2 & 2 & 1 \\ 4 & 4 & 2 & 5 & 4 \\ 4 & 4 & 1 & 4 & 4 \end{pmatrix}$$

5. Menghitung *goodness measure* sesuai persamaan 2.17.

$$g(A, B) = \frac{4}{(1+1)^{(1+2(0,98))} - (1)^{(1+2(0,98))} - (1)^{(1+2(0,98))}} = 0,692$$

Perhitungan *goodness measure* dilakukan untuk setiap pasangan objek hingga membentuk matriks \mathbf{G} seperti berikut.

$$\mathbf{G} = \mathbf{C} \begin{pmatrix} 0,692 & 0,692 & 0,172 & 0,692 & 0,692 \\ 0,692 & 0,692 & 0,172 & 0,692 & 0,692 \\ 0,173 & 0,172 & 0,346 & 0,346 & 0,172 \\ 0,692 & 0,692 & 0,346 & 0,865 & 0,692 \\ 0,692 & 0,692 & 0,172 & 0,692 & 0,692 \end{pmatrix}$$

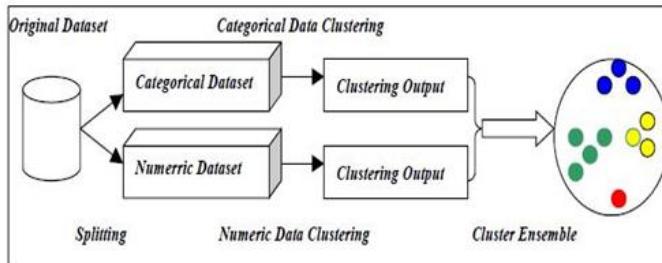
Setelah terbentuk matriks \mathbf{G} , objek dengan *goodness measure* terbesar membentuk satu kelompok. Dimana objek A dan B berkelompok. Kemudian kembali ke langkah 5 untuk menghitung *goodness measure* yang baru sampai diperoleh kelompok yang diinginkan.

2.4 Pengelompokan Data Numerik dan Kategorik

Kumpulan data dalam berbagai kasus di dunia nyata banyak mengandung data campuran dengan berbagai skala. *Cluster ensemble* merupakan metode untuk menggabungkan beberapa algoritma *clustering* yang berbeda untuk mendapatkan partisi umum dari kumpulan data asli. Salah satu algoritma yang ada menggunakan ukuran kemiripan seperti jarak *Euclidean* yang memberikan hasil yang baik untuk data numerik. Namun, hal tersebut tidak berlaku untuk data kategorik. Pada *cluster*

ensemble pendekatan untuk data numerik dan kategorik dilakukan secara terpisah (Reddy & Kavitha, 2012).

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* dengan data campuran dapat menggunakan algoritma CEBMDC (*Cluster Ensemble Based Mixed Data Clustering*) yang ditunjukkan pada Gambar 2.1.



Gambar 2.1 Kerangka Proses Analisis Cluster Ensemble

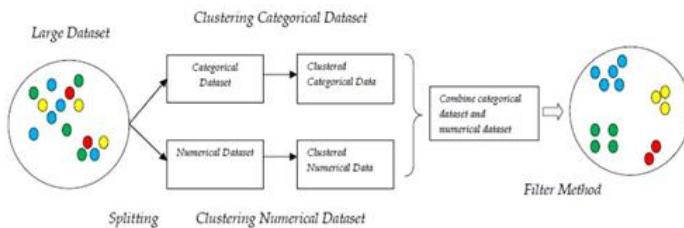
Gambar 2.1 menunjukkan kerangka konsep dalam analisis *cluster ensemble*, pada penelitian ini penjabaran langkah-langkah dalam analisis *cluster ensemble* adalah sebagai berikut.

1. Membagi data menjadi dua bagian yaitu data numerik dan data kategorik
2. Melakukan analisis klaster menggunakan metode *ward* untuk data numerik
3. Melakukan analisis klaster menggunakan metode ROCK untuk data kategorik
4. Menggabungkan hasil klaster untuk data numerik dan kategorik. Penggabungan ini disebut proses *ensemble*
5. Melakukan analisis klaster pada data campuran menggunakan *ensemble* ROCK dan SWFM. Tahap analisis terakhir pada metode *ensemble* ROCK sama

dengan tahap analisis klaster untuk data kategorik dimana hasil dari proses *ensemble* akan dianalisis dengan metode ROCK. Sedangkan untuk tahap analisis metode *ensemble* SWFM dijelaskan pada sub bab 2.5.

2.5 *Similarity Weight and Filter Method* (SWFM)

Pengelompokan menggunakan SWFM mempunyai konsep yang sama dengan *cluster ensemble* pada umumnya. Namun, pada metode ini memiliki perbedaan algoritma pada tahap pembentukan klaster terakhir. Pada metode *cluster ensemble* umumnya dalam pembentukan klaster terakhir menggunakan algoritma yang sama dengan pembentukan klaster untuk data kategorik, sementara SWFM menggunakan algoritma berdasarkan *similarity weight and filter method* (Reddy & Kavitha, 2012). Kerangka konsep dalam pengelompokan menggunakan SWFM dijelaskan pada Gambar 2.2.



Gambar 2.2 Kerangka Proses Pengelompokan SWFM

Gambar 2.2 menunjukkan proses pengelompokan menggunakan SWFM mulai dari pembagian data numerik dan kategorik sampai pembentukan akhir klaster. Pada tahap *similarity weight* digunakan ukuran kemiripan yang memasukkan pembobotan faktor pada rumus ukuran kemiripan. Pembobot yang diberikan berdasarkan jumlah anggota pengamatan (n_i) atau (n_j). Maka rumus yang diigunakan untuk menghitung ukuran kemiripan diperoleh berdasarkan persamaan 2.18.

$$\text{sim}(X_i, X_j) = \frac{|X_i \cap X_j|}{|X_i \cup X_j|} \quad (2.18)$$

dengan,

X_i : Himpunan pengamatan ke- i dengan $X_i = \{x_{1i}, x_{2i}, \dots, x_{ki}\}$

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

n : Total jumlah objek pengamatan

n_i : Jumlah anggota dalam kelompok ke- i

n_j : Jumlah anggota dalam kelompok ke- j

Tahap selanjutnya yaitu menggabungkan hasil pengelompokan berdasarkan *similarity weight* dengan menggunakan *filter algorithm* pada persamaan 2.19.

$$F(X_i, X_j) = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w_{ij} d(X_i, X_j) \quad (2.19)$$

dimana $d(X_i, X_j) = 1 - \text{sim}(X_i, X_j)$, sedangkan w_{ij} adalah bobot antara kelompok ke- i dan kelompok ke- j hasil pengelompokan dengan dua algoritma yang berbeda dimana pada umumnya bernilai 0,5 atau $0 < \text{sim}(X_i, X_j) \leq 1$.

Terdapat berbagai keunggulan dalam menggunakan SWFM untuk *cluster ensemble* antara lain penggunaan yang efisien untuk memotong dan mengolah properti, bentuk dari klaster mempunyai dampak yang sangat kecil terhadap kinerja algoritma *filter clustering*, efisien untuk data dengan segala dimensi dan mengurangi kompleksitas waktu, pencarian jarak terdekat digunakan untuk menentukan *filter* yang efisien, dan bekerja dengan efisien meskipun batas-batas kelompok tidak beraturan (Reddy & Kavitha, 2012). Ilustrasi pengelompokan menggunakan metode SWFM adalah sebagai berikut:

1. Menggabungkan hasil pengelompokan numerik dan kategorik. Dimana hasil pengelompokan dianggap sebagai variabel baru. Data yang digunakan disajikan pada Tabel 2.5.

Tabel 2.5 Data Ilustrasi Metode SWFM

Objek	A	B	C	D	E
X1	1	2	3	3	2
X2	2	2	1	2	2

2. Menghitung similaritas antar objek menggunakan persamaan 2.18.

$$\begin{aligned} sim(A, B) &= \frac{A \cap B}{A \cup B} = \frac{n(A \cap B)}{n(A) + n(B) - n(A \cap B)} \\ &= \frac{1}{2+2-1} = 0,333 \end{aligned}$$

Diperoleh similaritas antara objek A dan B sebesar 0,333. Perhitungan dilakukan sampai memperoleh seluruh similaritas antar objek. Dimana nilai similaritas disajikan dalam bentuk matriks sebagai berikut.

$$\mathbf{sim} = C \begin{pmatrix} 1 & & & & \\ 0,33 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0,33 & 0,33 & 1 & \\ 0,33 & 1 & 0 & 0,33 & 1 \end{pmatrix}$$

3. Menghitung *filter* menggunakan persamaan 2.18. Dimana bobot yang digunakan pada penelitian ini adalah 0,5. Hasil perhitungan *filter* antar objek disajikan dalam bentuk matriks F seperti berikut.

$$\mathbf{F} = C \begin{pmatrix} 0 & & & & \\ 0,33 & 0 & & & \\ 0,5 & 0,5 & 0 & & \\ 0,5 & 0,33 & 0,33 & 0 & \\ 0,33 & 0 & 0,5 & 0,33 & 0 \end{pmatrix}$$

Matriks F digunakan sebagai matriks jarak. Dimana dalam tahap awal penggabungan objek menjadi satu kelompok menggunakan nilai F terkecil sehingga objek B dan E berkelompok. Selanjutnya tahap penggabungan objek

berdasarkan metode numerik *ward*. Sampai diperoleh kelompok yang diinginkan.

2.6 Kinerja Pengelompokan

Pengukuran kinerja pengelompokan digunakan untuk menentukan jumlah kelompok optimum. hasil pengelompokan yang baik adalah antar anggota kelompok memiliki homogenitas yang tinggi dan antar kelompok memiliki heterogenitas yang tinggi (Hair, *et al*, 2010). Dimana pengukuran kinerja pengelompokan terbagi atas metode pengelompokan numerik dan kategorik seperti berikut ini.

2.6.1 Kinerja Pengelompokan Data Numerik

Indeks R^2 adalah salah satu indeks yang digunakan untuk tahap validasi pengelompokan (Sharma S., 1996). Dalam perhitungan indeks tersebut melibatkan perhitungan keragaman total, keragaman dalam kelompok dan keragaman antar kelompok berdasarkan beberapa persamaan berikut.

$$SST = \sum_{l=1}^{m_{numerik}} \sum_{i=1}^n (x_{il} - \bar{x}_l)^2 \quad SSW = \sum_{c=1}^C \sum_{l=1}^{m_{numerik}} \sum_{i=1}^{n_c} (x_{ilc} - \bar{x}_{lc})^2 \quad (2.20)$$

$$SSB = SST - SSW \quad (2.21)$$

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 ,

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

\bar{X}_{lc} : Rata-rata variabel ke- l pada kelompok ke- c ,

dimana $c = 1, 2, \dots, C$.

Nilai R^2 dihitung berdasarkan persamaan 2.22,

$$R^2 = \frac{SSB}{SST} = \frac{[SST - SSW]}{SST} \quad (2.22)$$

R^2 didefinisikan sebagai ukuran perbedaan antar kelompok dengan nilai berkisar antara 0 sampai 1. Nilai 0 menunjukkan bahwa tidak terdapat perbedaan antar kelompok, sedangkan nilai 1 menunjukkan adanya perbedaan yang signifikan antar kelompok yang terbentuk. Oleh karena itu, semakin besar nilai R^2 , mengindikasikan heterogenitas antar kelompok yang semakin tinggi. Setelah diperoleh nilai R^2 , dilakukan penentuan jumlah kelompok optimum dengan menggunakan *Pseudo-F*. Dimana nilai *Pseudo-F* diperoleh berdasarkan persamaan 2.23.

$$PseudoF = \frac{\left(\frac{R^2}{c-1}\right)}{\left(1 - \frac{R^2}{n-c}\right)} \quad (2.23)$$

Penentuan jumlah kelompok optimum menggunakan nilai maksimum dari *Pseudo-F*. sedangkan untuk menentukan metode terbaik menggunakan *icdrate* sesuai persamaan 2.24.

$$Icdrate = 1 - R^2 \quad (2.24)$$

Penentuan metode terbaik menggunakan *icdrate* terkecil. Ilustrasi perhitungan *pseudo-f* dan *icdrate* sebagai berikut.

1. Menghitung nilai SST sesuai dengan persamaan 2.20 yang disajikan pada Tabel 2.6.

Tabel 2.6 Perhitungan SST

Klaster	X_1	X_2	$(X_{i1} - \bar{X}_1)^2$	$(X_{i2} - \bar{X}_2)^2$
1	2.5	4	0.0441	0.0625
1	2	2	0.5041	4
3	4	3	1.6641	9
2	2	4	0.5041	16
2	1	8	2.9241	64
2	6	2	10.8241	4
1	1.5	3	1.4641	9
2	1.2	2.5	2.2801	6.25
1	2.9	4	0.0361	16
3	4	5	1.6641	25
\bar{X}_l	2.71	3.75	SST	175.22

Sehingga diperoleh SST sebesar 175.22. Selanjutnya menghitung SSW berdasarkan persamaan 2.20. sebelum dilakukan perhitungan, variabel dibagi sesuai dengan kelompok yang terbentuk. Hasil perhitungan disajikan pada Tabel 2.7.

Tabel 2.7 Perhitungan SSW

Klaster 1						
X₁	X₂	̄X₁₁	̄X₂₁	(X_{i11} - ̄X₁₁)²	(X_{i21} - ̄X₂₁)²	
2.5	4	2.23	3.25	0.08	0.56	
2	2			0.05	1.56	
1.5	3			0.53	0.06	
2.9	4			0.46	0.56	

Klaster 2						
X₁	X₂	̄X₁₁	̄X₂₁	(X_{i11} - ̄X₁₁)²	(X_{i21} - ̄X₂₁)²	
2	4	2.55	4.13	0.3	0.02	
1	8			2.4	15.02	
6	2			11.9	4.52	
1.2	2.5			1.82	2.64	

Klaster 3						
X₁	X₂	̄X₁₁	̄X₂₁	(X_{i11} - ̄X₁₁)²	(X_{i21} - ̄X₂₁)²	
4	3	4	4	0	1	
4	5			0	1	

SSW sebesar 44.475 dari penjumlahan kuadrat dari jarak objek terhadap rata-rata nilai variabel setiap kelompok.

2. Selanjutnya menghitung R^2 sesuai persamaan 2.22.

$$R^2 = \frac{(SST - SSW)}{SST} = \frac{(175.22 - 44.475)}{175.22} = 0.746$$

3. Selanjutnya melakukan perhitungan untuk mendapatkan *pseudo-f* dan *icdrate* sesuai persamaan 2.23 dan 2.24. Dimana pada ilustrasi ini jumlah kelompok yang terbentuk sebanyak 3 kelompok dengan 10 objek pengamatan.

$$pseudo-f = \frac{\left(\frac{R^2}{c-1} \right)}{\left(\frac{1-R^2}{n-c} \right)} = \frac{\left(\frac{0.746}{3-1} \right)}{\left(\frac{1-0.746}{10-3} \right)} = 10.279$$

$$ICD\ rate = 1 - R^2 = 1 - 0.746 = 0.254$$

2.6.2 Kinerja Pengelompokan Data Kategorik

Menentukan kelompok optimum pada metode pengelompokan untuk data kategorik menggunakan tingkat akurasi. Persamaan untuk menentukan tingkat akurasi adalah sebagai berikut (Huang & Ng, 1999).

$$r = \frac{1}{n} \sum_{l=1}^k a_l \quad (2.25)$$

dengan,

n : banyaknya pengamatan

a_l : banyaknya objek yang mendominasi pada kelompok ke- l

($l = 1, 2, \dots, k$)

Berikut akan disajikan ilustrasi perhitungan akurasi pada dua klaster hasil pengelompokan. Dimana data ilustrasi hasil pengelompokan disajikan pada Tabel 2.8.

Tabel 2.8 Data Ilustrasi Perhitungan Akurasi

Klaster 1			Klaster 2		
Objek	X ₁	X ₂	Objek	X ₁	X ₂
1	2	3	2	1	2
4	3	2	3	2	2
5	3	3	6	1	1
9	3	2	7	1	2
10	2	2	8	3	3

Klaster 1 dan 2 masing-masing terdiri atas 5 objek. Pada setiap klaster dan variabel ditentukan nilai yang paling sering muncul (modus). Misalnya nilai yang paling sering muncul pada klaster 1 dengan variabel X₁ adalah 3 dengan frekuensi sebanyak 3. Sedangkan pada klaster 1 variabel X₂ nilai yang paling sering muncul adalah 2 dengan frekuensi sebanyak 3. Untuk klaster 2 dengan variabel X₁ nilai yang paling sering muncul adalah 1 dengan frekuensi sebanyak 3. Nilai yang paling sering muncul pada klaster 2 dengan variabel X₂ adalah 2

dengan frekuensi sebanyak 3. Hasil akurasi yang disajikan pada Tabel 2.9.

Tabel 2.9 Hasil Ilustrasi Akurasi

Klaster	X ₁	X ₂
1	3	3
2	3	3
Jumlah	6	6
Rata-rata	6/10	6/10

Akurasi diperoleh berdasarkan nilai rata-rata pada Tabel 2.9 yang dirata-rata, sehingga diperoleh akurasi sebesar 0.6 atau 60%.

2.7 Uji Signifikansi Antar Klaster

Uji signifikansi dilakukan untuk mengetahui apakah terdapat perbedaan yang signifikan antar klaster yang terbentuk. Uji signifikansi pada penelitian ini menggunakan MANOVA, ANOVA dan uji *Kruskal-Wallis*. Pembahasan dimulai dengan MANOVA pada subbab 2.7.1.

2.7.1 Uji MANOVA

Setelah mendapatkan hasil pengelompokan yang optimum, berikutnya dilakukan uji perbedaan rata-rata dari kelompok yang diperoleh. Pada data multivariat pengujian rata-rata tersebut dilakukan dengan MANOVA (*Multivariate Analysis of Variance*). MANOVA merupakan perluasan dari teknik univariat *Analysis of Variance* (ANOVA) yang melibatkan lebih dari satu variabel (Johnson & Wichern, 2007). Rumusan hipotesis pada MANOVA adalah sebagai berikut,

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_g = \mu$$

$$H_1 : \text{minimal terdapat satu } \mu_i \neq \mu_j$$

Statistik uji yang digunakan dalam pengambilan keputusan dalam perbedaan antar kelompok adalah *wilk's lambda*. Nilai statistik uji *wilk's lambda* berkisar antara 0 sampai 1. Semakin rendah nilai statistik *wilk's lambda*, maka perbedaan antara kelompok semakin signifikan. Rumusan untuk statistik uji *wilk's lambda*, didefinisikan pada persamaan 2.26.

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} \quad (2.26)$$

dengan,

$$\mathbf{W} = \sum_{l=1}^g \sum_{j=1}^n (x_{lj} - \bar{x}_l) (x_{lj} - \bar{x}_l)^T \quad (2.27)$$

$$\mathbf{B} = \sum_{l=1}^g n_l (\bar{x}_l - \bar{x}) (\bar{x}_l - \bar{x})^T \quad (2.28)$$

Keterangan:

x_{lj} : objek pengamatan ke- j pada kelompok ke- l

\bar{x}_l : rata-rata seluruh objek pada kelompok ke- l

\bar{x} : rata-rata keseluruhan objek pengamatan

Jika asumsi tidak terpenuhi, digunakan statistik uji *Pillai's Trace* berikut ini.

$$V = \text{tr} \left[\mathbf{B} (\mathbf{W} + \mathbf{B})^{-1} \right] \quad (2.29)$$

Nilai statistik *Pillai's Trace* dapat diketahui dengan statistik uji F.

Dalam melakukan analisis dengan metode MANOVA, data yang dimiliki harus berdistribusi normal multivariat dan matriks varians kovarians homogen.

Uji homogenitas matriks varians kovarians dilakukan untuk mendeteksi dua atau lebih kelompok data sampel dari populasi memiliki matriks varians kovarians yang homogen. Pengujian yang dilakukan adalah dengan menggunakan *Box's M test*. Hipotesis yang digunakan dalam pengujian ini adalah sebagai berikut (Rencher, 2002).

H_0 : $\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$ (matriks varians kovarians bersifat homogen)

H_1 : minimal terdapat satu $\Sigma_i \neq \Sigma_j$ (matriks varians kovarians tidak homogen)

dimana g adalah banyaknya kelompok. Diperoleh keputusan tolak H_0 jika nilai $V^* > F_{\alpha;c-1;n-c}$ (Sharma, 1996). Statistik uji pada *Box's M test* adalah sebagai berikut.

$$c_1 = \frac{(k+1)(2p^2 + 3p - 1)}{6kv(p+1)}, \quad (2.30)$$

$$c_2 = \frac{(p-1)(p+2)}{6(k-1)} \left[\sum_{i=1}^k \frac{1}{v_i^2} - \frac{1}{\left(\sum_{i=1}^k v_i \right)^2} \right] \quad (2.31)$$

dan

$$a_1 = \frac{1}{2}(k-1)p(p+1) \quad a_2 = \frac{a_1 + 2}{|c_2 - c_1^2|} \quad (2.32)$$

$$b_1 = \frac{1-c_1-a_1/a_2}{a_1}, \quad b_2 = \frac{1-c_1-2/a_2}{a_2} \quad (2.33)$$

dengan kriteria yang digunakan,
jika $c_2 > c_1^2$, maka

$$F = -2b_1 \ln \mathbf{M} \quad (2.34)$$

jika $c_2 < c_1^2$, maka

$$F = \frac{2a_2 b_2 \ln \mathbf{M}}{a_1(1 + 2b_2 \ln \mathbf{M})} \quad (2.35)$$

dengan,

$$\mathbf{S}_{pl} = \frac{\sum_{i=1}^g v_i \mathbf{S}_i}{\sum_{i=1}^k v_i} \quad (2.36)$$

$$\ln \mathbf{M} = \frac{1}{2} \sum_{i=1}^g v_i \ln |\mathbf{S}_i| - \frac{1}{2} \left(\sum_{i=1}^g v_i \right) \ln |\mathbf{S}_i| \quad (2.37)$$

dimana n_i merupakan banyaknya data pada kelompok ke- i , dengan $i = 1, 2, \dots, g$, $v_i = n_i - 1$, \mathbf{S}_{pl} merupakan matriks varians kovarians kelompok gabungan, p adalah jumlah variabel independen, dan \mathbf{S}_i merupakan matriks varians kovarians kelompok ke- i . Persamaan untuk mendapatkan matriks \mathbf{S}_i adalah sebagai berikut.

$$\mathbf{S}_i = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}, \quad (2.38)$$

Keterangan :

$$v_i = n_i - 1$$

n_i : banyaknya data pada kelompok ke- i dengan $i=1,2,\dots,g$

g : banyaknya kelompok

\mathbf{S}_{pl} : matriks varians kovarians kelompok gabungan

p : jumlah variabel independen

\mathbf{S}_i : matriks varians kovarians kelompok ke- i

H_0 akan ditolak apabila $F > F_{\alpha(\alpha_1, \alpha_2)}$. Sedangkan untuk data kategorik dilakukan visualisasi analisis deskriptif untuk mengetahui perbedaan antar kelompok yang terbentuk. Setelah dilakukan uji signifikansi secara multivariat, dilakukan uji signifikansi secara univariat menggunakan *one-way* ANOVA.

One-Way ANOVA digunakan untuk mengetahui perbedaan antar klaster atau kelompok yang terbentuk. *One-Way* ANOVA adalah teknik untuk mengetahui perbedaan rata-rata pada 2 atau lebih k populasi dimana antara k populasi saling independent (Walpole, et al, 2007). Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : \mu_1 = \mu_2 = \dots = \mu_g$ (tidak ada perbedaan rata-rata dari k -populasi)

H_1 : paling sedikit ada 2 μ_i yang berbeda

Statistik uji yang digunakan dijelaskan pada Tabel 2.10. Untuk pengambilan keputusan, tolak H_0 jika $F_{hitung} > F_{(g-1,N-g)(\alpha)}$ yang artinya terdapat perbedaan rata-rata dari k populasi.

Tabel 2.10 Analysis of Variance

<i>Source of Variation</i>	<i>Sum of Square</i>	<i>Df</i>	<i>Mean Square</i>	<i>F</i>
<i>Treatment</i>	$SStr = \sum_{l=1}^g n_l (\bar{x}_l - \bar{x})^2$	$g - 1$	$\frac{SStr}{Df}$	$\frac{MStr}{MSE}$
<i>Residual</i>	$SSE = \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2$	$\sum_{l=1}^g n_l - g$	$\frac{SSE}{Df}$	
<i>Total</i>	$SST = \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x})^2$	$\sum_{l=1}^g n_l - 1$		

2.7.3 Uji Kruskal-Wallis

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

H_0 : Kedua populasi identik

H_1 : Tidak semua populasi identik

Statistik uji yang digunakan sesuai dengan persamaan 2.39.

$$H = \frac{12}{N(N+1)} \sum_{l=1}^g \frac{R_l^2}{n_l} - 3(N+1) \quad (2.39)$$

dengan,

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

n_i : banyaknya pengamatan di tiap kelompok

N : banyaknya seluruh pengamatan

K : banyaknya kelompok

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

2.8 Indikator Daerah Tertinggal

Menurut Peraturan Pemerintah No. 78 Tahun 2014, daerah tertinggal merupakan suatu daerah dengan Kabupaten yang masyarakat dan wilayahnya relatif kurang berkembang dibandingkan dengan daerah lain dalam skala nasional. Keberadaan daerah tertinggal merupakan indikator adanya kesenjangan dalam pembangunan. Suatu daerah ditetapkan sebagai daerah tertinggal berdasarkan beberapa kriteria antara lain perekonomian, sumber daya manusia, kemampuan keuangan daerah, infrastruktur/sarana prasarana, aksesibilitas dan karakteristik daerah. Berbagai permasalahan menyebabkan suatu Kabupaten/Kota masuk kategori daerah tertinggal adanya regulasi yang tidak memihak terhadap percepatan pembangunan, lemahnya kordinasi antar pelaku pembangunan, belum optimalnya kebijakan yang afirmatif, rendahnya kualitas sumber daya manusia dan tingkat kesejahteraan masyarakat, terbatasnya ketersediaan sarana dan prasarana publik dasar, rendahnya produktivitas masyarakat, belum optimalnya pengelolaan potensi sumber daya lokal dalam pengembangan ekonomi, kurangnya aksesibilitas daerah terhadap pusat pertumbuhan dan belum adanya insentif terhadap sektor swasta dan pelaku usaha untuk berinvestasi.

Jumlah daerah tertinggal sejak tahun 2005 sampai sekarang mengalami penurunan, pada periode 2005-2009 terdapat daerah tertinggal sebanyak 199 daerah. Berkurang pada periode 2010-2014 menjadi 183 daerah tertinggal. Di Provinsi Jawa Timur memiliki jumlah daerah tertinggal paling banyak diantara semua Provinsi di pulau Jawa. Pada periode 2010-2014 pemerintah gagal memenuhi target dalam pengurangan jumlah daerah tertinggal di Jawa Timur. Berdasarkan Peraturan Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigrasi No. 3 Tahun 2016, untuk identifikasi status ketertinggalan suatu Kabupaten/Kota digunakan 6 kriteria yaitu perekonomian, sumber daya manusia (SDM), kemampuan keuangan daerah, infrastruktur sarana/prasarana, aksesibilitas, dan karakteristik

daerah. Penentuan status ketertinggalan daerah menggunakan indeks komposit dengan perhitungan seperti pada persamaan 2.40.

$$IK = \sum_{j=1}^n z_j a_j \quad (2.40)$$

dengan,

IK : indeks komposit

z_j : nilai indikator ke- j yang telah distandarisasi

a_j : arah indikator (+ atau -) ke- j

Penelitian yang dilakukan Kementerian DPDT dalam pengentasan ketertinggalan kabupaten/kota dilakukan dengan intervensi terhadap kriteria/indikator yang menjadi penyebab utama kabupaten tersebut termasuk daerah tertinggal dengan beberapa tahap yaitu menentukan batas maksimum indeks komposit kabupaten potensi maju, menentukan parameter masing-masing kriteria/indikator dan menentukan kriteria/indikator yang perlu diintervensi.

Batas Maksimum Indeks Komposit Kabupaten Potensi Maju (BMIKKPM) adalah 0,085666. Kemudian selanjutnya menentukan batas maksimum indeks kriteria (BMIK) dan indikator (BMII) sesuai persamaan 2.41.

$$BMIK_i = \frac{\bar{a}_i}{BMIKKPM}, \quad BMII_j = \frac{\bar{e}_j}{BMIKKPM} \quad (2.41)$$

dimana \bar{a}_i merupakan rata-rata indeks kriteria ke- i dan \bar{e}_j adalah nilai rata-rata indeks indikator ke- j . Identifikasi kriteria/indikator penyebab ketertinggalan dilakukan dengan membandingkan indeks masing-masing kriteria/indikator dengan nilai BMIK/BMII dengan ketentuan:

- Jika nilai indeks kriteria ke- i (a_i) $>$ BMIK (\hat{a}_i) maka kriteria tersebut perlu diintervensi
- Jika nilai indeks kriteria ke- i (a_i) \leq BMIK (\hat{a}_i) maka kriteria tidak perlu diintervensi
- Jika nilai indeks kriteria ke- i (e_j) $>$ BMII (\hat{e}_j) maka indikator tersebut perlu diintervensi

(d) Jika nilai indeks kriteria ke- i (e_j) \leq BMII (\hat{e}_j) maka indikator tidak perlu diintervensi

a. Kriteria Infrastruktur

Kriteria infrastruktur mewakili kondisi infrastruktur dan fasilitas pendidikan serta kesehatan yang ada di daerah. Indikator yang digunakan pada penelitian ini adalah presentase banyaknya desa dengan jalan terluas aspal, rasio rumah sakit per 1000 penduduk, rasio jumlah SMA sederajat terhadap penduduk usia SMA, presentase rumah tangga pengguna listrik PLN dan sumber air minum utama. Sumber air minum utama terbagi atas 6 kategori yaitu air hujan, danau/sungai/kolam, mata air, sumur, ledeng tanpa meteran dan air kemasan.

b. Kriteria Karakteristik Daerah

Karakteristik daerah diwakili oleh beberapa indikator yaitu presentase lahan kritis dan keberadaan Kawasan industri di setiap Kabupaten/Kota.

c. Kriteria Perekonomian

Kriteria perekonomian berhubungan dengan kondisi perekonomian penduduk di suatu daerah. Indikator yang mewakili perekonomian suatu Kabupaten/Kota dalam pemetaan daerah adalah presentase penduduk miskin dan pengeluaran per kapita penduduk. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan dibawah garis kemiskinan. Dimana untuk menghitung presentase penduduk miskin digunakan persamaan berikut.

$$P_\alpha = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^\alpha \quad (2.42)$$

dengan,

$\alpha = 0$

$z =$ garis kemiskinan

$y_i =$ rata-rata pengeluaran per kapita sebulan penduduk yang berada di bawah garis kemiskinan

$q =$ banyaknya penduduk yang berada di bawah garis kemiskinan

$n =$ jumlah penduduk

Pengeluaran rata-rata per kapita adalah biaya yang dikeluarkan untuk konsumsi anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga dalam rumah tangga tersebut.

d. Kriteria Sumber Daya Manusia

Angka harapan hidup, angka melek huruf dan presentase pengangguran mewakili kriteria sumber daya manusia dalam indikator daerah tertinggal. Angka harapan hidup merupakan rata-rata tahun hidup yang masih akan dijalani oleh seseorang yang telah berhasil mencapai umur x pada suatu tahun tertentu. Angka Harapan Hidup merupakan alat untuk mengevaluasi kinerja pemerintah dalam meningkatkan kesejahteraan penduduk dan derajat kesehatan. Sedangkan angka melek huruf adalah proporsi penduduk usia 15 tahun ke atas yang mempunyai kemampuan membaca dan menulis huruf latin dan huruf lainnya, tanpa harus mengerti apa yang dibaca/ditulisnya terhadap penduduk usia 15 tahun ke atas. Selain itu, terdapat presentase pengangguran terhadap angkatan kerja.

e. Kriteria Kemampuan Keuangan Daerah

Kemampuan keuangan daerah diwakili oleh tingkat desentralisasi fiskal. Tingkat desentralisasi fiskal dihitung berdasarkan perbandingan antara pendapatan asli daerah (PAD) dengan total penerimaan daerah. Rasio ini menunjukkan kontribusi PAD terhadap total penerimaan daerah. Menghitung rasio desentralisasi fiskal sesuai dengan persamaan 2.43.

$$\text{Derajat Desentralisasi} = \frac{\text{Pendapatan Asli Daerah}}{\text{Total Pendapatan Daerah}} \times 100\% \quad (2.43)$$

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Sumber data yang digunakan pada penelitian ini adalah data sekunder yang didapatkan dari Badan Pusat Statistik Provinsi Jawa Timur yang dapat diakses di situs jatim.bps.go.id. Data yang digunakan adalah data Potensi Desa (PODES) dan Survei Sosial dan Ekonomi Nasional (SUSENAS) tahun 2014 untuk seluruh Kabupaten/Kota di Jawa Timur. Data yang digunakan adalah data dengan skala numerik dan kategorik.

Terdapat tingkatan pada data potensi desa yaitu desa, kecamatan dan Kabupaten. Penelitian ini menggunakan data potensi desa tingkat Kabupaten/Kota. Periode survei untuk data PODES dilakukan 3 kali dalam kurun waktu 10 tahun. Survei Sosial Ekonomi Nasional merupakan kegiatan survei mengenai bidang kependudukan, kesehatan, pendidikan, keluarga berencana, perumahan, konsumsi dan pengeluaran yang dilakukan tiap tahun. Dimana pengumpulan data setiap tahunnya terbagi atas 4 triwulan. Unit pengamatan pada penelitian ini adalah 38 Kabupaten/Kota di Jawa Timur

3.2 Variabel Penelitian

Variabel penelitian yang digunakan pada penelitian ini adalah variabel indikator penetapan daerah tertinggal di Jawa Timur yang berisi data campuran yaitu numerik dan kategorik. Variabel penelitian disajikan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Data Numerik		
Kriteria	Variabel	Deskripsi
Infrastruktur	X_1	Presentase Desa dengan Permukaan Jalan Terluas Aspal
	X_2	Rasio Jumlah Rumah Sakit per 1000 Penduduk
	X_3	Rasio SMA Terhadap Penduduk Usia SMA

Tabel 3.1 Variabel Penelitian (Lanjutan)

Variabel Numerik		
Kriteria	Variabel	Deskripsi
Infrastruktur	X_4	Presentase Rumah Tangga Pengguna Listrik PLN
Karakteristik Daerah	X_5	Presentase Luas Lahan Kritis
Ekonomi	X_6	Presentase Penduduk Miskin
	X_7	Pengeluaran per Kapita Penduduk
Sumber Daya Manusia	X_8	Angka Harapan Hidup
	X_9	Angka Melek Huruf
	X_{10}	Presentase Pengangguran
Variabel Kategorik		
Kriteria	Variabel	Keterangan
Karakteristik Daerah	X_{11}	Keberadaan Kawasan Industri
Kemampuan Keuangan Daerah	X_{12}	Tingkat Desentralisasi Fiskal
Infrastruktur	X_{13}	Sumber Air Minum Utama

Struktur data dari variabel-variabel yang telah tercantum pada Tabel 3.1 disajikan pada Tabel 3.2.

Tabel 3.2 Struktur Data

Kabupaten/Kota ke-	Numerik						Kategorik
	X_1	X_2	...	X_{10}	X_{11}	...	X_{13}
1	$x_{1,1}$	$x_{2,1}$...	$x_{10,1}$	$x_{11,1}$		$x_{13,1}$
2	$x_{1,2}$	$x_{2,2}$...	$x_{10,2}$	$x_{11,2}$		$x_{13,2}$
3	$x_{1,3}$	$x_{2,3}$...	$x_{10,3}$	$x_{11,3}$		$x_{13,3}$
:	:	:	...	:	:		:
38	$x_{1,38}$	$x_{2,38}$...	$x_{10,38}$	$x_{11,38}$		$x_{13,38}$

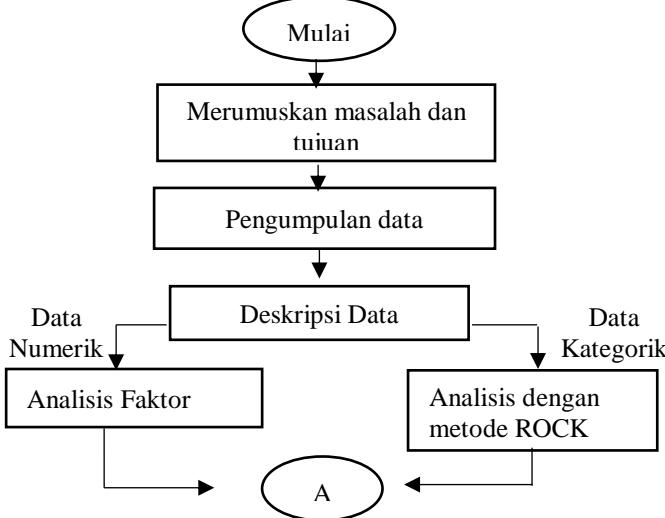
3.3 Langkah Analisis

Berikut adalah langkah-langkah analisis yang akan dilakukan pada penelitian berdasarkan tujuan penelitian yang telah dirumuskan.

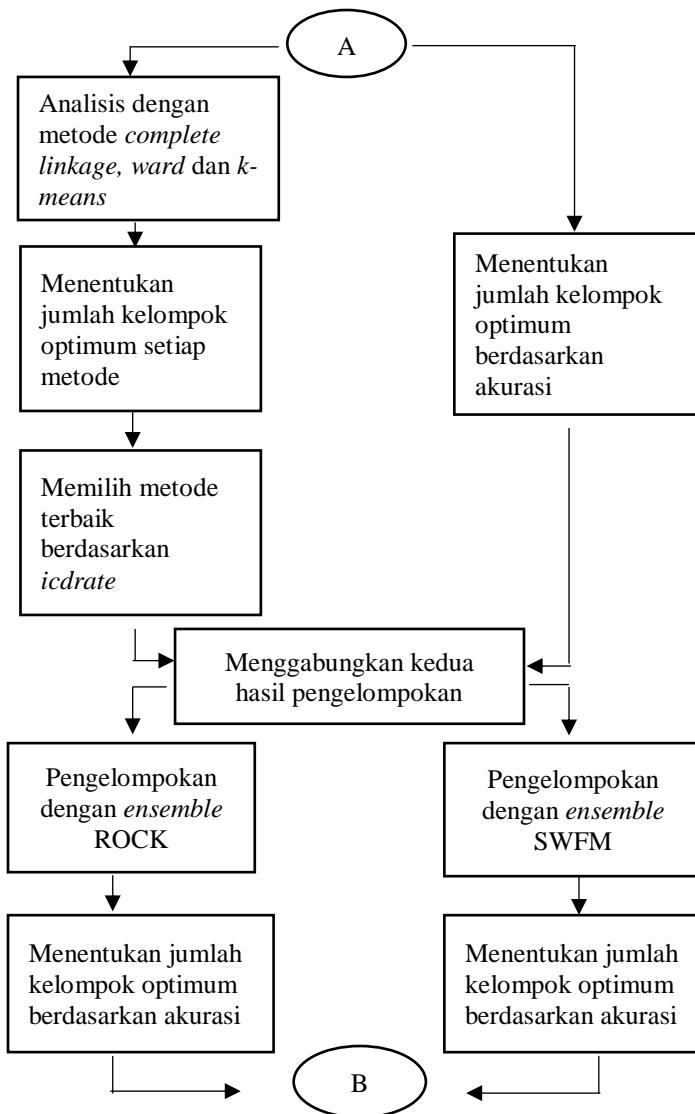
1. Mendeskripsikan data menggunakan analisis statistika deskriptif terhadap data indikator daerah tertinggal berdasarkan tipe data numerik dan kategorik yang meliputi *mean*, modus, dan varians sesuai dengan subbab 2.1.
2. Melakukan pengelompokan data campuran menggunakan metode *ensemble* ROCK. Adapun tahap analisis adalah sebagai berikut.
 - (a) Memisahkan antara data numerik dan kategorik.
 - (b) Melakukan analisis faktor untuk data numerik beserta pengujian asumsi yang diperlukan. Dipilih variabel dengan *loading factor* terbesar yang akan dianalisis pada tahap pengelompokan.
 - (c) Melakukan analisis pengelompokan data numerik menggunakan metode *ward*, *complete linkage*, dan *k-means* dengan merujuk pada persamaan 2.13, 2.14 dan 2.15 serta menentukan jumlah kelompok optimum dari setiap metode menggunakan nilai *pseudo-f* sesuai persamaan 2.23. memilih metode

- terbaik berdasarkan *icdrate* terkecil yang merujuk persamaan 2.24.
- (d) Melakukan analisis pengelompokan menggunakan metode ROCK untuk data kategorik.
- i. Menentukan inisialisasi untuk masing-masing data poin sebagai *cluster* awal.
 - ii. Menghitung similaritas. Ukuran kemiripan antara pasangan pengamatan ke-*i* dan ke-*j* sesuai persamaan 2.16.
 - iii. Menentukan Tetangga (*Neighbors*)
 - iv. Menghitung *Link*
 - v. Menghitung *Goodness Measure* sesuai persamaan 2.17.
 - vi. Membandingkan hasil untuk masing-masing *threshold* (θ) yang telah ditentukan dan menentukan jumlah kelompok optimum berdasarkan nilai akurasi tertinggi sesuai dengan persamaan 2.25.
- (e) Menggabungkan hasil pengelompokan untuk data numerik dan kategorik.
- (f) Melakukan penggabungan hasil pengelompokan menggunakan *ensemble* ROCK sesuai dengan langkah nomor 2D dalam melakukan pengelompokan menggunakan metode ROCK untuk data kategorik.
3. Melakukan penggabungan hasil pengelompokan data campuran menggunakan SWFM. Adapun tahap analisis adalah sebagai berikut.
- (a) Melakukan analisis menggunakan *similarity weight method* untuk menentukan nilai $sim(X_i, X_j)$ sesuai dengan persamaan 2.18.

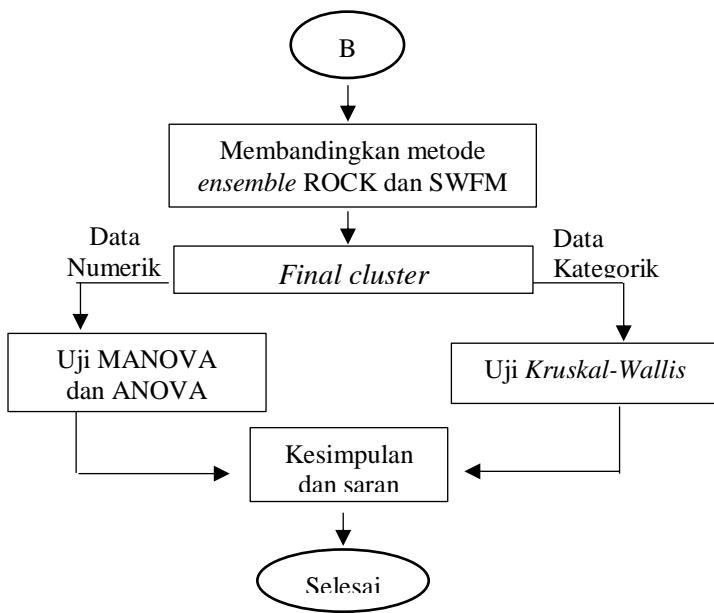
- (b) Melakukan analisis menggunakan *filter algorithm* untuk memperoleh nilai $F(X_i, X_j)$ pada data numerik dan kategorik sesuai dengan persamaan 2.19.
- (c) Menentukan jumlah kelompok optimum menggunakan nilai akurasi tertinggi berdasarkan persamaan 2.25.
4. Membandingkan hasil pengelompokan dari metode *ensemble* ROCK dan SWFM untuk menentukan metode terbaik dengan menggunakan nilai akurasi sesuai persamaan 2.24. Selanjutnya dilakukan uji MANOVA pada data numerik untuk mengetahui apakah terdapat perbedaan kelompok yang dihasilkan sesuai dengan subbab 2.7.1 Sedangkan data kategorik dilakukan pengujian Kruskal-Wallis sesuai dengan persamaan 2.38. Langkah analisis sebagaimana yang telah dijelaskan diatas dapat disajikan dalam diagram alir seperti berikut ini.



Gambar 3.1 Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian (Lanjutan)



Gambar 3.1 Diagram Alir Penelitian (Lanjutan)

(Halaman ini sengaja dikosongkan)

BAB IV **ANALISIS DAN PEMBAHASAN**

Analisis dan pembahasan yang diuraikan pada bab ini mencakup hasil pemetaan Kabupaten/Kota di Jawa Timur yaitu karakteristik Kabupaten/Kota di Jawa Timur, pengelompokan Kabupaten/Kota di Jawa Timur menggunakan metode *ensemble* ROCK dan SWFM dan perbandingan hasil pengelompokan antara *ensemble* ROCK dan SWFM. Analisis akan dimulai dengan analisis deskriptif untuk mengetahui karakteristik indikator daerah tertinggal Kabupaten/Kota di Jawa Timur.

4.1 Karakteristik Kabupaten/Kota di Jawa Timur

Terdapat 38 Kabupaten/Kota di Jawa Timur. Dalam penetapan daerah tertinggal digunakan indikator daerah tertinggal yang mengandung data campuran numerik dan kategorik. Dimana terdapat 5 kriteria yang digunakan sebagai indikator daerah tertinggal yaitu ekonomi, sumber daya manusia, infrastruktur, kapasitas keuangan daerah dan karakteristik daerah. Karakteristik data numerik indikator daerah tertinggal Kabupaten/Kota di Jawa Timur disajikan pada Tabel 4.1 dan Gambar 4.1

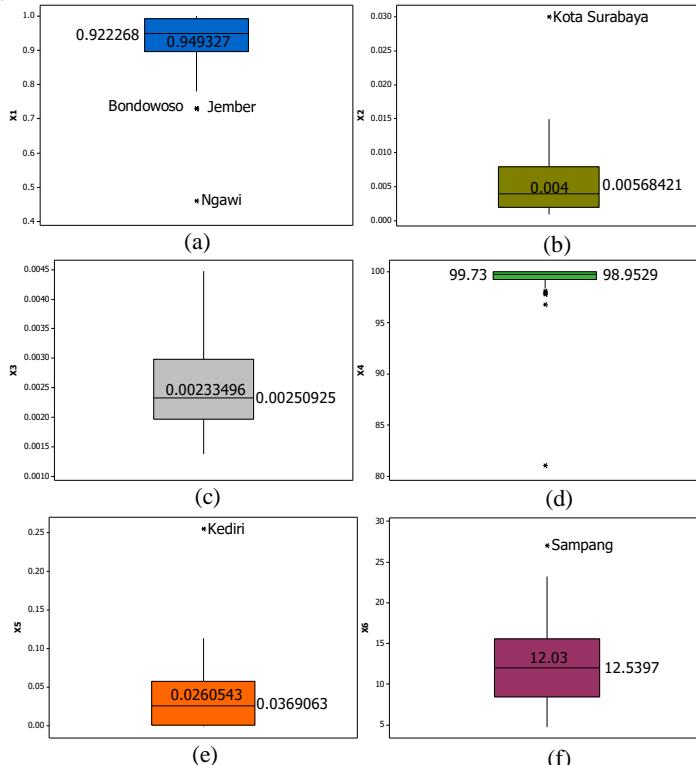
Tabel 4.1 Karakteristik Data Numerik Kabupaten/Kota di Jawa Timur

Variabel	Median	Mean	Min	Max
Presentase Desa Dengan Jalan Terluas Aspal	0.949	0.922	0.461	1
Rumah Sakit per 1000 penduduk	0.004	0.005	0.001	0.03
Rasio SMA terhadap penduduk usia SMA	0.002	0.002	0.001	0.004
Presentase Pengguna Listrik PLN	99.73	98.95	81.04	100
Presentase Luas Lahan Kritis	0.026	0.037	0	0.255
Presentase Penduduk Miskin	12.03	12.54	4.77	27.08

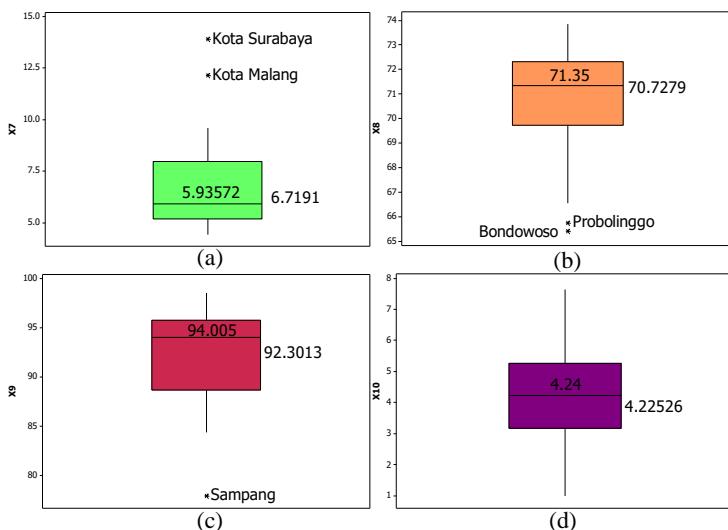
Tabel 4.1 Karakteristik Data Numerik Kabupaten/Kota di Jawa Timur (Lanjutan)

Variabel	Median	Mean	Min	Max
Pengeluaran per Kapita Penduduk	5,936	6.719	4.445	13.89
Angka Harapan Hidup	71,35	70.728	65.43	73.85
Angka Melek Huruf	94,005	92.301	77.93	98.52
Presentase Pengangguran	4,24	4.225	1.01	7.66

Untuk visualisasi analisis deskriptif dapat dilihat pada Gambar 4.1



Gambar 4.1 Boxplot (a)Presentase Jalan Aspal (b)Rasio SMA per 1000 penduduk (c)Rasio SMA per Penduduk Usia SMA (d)Presentase Pengguna Listrik PLN (e)Prsentase Lahan Kritis (f)Presentase Penduduk Miskin



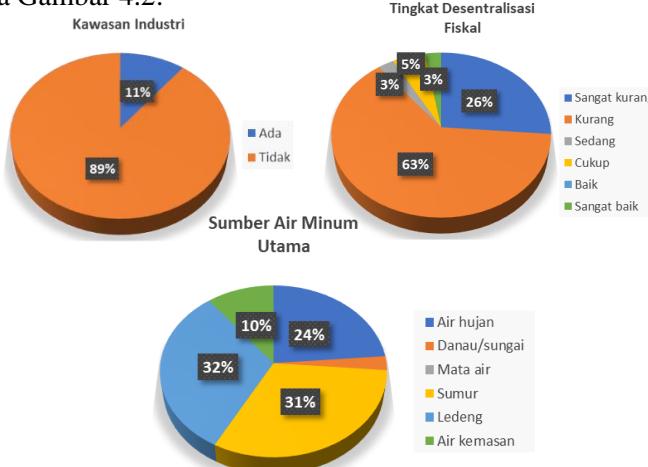
Gambar 4.1 Boxplot (a) Pengeluaran per Kapita Penduduk (b) Angka Harapan Hidup (c) Angka Melek Huruf (d) Presentase Pengangguran

Berdasarkan Tabel 4.1 dan Gambar 4.1, rata-rata presentase jumlah wilayah dengan jenis jalan aspal di Jawa Timur cukup tinggi yaitu mencapai 92.2%. Dari kriteria kesehatan dan Pendidikan, rata-rata rasio banyaknya rumah sakit per 1000 penduduk yaitu 0.005 yang artinya lima rumah sakit menampung seribu penduduk dan rata-rata rasio banyaknya SMA sederajat terhadap penduduk usia SMA sederajat yaitu 0,002. Rata-rata presentase rumah tangga pengguna listrik PLN di Jawa Timur mencapai 98.95% dimana terdapat 10 Kabupaten/Kota yang keseluruhan rumah tangga menggunakan listrik PLN. Untuk rata-rata presentase lahan kritis di Jawa Timur cukup kecil yaitu 3.7%. Terdapat 9 Kabupaten/Kota yang tidak memiliki lahan kritis.

Dari kriteria ekonomi, rata-rata presentase penduduk miskin di Jawa Timur mencapai 12.54% dengan Kabupaten Sampang memiliki jumlah penduduk miskin terbesar dengan presentase mencapai 27.08%. Hal ini menunjukkan ketimpangan karena presentase penduduk miskin terkecil hanya

4.77% yaitu di Kota Batu. Rata-rata pengeluaran per kapita penduduk Jawa Timur mencapai Rp. 671.900,00. Berdasarkan kriteria SDM, rata-rata Angka Harapan Hidup dan Angka Melek Huruf di Jawa Timur masing-masing mencapai 70.728 dan 92.301. Sedangkan rata-rata presentase pengangguran di Jawa Timur mencapai 4.225%. Kota Kediri memiliki jumlah pengangguran tertinggi dengan presentase mencapai 7.66%.

Selain data numerik, dilakukan analisis deskriptif untuk mengetahui karakteristik data kategorik indikator daerah tertinggal Kabupaten/Kota di Jawa Timur. Dimana data kategorik diwakili oleh tiga variabel yaitu keberadaan kawasan industri, tingkat desentralisasi fiskal, dan sumber air minum utama. Dimana karakteristik data kategorik divisualisasikan pada Gambar 4.2.



Gambar 4.2 Pie Chart Data Kategorik

Selain visualisasi dengan gambar, karakteristik data kategorik juga disajika pada Tabel 4.2.

Tabel 4.2 Karakteristik Data Kategorik

Variabel	Kategori	Jumlah	Persentase
Kawasan Industri	1	4	11%
	2	34	89%

Tabel 4.2 Karakteristik Data Kategorik (Lanjutan)

Variabel	Kategori	Jumlah	Percentase
Tingkat Desentralisasi Fiskal	1	10	26%
	2	24	63%
	3	1	3%
	4	2	5%
	5	0	0%
	6	1	3%
Sumber Air Minum	1	9	24%
	2	1	3%
	3	0	0%
	4	12	31%
	5	12	31%
	6	4	11%

Gambar 4.2 dan Tabel 4.2 menunjukkan karakteristik data kategorik indikator daerah tertinggal Kabupaten/Kota di Jawa Timur. Sebagian besar Kabupaten/Kota di Jawa Timur tidak memiliki Kawasan industri, hal ini ditunjukkan dengan presentase Kabupaten/Kota di Jawa Timur yang memiliki kawasan industri (1) cukup kecil yaitu sebesar 11% atau sebanyak 4 Kabupaten/Kota. Berdasarkan kriteria kapasitas kemampuan keuangan daerah yang diukur menggunakan tingkat desentralisasi fiskal, sebagian besar Kabupaten/Kota di Jawa Timur dengan presentase 63% masuk dalam kategori kurang (2). Hanya satu Kabupaten/Kota yang memiliki tingkat desentralisasi fiskal yang sangat baik (4) yaitu Surabaya. Sedangkan sumber air minum utama yang paling banyak digunakan di Kabupaten/Kota di Jawa Timur adalah bersumber dari sumur (4) dan ledeng tanpa meteran/bor/pompa (5) dengan presentase sebesar 31%.

4.2 Analisis Faktor Indikator Daerah Tertinggal Kabupaten/Kota di Jawa Timur

Sebelum dilakukan analisis klaster, terlebih dahulu akan dilakukan analisis faktor pada data numerik. Analisis faktor digunakan untuk mengurangi dampak multikolinearitas dengan memilih variabel yang memiliki kontribusi terbesar pada setiap

faktor yang terbentuk. Untuk mengetahui kelayakan analisis faktor, dilakukan pemeriksaan kecukupan korelasi antar variabel, pemeriksaan dependensi antar variabel, dan pemeriksaan distribusi normal multivariat.

b. Identifikasi Kecukupan Korelasi Antar Variabel

Pemeriksaan kecukupan korelasi antar variabel menggunakan nilai *Kaiser Meyer-Olkin* (KMO). Dimana data dikatakan memenuhi kecukupan korelasi antar variabel jika nilai KMO lebih besar dari 0.5. Hasil identifikasi kecukupan korelasi antar variabel menunjukkan nilai KMO sebesar 0.705. Hal tersebut menunjukkan bahwa data telah memenuhi kecukupan korelasi antar variabel.

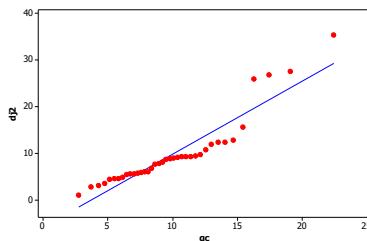
c. Uji Dependensi Antar Variabel

Analisis faktor layak dilakukan ketika terdapat dependensi antar variabel. Hal ini dapat diketahui dengan uji dependensi/korelasi antar variabel dengan menggunakan uji *Bartlett*. Adapun hipotesis untuk uji *Bartlett* sebagai berikut.

$$H_0 : \rho = I \text{ (tidak terdapat korelasi antar variabel)}$$

$$H_1 : \rho \neq I \text{ (terdapat korelasi antar variabel)}$$

Uji *Bartlett* menghasilkan nilai χ^2_{hitung} sebesar 173.305 dengan *p-value* 0.000 lebih kecil dari $\alpha(0.05)$ sehingga diperoleh kesimpulan bahwa terdapat korelasi antar variabel. Pada uji dependensi diperlukan pemeriksaan asumsi distribusi normal multivariat menggunakan uji *sapiro wilks* diperoleh *p-value* sebesar 1.536×10^{-12} yang berarti data tidak berdistribusi normal multivariat. Salah satu alasan data tidak berdistribusi normal multivariat karena terindikasi mengandung *outlier* secara multivariat. Sehingga dilakukan pemeriksaan *outlier* secara multivariat menggunakan *Qqplot* yang disajikan pada Gambar 4.3.



Gambar 4.3 *Qq Plot* Data Numerik

Secara visual, pada Gambar 4.3 menunjukkan bahwa terdapat data yang berada di luar garis normal yang mengindikasikan terdapat banyak data *outlier* secara multivariat. Hal tersebut yang menyebabkan data tidak berdistribusi normal multivariat.

Statistik uji yang digunakan untuk uji *Bartlett* menggunakan *chi-square* yang merupakan derivatif dari distribusi normal, sehingga menyebabkan hasil pengujian tidak valid karena data tidak berdistribusi normal multivariat. Oleh karena itu, pemeriksaan kelayakan analisis faktor pada penelitian ini hanya didasarkan pada pendekatan deskriptif menggunakan nilai KMO. Dimana hasil pemeriksaan menunjukkan bahwa data memenuhi kecukupan korelasi antar variabel. Sehingga analisis faktor masih bisa dilanjutkan. Hasil analisis faktor dapat dilihat pada Tabel 4.3.

Tabel 4.3 Proporsi Keragaman

<i>Component</i>	<i>Eigenvalues</i>		
	Total	Proporsi Varians (%)	Kumulatif Varians (%)
1	4.542	45.420	45.420
2	1.240	12.397	57.817
3	1.079	10.794	68.611
4	0.767	7.673	76.284
5	0.711	7.107	83.392
6	0.628	6.28	89.671
7	0.436	4.364	94.035

Tabel 4.3 Proporsi Keragaman (Lanjutan)

<i>Component</i>	<i>Eigenvalues</i>		
	Total	Proporsi Varians (%)	Kumulatif Varians (%)
8	0.291	2.908	96.943
9	0.216	2.165	99.108
10	0.089	0.892	100

Penentuan jumlah komponen yang terbentuk berdasarkan *eigenvalue* lebih besar dari satu atau proporsi varians kumulatif di atas 60% (Hair, *et al*, 2010). Sehingga berdasarkan Tabel 4.3, dapat disimpulkan jumlah komponen yang terbentuk sebanyak tiga komponen yang saling independen. Meskipun hasil uji *Bartlett* tidak valid, namun hasil analisis faktor masih bisa ditoleransi karena jumlah faktor yang terbentuk lebih kecil dari jumlah variabel yang digunakan. Hal ini menunjukkan terdapat korelasi yang tinggi antar anggota dalam satu faktor.

Proporsi keragaman yang dihasilkan sebesar 68.611% yang artinya ketiga komponen yang terbentuk mampu menjelaskan variabilitas dari seluruh variabel sebesar 68.611%. Sehingga untuk analisis klaster akan digunakan tiga variabel yang mewakili 3 komponen tersebut. Untuk mengetahui variabel yang akan digunakan untuk analisis klaster dapat dilihat berdasarkan nilai *loading factor* pada Tabel 4.4.

Tabel 4.4 Loading Factor

No	Variabel	Faktor		
		1	2	3
1	Presentase Jalan Aspal	0.65	0.104	0.527
2	Rasio Rumah Sakit per 1000 penduduk	0.679	-0.095	-0.163
3	Rasio SMA sederajat terhadap penduduk usia SMA	-0.149	-0.857*	-0.071
4	Presentase Pengguna Listrik PLN	0.089	0.764	-0.065

Tabel 4.4 Loading Factor (Lanjutan)

No	Variabel	Faktor		
		1	2	3
5	Presentase Lahan Kritis	-0.248	-0.047	0.805*
6	Presentase Penduduk Miskin	-0.755	-0.436	0.054
7	Pengeluaran per Kapita Penduduk	0.852*	0.106	-0.284
8	Angka Harapan Hidup	0.725	0.354	-0.031
9	Angka Melek Huruf	0.78	0.414	-0.122
10	Presentase Pengangguran	0.437	0.473	-0.376

Tabel 4.4 menunjukkan nilai *loading factor*. Dimana faktor 1 terdiri atas enam indikator yaitu presentase jalan aspal, rasio rumah sakit per 1000 penduduk, presentase penduduk miskin, pengeluaran penduduk per kapita, angka harapan hidup, dan angka melek huruf. Faktor 2 terdiri atas 3 indikator yaitu rasio SMA sederajat per penduduk usia SMA, presentase pengguna listrik PLN dan presentase pengangguran. Sisanya satu indikator yaitu presentase lahan kritis termasuk dalam faktor 3. Dari ketiga faktor yang terbentuk akan dipilih variabel yang memiliki kontribusi/bobot faktor terbesar dari setiap faktor yang terbentuk. Variabel yang terpilih untuk masing-masing faktor adalah pengeluaran per kapita penduduk, rasio banyaknya SMA sederajat terhadap penduduk usia SMA sederajat, dan presentase lahan kritis. Ketiga variabel numerik tersebut akan digunakan untuk analisis klaster menggunakan metode *ward*. Dari ketiga faktor yang terbentuk memiliki korelasi yang kecil antar faktor. Namun, antar anggota memiliki korelasi yang besar. Hal ini sesuai dengan matriks korelasi yang terbentuk berikut ini.

$$X_1 \begin{pmatrix} 1 \\ X_2 & 0.21 & 1 \\ X_3 & -0.2 & \mathbf{-0.84} & 1 \\ X_4 & 0.182 & 0.17 & -0.48 & 1 \\ X_5 & 0.036 & -0.15 & 0.068 & -0.13 & 1 \\ X_6 & -0.44 & -0.36 & 0.433 & -0.34 & 0.225 & 1 \\ X_7 & 0.442 & \mathbf{0.578} & -0.2 & 0.232 & -0.41 & \mathbf{-0.69} & 1 \\ X_8 & 0.385 & 0.432 & -0.49 & 0.208 & -0.24 & \mathbf{-0.62} & \mathbf{0.573} & 1 \\ X_9 & 0.435 & 0.399 & -0.39 & 0.345 & -0.32 & \mathbf{-0.84} & \mathbf{0.641} & \mathbf{0.738} & 1 \\ X_{10} & 0.120 & 0.281 & -0.38 & 0.31 & -0.23 & \mathbf{-0.52} & \mathbf{0.562} & \mathbf{0.401} & \mathbf{0.545} & 1 \end{pmatrix}$$

Berdasarkan matriks korelasi di atas, terdapat sebelas nilai korelasi lebih dari 0.5.

4.2 Pengelompokan Data Numerik Indikator Daerah Tertinggal

Pengelompokan data numerik dilakukan dengan menggunakan metode hierarki *complete linkage*, *ward* dan *non hierarki k-means* dengan jarak *euclidean*. Sebelum dilakukan perbandingan antar metode, dilakukan pemilihan klaster optimum dari masing-masing metode menggunakan nilai *Pseudo-F* tertinggi. Kemudian dilakukan pemilihan metode terbaik menggunakan *icdrate* terkecil dari klaster optimum masing-masing metode. Hasil pengelompokan data numerik dapat dilihat pada Tabel 4.5.

Tabel 4.5 Hasil Pengelompokan Data Numerik

Metode	Jumlah Klaster	Pseudo-F	Icdrate
<i>Complete linkage</i>	2	35.907	0.5
	3	117.256	0.13
	4	166.833	0.064
	5	142.669	0.055
<i>Ward's*</i>	2	83.33	0.302
	3	150.354	0.104
	4	162.394	0.065
	5	163.408	0.048
<i>K-means</i>	2	83.33	0.302
	3	150.354	0.104
	4	166.83	0.064
	5	142.669	0.055

Pengelompokan data numerik menggunakan *complete linkage*, *ward* dan *k-means* pada Tabel 4.5 menghasilkan kelompok optimum berturut-turut empat, lima dan empat kelompok optimum dari setiap metode. Dari jumlah kelompok optimum yang terpilih ditentukan metode terbaik, dimana metode *ward* adalah metode terbaik karena memiliki *icdrate* terkecil yaitu 0.048. Selanjutnya untuk mengetahui perbandingan analisis klaster antara menggunakan analisis faktor dan tidak, dapat dilihat pada Tabel 4.6.

Tabel 4.6 Perbandingan Analisis Klaster Data Numerik

Metode	Keterangan	Icdrate
<i>Complete linkage</i>	Analisis Faktor	0.064
	Tanpa Analisis Faktor	0.5
<i>Ward</i>	Analisis Faktor	0.048
	Tanpa Analisis Faktor	0.55
<i>K-means</i>	Analisis Faktor	0.064
	Tanpa Analisis Faktor	0.488

Berdasarkan Tabel 4.6, analisis klaster dengan menggunakan variabel hasil analisis faktor lebih baik karena memiliki *icdrate* yang lebih kecil.

4.3 Pengelompokan Data Kategorik Indikator Daerah Tertinggal

Pengelompokan data kategorik indikator daerah tertinggal dilakukan menggunakan metode ROCK. Nilai *threshold* (θ) yang digunakan adalah 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, dan 0.5. Dari kesepuluh nilai *threshold* tersebut akan dipilih klaster yang menghasilkan kelompok optimum dengan kriteria akurasi tertinggi. Semakin besar akurasi, maka kinerja pengelompokan semakin baik. Hal ini menunjukkan bahwa terdapat homogenitas maksimum dalam kelompok dan heterogenitas maksimum antar kelompok. Hasil analisis klaster untuk data kategorik menggunakan metode ROCK dapat dilihat pada Tabel 4.7.

Tabel 4.7 Hasil Pengelompokan Metode ROCK

<i>Threshold</i>	Jumlah Klaster	Akurasi
0.05	4	0.711
0.15	2	0.667
0.2	2	0.614

Tabel 4.7 merupakan hasil pengelompokan menggunakan metode ROCK. Dimana jumlah klaster optimum yang dihasilkan adalah empat klaster dengan nilai *threshold* sebesar 0.05 karena memiliki nilai akurasi terbesar yaitu 0.711. Sedangkan nilai *threshold* (θ) lainnya yang tidak tercantum dalam tabel menghasilkan pengelompokan yang sama dengan nilai *threshold* (θ) yang tercantum pada Tabel 4.7.

4.4 Pengelompokan Data Campuran Indikator Daerah Tertinggal di Jawa Timur Menggunakan Metode Ensemble ROCK dan SWFM

Pengelompokan data campuran menggunakan metode *ensemble* ROCK dilakukan dengan menggunakan hasil pengelompokan data numerik menggunakan *ward* dan data kategorik menggunakan ROCK. Hasil pengelompokan data numerik menggunakan metode *ward* terdiri atas lima kelompok. Sedangkan hasil pengelompokan data kategorik menggunakan metode ROCK menghasilkan empat kelompok. Kedua hasil pengelompokan digabung untuk kemudian dianalisis menggunakan *ensemble* ROCK. Struktur data untuk analisis data campuran disajikan pada Tabel 4.8.

Tabel 4.8 Struktur Data Analisis Ensemble ROCK dan SWFM

No	Kabupaten/Kota	Hasil Klaster Numerik	Hasil Klaster Kategorik
1	Pacitan	1	1
2	Ponorogo	1	2
:	:	:	:
36	Kota Malang	5	3
37	Kota Surabaya	4	4
38	Kota Batu	4	1

Hasil pengelompokan untuk data numerik dan kategorik dianggap sebagai variabel baru dengan skala kategorik yang kemudian dianalisis menggunakan metode ROCK. Nilai *threshold* yang digunakan sama dengan nilai *threshold* pada pengelompokan data kategorik. Hasil pengelompokan data campuran menggunakan metode *ensemble* ROCK dapat dilihat pada Tabel 4.9.

Tabel 4.9 Hasil Pengelompokan Ensemble ROCK

Threshold	Jumlah Klaster	Akurasi
0.05	3	0.618
0.1	4	0.724
0.2	4	0.671
0.25	3	0.553
0.3	4	0.618

Tabel 4.9 merupakan hasil pengelompokan metode *ensemble* ROCK. Dimana jumlah klaster optimum adalah 4 klaster dengan *threshold* 0.1. Jumlah klaster optimum tersebut terpilih karena memiliki nilai akurasi tertinggi diantara nilai *threshold* lainnya. Untuk nilai *threshold* 0.15 menghasilkan pengelompokan yang sama dengan nilai *threshold* 0.1.

Selanjutnya dilakukan pengelompokan data campuran menggunakan metode *ensemble* SWFM dengan menggunakan hasil pengelompokan data numerik menggunakan *ward* dan data kategorik menggunakan ROCK. Struktur data yang digunakan sama dengan struktur data pada pengelompokan data campuran menggunakan metode *ensemble* ROCK. Hasil pengelompokan data campuran menggunakan metode *ensemble* SWFM dapat dilihat pada Tabel 4.10.

Tabel 4.10 Hasil Pengelompokan Ensemble SWFM

Jumlah Klaster	Akurasi
2	0.461
3	0.566
4	0.618
5	0.539

Berdasarkan Tabel 4.10, jumlah klaster optimum yang terpilih untuk metode *ensemble* SWFM pada data campuran adalah

empat klaster karena memiliki akurasi terbesar yaitu 0.618. Dimana anggota klaster hasil pengelompokan menggunakan metode *ensemble* SWFM dapat dilihat pada lampiran 13. Selanjutnya dilakukan perbandingan antara metode *ensemble* ROCK dengan SWFM untuk menentukan metode pengelompokan untuk data campuran terbaik. Dimana perbandingan kedua metode dapat dilihat berdasarkan Tabel 4.11.

Tabel 4.11 Perbandingan Metode Ensemble ROCK dan SWFM

Metode	Jumlah Klaster Optimum	Akurasi
<i>Ensemble</i> ROCK	4	0.724
SWFM	4	0.618

Jika dilihat pada Tabel 4.11 dapat disimpulkan bahwa metode pengelompokan terbaik untuk data campuran adalah metode *ensemble* ROCK karena memiliki akurasi lebih baik dibandingkan dengan akurasi dari metode *ensemble* SWFM dengan akurasi yang tertera pada Tabel 4.11.

4.5 Karakteristik Klaster Hasil Pengelompokan Data Campuran Menggunakan Metode Terbaik

Sebelum dilakukan analisis deskriptif terhadap setiap klaster yang terbentuk, akan dilakukan uji signifikansi menggunakan MANOVA dan ANOVA untuk data numerik. Uji Kruskal-Wallis untuk data kategorik. Dimana uji MANOVA dan ANOVA dijelaskan pada subbab 4.5.1.

4.5.1 One-way MANOVA dan ANOVA Data Numerik Hasil Pengelompokan Metode Terbaik

Setelah diperoleh hasil pengelompokan terbaik, akan dilakukan analisis menggunakan metode *one-way* MANOVA pada data numerik untuk mengetahui apakah terdapat perbedaan yang signifikan antara klaster yang terbentuk. Sebelum dilakukan analisis *one-way* MANOVA, dilakukan uji Box'M untuk mengetahui apakah asumsi homogenitas matriks varian kovarian terpenuhi. Hasil uji Box'M disajikan pada Tabel 4.12.

Tabel 4.12 Hasil Uji Box'M

Keterangan	Nilai
F_{hitung}	1.937
$P-value$	0.000

Merujuk pada Tabel 4.12 dapat diketahui bahwa keputusan untuk uji *Box'M* adalah tolak H_0 karena F_{hitung} lebih besar dari $F_{0,05(55;1398,327)}$ yaitu $1.937 > 1.343$ dan $p\text{-value}$ sebesar 0.000 lebih kecil dari $\alpha(0,05)$ yang berarti matriks varian kovarian tidak bersifat homogen dan asumsi tidak terpenuhi.

Mengacu pada uji distribusi normal multivariat, bahwa data tidak berdistribusi normal multivariat dan tidak memenuhi asumsi homogenitas matriks varians kovarians, sehingga menggunakan statistik uji *Pillai's Trace*. Hasil analisis menggunakan *one-way* MANOVA menggunakan *pillai's trace* diperoleh $p\text{-value}$ sebesar 0.015 lebih kecil dari $\alpha(0,05)$ dan F_{hitung} sebesar 1.856 lebih besar dari $F_{0,05(9,29)}$ yang bernilai 0.349. Sehingga diperoleh keputusan tolak H_0 yang artinya hasil pengelompokan Kabupaten/Kota berdasarkan sepuluh variabel numerik secara multivariat memiliki perbedaan yang signifikan. Untuk mengetahui variabel apa saja yang berbeda signifikan antar klaster yang terbentuk dilakukan analisis menggunakan *one-way* ANOVA. Dimana hasil analisis disajikan pada Tabel 4.13.

Tabel 4.13 Hasil One-way ANOVA

Variabel	F_0	$P\text{-value}$
Presentase Jalan Aspal	2.11	0.117
Rumah Sakit per 1000 penduduk	3.20	0.035
Rasio SMA/MA terhadap penduduk usia SMA	2.59	0.069
Presentase Pengguna Listrik PLN	1.29	0.293
Presentase Lahan Kritis	1.99	0.134
Presentase Penduduk Miskin	5.54	0.003
Pengeluaran per Kapita Penduduk	15.52	0.000
Angka Harapan Hidup	8.93	0.000

Tabel 4.13 Hasil One-way ANOVA (Lanjutan)

Variabel	F_0	P-value
Angka Melek Huruf	10.43	0.000
Presentase Pengangguran	3.11	0.039

Berdasarkan hasil *one-way* ANOVA pada Tabel 4.13, dapat diketahui bahwa rasio rumah sakit per 1000 penduduk dari setiap klaster yang terbentuk memiliki perbedaan yang signifikan karena nilai $F_0 > F_{0,05;3,34}$ yaitu $3.20 > 2.88$ dan *p-value* 0.035 lebih kecil dari $\alpha(0.05)$ dimana terdapat minimal sepasang perbedaan. Hal ini juga berlaku pada indikator daerah tertinggal lainnya yaitu presentase penduduk miskin, pengeluaran perkapita penduduk, angka harapan hidup, angka melek huruf dan presentase pengangguran.

4.5.2 Uji Kruskal-Wallis Data Kategorik Hasil Pengelompokan dengan Metode Terbaik

Uji *Kruskal-Wallis* digunakan untuk mengetahui apakah terdapat perbedaan antar klaster yang terbentuk dalam hal indikator daerah tertinggal yang bersifat kategorik yaitu keberadaan kawasan industri, tingkat desentralisasi fiskal dan sumber air minum utama. Hasil uji *Kruskal-Wallis* dapat dilihat pada Tabel 4.14.

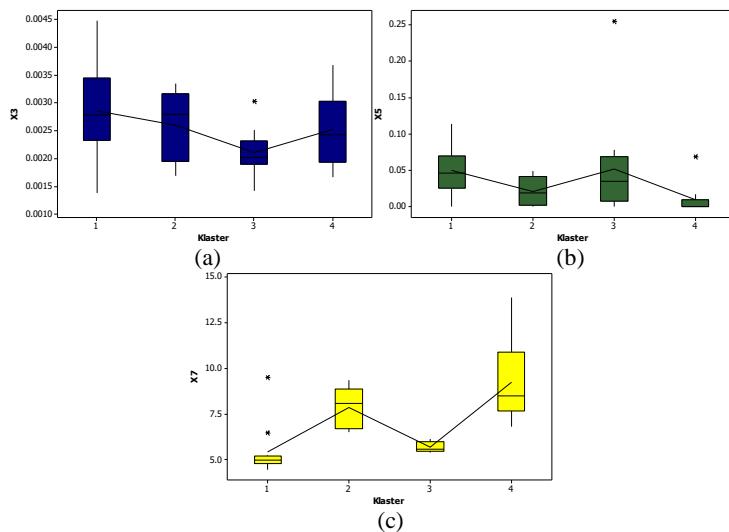
Tabel 4.14 Uji Kruskal-Wallis

Variabel	Chi-Square	P-value
Kawasan Industri	6.847	0.077
Tingkat Desentralisasi Fiskal	11.662	0.009
Sumber Air Minum	20.237	0.000

Tabel 4.14 merupakan hasil uji *Kruskal-Wallis* untuk data kategorik hasil pengelompokan menggunakan metode *ensemble* ROCK. Dimana kesimpulannya terdapat perbedaan yang signifikan antara klaster yang terbentuk jika dilihat pada tingkat desentralisasi fiskal dan sumber air minum karena memiliki *p-value* kurang dari $\alpha(0.05)$. Sedangkan jika dilihat dari keberadaan kawasan industri tidak ada perbedaan yang signifikan antar klaster yang terbentuk.

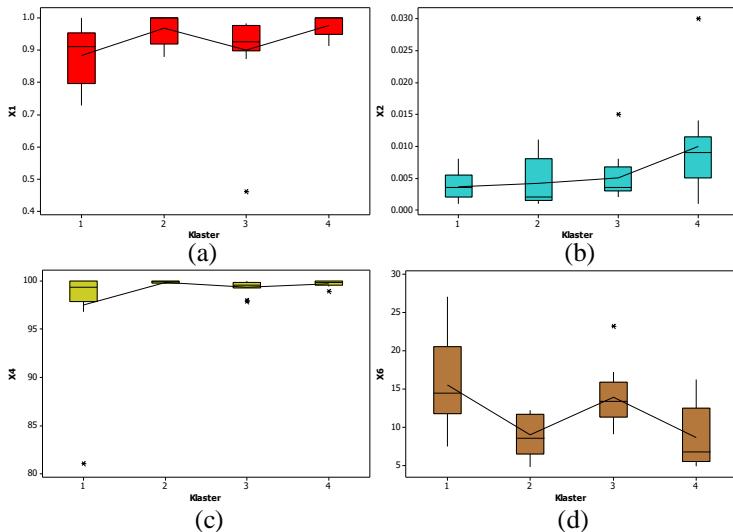
Setelah dilakukan uji signifikansi, selanjutnya dilakukan analisis deskriptif pada setiap klaster yang terbentuk. Gambar

4.2 merupakan perbandingan *boxplot* dari setiap klaster yang terbentuk berdasarkan variabel rasio SMA per penduduk usia SMA, presentase luas lahan kritis dan pengeluaran penduduk per kapita. Sedangkan Gambar 4.4 merupakan perbandingan karakteristik presentase banyaknya desa dengan jalan terluas aspal, rasio rumah sakit per 1000 penduduk, presentase pengguna listrik PLN, dan presentase penduduk miskin.



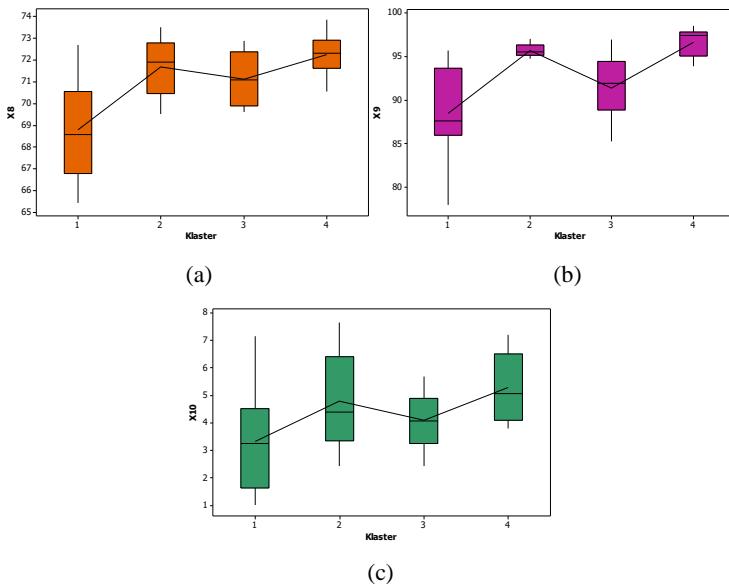
Gambar 4.4 Perbandingan (a) Rasio SMA per Penduduk Usia SMA (b) Presentase Luas Lahan Kritis (c) Pengeluaran Penduduk per Kapita Berdasarkan Gambar 4.4 (a) klaster 3 merupakan klaster dengan rasio SMA per penduduk usia SMA terendah dibandingkan dengan klaster lainnya yaitu 0.0021. Jika dilihat berdasarkan Gambar 4.4(b) rata-rata presentase lahan kritis paling banyak ada pada klaster 1 yaitu 4.96%. Namun, untuk rasio SMA per penduduk usia SMA dan rata-rata presentase luas lahan kritis dari keempat klaster memiliki perbedaan yang tidak signifikan. Hal ini didukung dengan hasil uji signifikansi menggunakan ANOVA dimana dari kedua variabel tersebut memiliki perbedaan yang tidak signifikan. Sedangkan pada Gambar

4.4(c), rata-rata pengeluaran per kapita penduduk klaster 3 dan 4 sangat tinggi yaitu berturut-turut 783400 dan 923800 rupiah. Hal ini menunjukkan penduduk pada klaster 3 dan 4 lebih sejahtera dibandingkan penduduk pada klaster 1 dan 2, jika dilihat berdasarkan rata-rata pengeluaran penduduk per kapita.



Gambar 4.5 Perbandingan Karakteristik (a) Presentase Jalan Aspal
 (b) Rasio rumah sakit per 1000 penduduk (c) Presentase Pengguna Listrik PLN (d) Presentase Penduduk Miskin

Klaster 1 merupakan klaster paling tertinggal karena rata-rata presentase jalan aspal, rasio rumah sakit per penduduk dan presentase pengguna listrik PLN paling rendah berturut-turut 88.29%, 0.003 dan 97.54% sesuai Gambar 4.5(a), (b) dan (c) . Jika dilihat berdasarkan Gambar 4.5 (d), presentase penduduk miskin paling tinggi pada klaster 1 mencapai 15.54%. Hal ini menunjukkan bahwa klaster 1 merupakan klaster paling tertinggal dibandingkan klaster lainnya. Selanjutnya analisis deskriptif perbandingan dari kriteria SDM yaitu angka harapan hidup, angka melek huruf dan presentase pengangguran yang dapat dilihat pada Gambar 4.6.



Gambar 4.6 Perbandingan Karakteristik (a) Angka Harapan Hidup (b) Angka Melek Huruf (c) Presentase Pengangguran

Jika dilihat pada Gambar 4.6 (a) dan (b) klaster 1 merupakan klaster paling tertinggal dibandingkan klaster lainnya karena memiliki rata-rata angka harapan hidup dan angka melek huruf paling rendah yaitu 68.78% dan 88.48%. Hal ini berbanding terbalik dengan klaster 4 yang merupakan klaster paling maju karena memiliki rata-rata angka melek huruf dan angka harapan hidup paling tinggi. Jika dilihat dari presentase pengangguran pada Gambar 4.6(c) yang menunjukkan klaster 4 merupakan klaster dengan rata-rata presentase pengangguran tertinggi yaitu 5.28%. Urutan ketertinggalan masing-masing klaster jika dilihat berdasarkan tiap variabel disajikan pada Tabel 4.15.

Tabel 4.15 Urutan Klaster Tiap Variabel

Variabel	Klaster	Urutan Ketertinggalan			
		1	2	3	4
Rasio SMA per Penduduk Usia SMA (X_3)	Klaster 1				✓
	Klaster 2			✓	
	Klaster 3	✓			
	Klaster 4		✓		
Presentase Luas Lahan Kritis (X_5)	Klaster 1	✓			
	Klaster 2			✓	
	Klaster 3		✓		
	Klaster 4				✓
Pengeluaran Penduduk per Kapita (X_7)	Klaster 1	✓			
	Klaster 2			✓	
	Klaster 3		✓		
	Klaster 4				✓
Presentase Jalan Terluas Aspal (X_1)	Klaster 1	✓			
	Klaster 2			✓	
	Klaster 3		✓		
	Klaster 4				✓
Rasio Rumah Sakit per 1000 Penduduk (X_2)	Klaster 1	✓			
	Klaster 2			✓	
	Klaster 3			✓	
	Klaster 4				✓
Presentase Pengguna Listrik PLN (X_4)	Klaster 1	✓			
	Klaster 2				✓
	Klaster 3		✓		
	Klaster 4			✓	

Tabel 4.15 Urutan Klaster Tiap Variabel (Lanjutan)

Variabel	Klaster	Urutan Ketertinggalan			
		1	2	3	4
Presentase Penduduk Miskin (X ₆)	Klaster 1	✓			
	Klaster 2			✓	
	Klaster 3		✓		
	Klaster 4				✓
Angka Harapan Hidup (X ₈)	Klaster 1	✓			
	Klaster 2			✓	
	Klaster 3		✓		
	Klaster 4				✓
Angka Melek Huruf (X ₉)	Klaster 1	✓			
	Klaster 2			✓	
	Klaster 3		✓		
	Klaster 4				✓
Presentase Pengangguran (X ₁₀)	Klaster 1				✓
	Klaster 2		✓		
	Klaster 3			✓	
	Klaster 4	✓			
Kawasan Industri (X ₁₁)	Klaster 1	✓			
	Klaster 2	✓			
	Klaster 3			✓	
	Klaster 4				✓
Tingkat Desentralisasi Fiskal (X ₁₂)	Klaster 1	✓			
	Klaster 2				✓
	Klaster 3			✓	
	Klaster 4				✓

Tabel 4.15 Urutan Klaster Tiap Variabel (Lanjutan)

Variabel	Klaster	Urutan Ketertinggalan			
		1	2	3	4
Sumber Air	Klaster 1	✓			
Minum Utama (X ₁₃)	Klaster 2			✓	
	Klaster 3		✓		
	Klaster 4				✓

Keterangan urutan: 1. Tertinggal 3. Agak Maju
2. Agak Tertinggal 4. Maju

Jika dilihat berdasarkan variabel rasio SMA per penduduk usia SMA pada Tabel 4.15, urutan klaster dari klaster tertinggal ke maju berturut-turut klaster tiga, empat, dua dan satu. Secara keseluruhan klaster 4 cenderung klaster dengan indikator paling baik dibandingkan dengan klaster lainnya karena unggul dalam sepuluh variabel yaitu presentase lahan kritis, pengeluaran penduduk per kapita, presentase jalan terluas aspal, rasio rumah sakit per penduduk, presentase penduduk miskin, angka harapan hidup, angka melek huruf, keberadaan kawasan industri, tingkat desentralisasi fiskal dan sumber air minum utama. Pada variabel keberadaan Kawasan industri, klaster satu dan dua berada pada urutan yang sama karena memiliki presentase Kabupaten/Kota termasuk kategori 2 (tidak ada kawasan industri) yang sama. Berikut variabel yang menyebabkan ketertinggalan dari setiap klaster disajikan tabel rangkuman Tabel 4.16.

Tabel 4.16 Variabel Terendah Setiap Klaster

Berdasarkan Tabel 4.16, klaster 1 diduga merupakan klaster yang berisikan Kabupaten/Kota tertinggal karena memiliki nilai paling rendah pada sebelas variabel. Tabel 4.17 digunakan untuk melihat anggota dari masing-masing klaster.

Tabel 4.17 Anggota Klaster

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

Klaster satu terdiri atas 12 Kabupaten/Kota seperti yang terlihat pada Tabel 4.13. Klaster dua dan tiga masing-masing terdiri atas 5 dan 12 Kabupaten/Kota. Sedangkan untuk klaster 4 terdiri atas 9 Kabupaten/Kota. Selanjutnya dilakukan pemetaan yang dapat dilihat pada Gambar 4.7.

**Gambar 4.7 Hasil Pemetaan Menggunakan Ensemble ROCK**

Gambar 4.7 merupakan hasil pemetaan Kabupaten/Kota di Jawa Timur menggunakan metode *ensemble* ROCK. Dari 38 Kabupaten/Kota di Jawa Timur, terdapat 32% Kabupaten/Kota yang termasuk klaster 1 (tertinggal). Selain itu, 50% Kabupaten/Kota yang termasuk klaster 1 (tertinggal) berada di wilayah timur provinsi Jawa Timur dan 25% merupakan Kabupaten/Kota yang berada di Pulau Madura. Total ada 75% Kabupaten di Pulau Madura termasuk pada klaster 1.

(Halaman ini sengaja dikosongkan)

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang didapatkan dari hasil penelitian ini adalah sebagai berikut.

1. Karakteristik Kabupaten/Kota di Jawa Timur berdasarkan indikator daerah tertinggal dengan lima kriteria yaitu
 - a. Kriteria infrastruktur, untuk presentase banyaknya desa dengan jalan terluas aspal sudah baik yaitu mencapai 92.2%. Untuk infrastruktur kesehatan dan pendidikan. Rata-rata rumah sakit di Jawa Timur setiap rumah sakit menampung 200 penduduk. Sedangkan standar pendidikan Kabupaten/Kota di Jawa Timur rasio SMA per penduduk usia SMA mencapai 0.002. Rumah tangga pengguna listrik di Jawa Timur cukup merata karena rata-rata mencapai 98.95%. Sumber air minum utama di Jawa Timur adalah bersumber dari sumur dan ledeng tanpa meteran/bor/pompa.
 - b. Berdasarkan karakteristik daerah, penanganan lahan kritis sudah baik mengingat rata-rata presentase luas lahan kritis sangat kecil. Keberadaan Kawasan industri di Jawa Timur sangat sedikit, hal ini ditunjukkan dengan presentase hanya 11% Kabupaten/Kota yang memiliki kawasan industri.
 - c. Kriteria ekonomi diukur dengan presentase penduduk miskin dan pengeluaran per kapita penduduk. Rata-rata presentase penduduk miskin di Jawa Timur mencapai 12.54%. Rata-rata pengeluaran per kapita cukup tinggi yaitu Rp. 671900,-.
 - d. Segi sumber daya manusia. Rata-rata angka harapan hidup dan angka melek huruf mencapai 70.728 dan 92.301. Presentase pengangguran mencapai 4.23%.

- e. Berdasarkan kemampuan keuangan daerah, Kabupaten/Kota di Jawa Timur kemampuannya sangat buruk karena 89% berada pada kategori sangat kurang dan kurang.
- 2. Pengelompokan data campuran menggunakan *ensemble* ROCK menghasilkan empat kelompok optimum dengan akurasi sebesar 0.724. Dimana untuk data numerik, hasil pengelompokan terbaik diperoleh menggunakan metode *ward* dengan tiga variabel hasil analisis faktor yang menghasilkan lima kelompok optimum dengan *pseudo-f* dan *icdrate* masing-masing sebesar 163.408 dan 0.048. Sedangkan pengelompokan data kategorik menggunakan metode ROCK yang menghasilkan empat kelompok optimum dengan akurasi sebesar 0.711.
- 3. Hasil pengelompokan menggunakan *ensemble* SWFM adalah lima kelompok optimum dengan akurasi sebesar 0.618 .
- 4. Perbandingan pengelompokan data campuran dengan *ensemble* ROCK dan SWFM menunjukkan bahwa metode *ensemble* ROCK lebih baik karena memiliki akurasi yang lebih tinggi.

Berdasarkan hasil akhir pengelompokan menggunakan *ensemble* ROCK menunjukkan bahwa klaster 1 terindikasi Kabupaten/Kota yang tergolong tertinggal dibanding Kabupaten/Kota pada klaster lain karena memiliki nilai terendah pada sebelas indikator. Dari 38 Kabupaten/Kota di Jawa Timur, terdapat 32% Kabupaten/Kota yang termasuk klaster 1 (tertinggal). Selain itu, 50% Kabupaten/Kota yang termasuk klaster 1 (tertinggal) berada di wilayah timur dan 25% merupakan Kabupaten/Kota yang berada di Pulau Madura. Total 75% Kabupaten di Pulau Madura termasuk pada klaster 1.

5.2 Saran

Berdasarkan hasil penelitian, saran yang dapat diberikan sebagai berikut:

1. Diperlukannya kebijakan pemerintah untuk mengentaskan permasalahan daerah tertinggal sesuai dengan karakteristik masing-masing klaster.
2. Penelitian selanjutnya disarankan menggunakan unit sampling lebih kecil yaitu desa di Jawa Timur, untuk memperoleh informasi yang lebih detail terkait desa tertinggal sehingga target pembangunan tercapai.
3. Pemeriksaan kelayakan analisis faktor menggunakan uji *Bartlett* ketika data tidak berdistribusi normal multivariat menyebabkan hasil pengujian diduga tidak valid. Sehingga untuk penelitian selanjutnya, dipertimbangkan untuk menggunakan metode pemeriksaan atau pendekatan lainnya.
4. Dalam penerapan metode *ensemble* ROCK, pengelompokan menggunakan nilai *theta* yang lebih bervariasi sehingga bisa didapatkan kelompok yang lebih optimum.
5. Hasil penelitian menunjukkan akurasi yang diperoleh tidak terlalu tinggi, sehingga bisa dipertimbangkan untuk menggunakan metode pengelompokan data campuran yang lain.

(Halaman ini sengaja dikosongkan)

DAFTAR PUSTAKA

- Agresti, A. (2007). *Categorical Data Analysis (2nd ed.)*. New York: John Wiley & Sons, INC.
- Alva, J., & Estrada, E. (2009). A Generalization of Shapiro-Wilk's Test for Multivariate-Normality. *Communication in Statistics-Theory and Methods*, 38(11), 1870-1883.
- Alvionita. (2017). *Metode Ensemel ROCK dan SWFM untuk Pengelompokan Data Campuran Numerik dan Kategorik pada Kasus Akses Jeruk*. Surabaya: Statistika, Institut Teknologi Sepuluh Nopember.
- Anderson, T., & Sclove, S. (1974). *Introductory Statistical Analysis*. Boston: Houghton Mifflin,.
- Badan Pusat Statistik. (2015). *Indikator Pembangunan Berkelanjutan*. Jakarta: Badan Pusat Statistik.
- Bappenas. (2016). *Pemantauan Pelaksanaan Pembangunan Daerah tertinggal dan Kawasan Perbatasan dalam Lingkup RPJMN 2015-2019*. Jakarta: Direktorat Daerah Tertinggal, Transmigrasi, dan Perdesaan.
- Bappenas, & BPS. (2015). *Indeks Pembangunan Desa*. Jakarta: Bappenas.
- Ben-Hur, A., & Guyon, I. (2003). *Detecting Stable Clusters Using Principal Component Analysis. In Functional Geonomics: Methods and Protocols*. Humana Press.
- Black, W. C., Hair, J. F., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis Seventh Edition*. New Jersey: Prentice Hall.
- Bunkers, M. J., & James, R. M. (1996). Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. *J. Climate*.
- Daniel, W. (1989). *Statistika Nonparametrik Terapan*. (A. T. W., Trans.) Jakarta: PT. Gramedia.
- Dewangan, R. R., Sharma, L., & Akasapu, A. K. (2010). Fuzzy Clustering Technique for Numerical and Categorical

- Dataset. *International Journal on Computer Science and Engineering*, hal 75-80.
- Dewi, A. (2012). *Metode Cluster Ensemble untuk Pengelompokan Desa Pedesaan di Provinsi Riau*. Surabaya: Statistika, Institut Teknologi Sepuluh Nopember.
- Dutta, M., Mahanta, A. K., & Arun, K. P. (2005). QROCK : A Quick Version of the ROCK Algorithm for Clustering of Categorical Data. *Proceeding of the 15th IEEE International Conference on Data Engineering*.
- Gong, X., & Richman, M. (1995). *On the Application of Cluster Analysis to Growing Season Precipitation Data in North America East of the Rockies*. North America: J. Climate.
- Guha, S., & Rastogi, R. &. (2000). *ROCK : A Robust Clustering Algorithm for Categorical Attributes*. *Proceedings of the 15th Internasional Conference on Data Engineering*.
- Hair, J. F., Black, W. C., Babin, J., & Anderson, E. (2009). *Multivariate Data Analysis (seventh ed.)*. New Jersey: Prentice Hall Inc.
- Halkidi, Batistakis, & Vizirgiannis. (2001). On Clustering Validation Techniques. *Journal of Intelligent Systems*, 17:2/3, hal 107-145.
- He, Z., Xu, X., & Deng, S. (2005). A Cluster Ensemble Method For Clustering Categorical Data. 143-151.
- Huang, Z., & Ng, M. K. (1999). A Fuzzy K-Modes Algorithm for Clustering Categorical Data. *IEEE Transactions On Fuzzy Systems*, 7(4).
- Jain, A., & Dubes, R. (1988). *Algorithms for Clustering Data*. New Jersey: Prentice Hall.
- Johnson, R. A., & Bhattacharyya, G. K. (2010). *Statistics Principle and Method* (6th ed.). United State of America: John Wiley & Sons, Inc.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Pearson Prentice Hall.

- Kader, D. G., & Perry, M. (2007). *Journal of Statistics Education*, 15 2.
- Kementerian DPDTT. (2013). *Kebijakan dan Strategi Pengembangan Kawasan Tertinggal dalam Mendukung Upaya Pemerataan dan Pertumbuhan Ekonomi Indonesia*. Jakarta: Kementerian Pembangunan Daerah Republik Indonesia.
- Kementerian DPDTT. (2016). *Petunjuk Pelaksanaan Identifikasi Masalah-Masalah Ketertinggalan Kabupaten Daerah Tertinggal*. Jakarta: Direktorat Jenderal Pembangunan Daerah Tertinggal.
- Khan, S. S., & Kant, S. (2007). Computation of Initial Modes for K-modes Clustering Algorithm using Evidence. *IJCAI (International Joint Conference on Artificial Intelligence Journal Division)*, 2.
- Prakoso, A. J. (2017). *Pengelompokan Sekolah Menengah Atas (SMA) di Sidoarjo Menggunakan Similarity Weight and Filter Method*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Purwadi, I. (2009). *Penerapan Bayesian Network dalam Penetapan Daerah Tertinggal*. Bogor: IPB.
- Putri, I. P. (2017). *Pengelompokan Perguruan Tinggi Negeri Di Indonesia Menggunakan Metode Ensemble Robust Clustering Using Links*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Ramdhany, R. (2018). *Pengelompokan Desa Di Kabupaten Bondowoso Berdasarkan Data Campuran Numerik dan Kategorik Menggunakan Metode Ensemble ROCK*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Reddy, M. J., & Kavitha, B. (2012). Clustering The Mixed Numerical and Categorical Dataset Using Similarity Weight and Filter Method. *International Journal of Database Theory and Application*.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2 ed.). Canada: John Wiley & Sons, Inc.

- Sentosa, P. B., & Hamdani, M. (2007). *Statistika Deskriptif dalam Bidang Ekonomi dan Niaga*. Jakarta: Erlangga.
- Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley and Sons Inc.
- Siagian, S. (2008). *Administrasi Pembangunan*. Jakarta: PT. Bumi Aksara.
- Suguna, J., & Selvi, M. A. (2012). Ensemble Fuzzy Clustering for Mixed for Numerical and Categorical Data. *Intenational Journal of Computer Application*, 42, 19.
- Tyagi, A., & Sharma, S. (2012). Implementation of ROCK Clustering Algorithm for the Optimazation of Query Searching Time. *International Journal on Computer Science and Engineering*, Vol 4, No 05.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2007). *Probability & Statistics for Engineers & Scientist*. New Jersey: Pearson Prentice Hall.
- Weiss, N. A. (2008). *Introductory Statistics Eighth Edition*. Boston: Pearson Education, Inc.

LAMPIRAN

Lampiran 1 Data Penelitian Skala Numerik

Kab/Kota	X1	X2	X3	...	X9	X10
Pacitan	0.9357	0.002	0.0023	...	89.57	1.08
Ponorogo	0.899	0.006	0.0029	...	95.02	3.66
Trenggalek	0.8726	0.003	0.0023	...	95.4	4.2
Tulungagung	0.9705	0.007	0.0014	...	96.97	2.42
Blitar	0.9798	0.004	0.0018	...	92.21	3.08
Kediri	0.9826	0.006	0.0019	...	93.16	4.91
Malang	0.9718	0.015	0.0022	...	93.27	4.83
Lumajang	0.922	0.003	0.0023	...	87.03	2.83
Jember	0.7298	0.006	0.0025	...	89.77	4.64
Banyuwangi	0.9309	0.008	0.002	...	94.99	7.17
Bondowoso	0.7306	0.003	0.0035	...	86.91	3.72
Situbondo	0.9706	0.004	0.003	...	85.77	4.15
Probolinggo	0.9606	0.004	0.0034	...	86.41	1.47
Pasuruan	0.8904	0.003	0.0019	...	94.8	4.43
Sidoarjo	0.9914	0.014	0.0018	...	98.05	3.88
Mojokerto	0.9408	0.009	0.0029	...	94.09	3.81
Jombang	0.8791	0.011	0.0033	...	95.66	4.39
Nganjuk	0.9296	0.005	0.0023	...	91.71	3.93
Madiun	0.9223	0.002	0.002	...	88.79	3.38
Magetan	0.9617	0.001	0.0022	...	95.57	4.28
Ngawi	0.4608	0.003	0.002	...	89.89	5.61
Bojonegoro	0.9233	0.008	0.003	...	89.24	3.21
Tuban	0.9817	0.003	0.0019	...	86.3	3.63
Lamongan	0.9135	0.009	0.0037	...	93.92	4.3
Gresik	0.9579	0.009	0.0028	...	97.54	5.06
Bangkalan	0.9217	0.002	0.0025	...	85.27	5.68

Lampiran 1 Data Penelitian Skala Numerik (Lanjutan)

Kab/Kota	X1	X2	X3	...	X9	X10
Sampang	0.7796	0.001	0.0027	...	77.93	2.22
Pamekasan	0.8836	0.002	0.0039	...	88.27	2.14
Sumenep	0.8524	0.001	0.0045	...	84.37	1.01
Kota Kediri	1	0.005	0.0017	...	97.04	7.66
Kota Blitar	1	0.004	0.0014	...	95.67	5.71
Kota Malang	1	0.008	0.0024	...	97.45	7.22
Kota Probolinggo	1	0.002	0.003	...	95.57	5.16
Kota Pasuruan	1	0.001	0.0017	...	98.52	6.09
Kota Mojokerto	1	0.005	0.002	...	97.44	4.42
Kota Madiun	1	0.005	0.0032	...	96.05	6.93
Kota Surabaya	1	0.03	0.0022	...	97.04	5.82
Kota Batu	1	0.002	0.0028	...	94.79	2.43

Lampiran 2 Data Penelitian Skala Kategorik

Kab/Kota	X11	X12	X13
Kabupaten Pacitan	2	1	6
Kabupaten Ponorogo	2	2	4
Kabupaten Trenggalek	2	1	5
Kabupaten Tulungagung	2	2	5
Kabupaten Blitar	2	1	5
Kabupaten Kediri	2	2	4
Kabupaten Malang	2	2	6
Kabupaten Lumajang	2	2	5
Kabupaten Jember	2	2	5
Kabupaten Banyuwangi	2	2	5
Kabupaten Bondowoso	2	1	6
Kabupaten Situbondo	2	1	4
Kabupaten Probolinggo	2	2	4
Kabupaten Pasuruan	1	2	4
Kabupaten Sidoarjo	2	4	1
Kabupaten Mojokerto	1	2	4
Kabupaten Jombang	2	2	4
Kabupaten Nganjuk	2	2	4
Kabupaten Madiun	2	1	4
Kabupaten Magetan	2	1	2
Kabupaten Ngawi	2	2	5
Kabupaten Bojonegoro	2	2	4
.	.	.	.
.	.	.	.
.	.	.	.
Kota Madiun	2	2	1
Kota Surabaya	1	6	1
Kota Batu	2	2	6

Lampiran 3 Syntax Pengelompokan Data Numerik

```

#Metode Ward 2 Kelompok
klasterdua=function(data){
  datanumerik=data.frame(data$X3, data$X5, data$X7)
  #Jarak yang digunakan
  d = dist(datanumerik, method = "euclidean") # distance matrix
  #Metode yang digunakan
  fit = hclust(d, method="ward")
  plot(fit) # display dendrogram
  #Jumlah klaster
  rect.hclust(fit, k=2, border="red")
  ward = cutree(fit, k=2) # cut tree into 2 clusters
  hasil=ward
  y=c((data$X3-mean(data$X3))^2,(data$X5-mean(data$X5))^2,(data$X7-
mean(data$X7))^2) SST=sum(y)
  #Menghitung kinerja pengelompokan
  kelompok.ward=hasil
  data.ward=data.frame(ward, data$X3, data$X5, data$X7)
  data.ward.sort=data.ward[order(data.ward$ward),]
  a=length(data.ward.sort$ward[data.ward.sort$ward==1])
  b=length(data.ward.sort$ward[data.ward.sort$ward==2])
  data.c1.ward=data.ward.sort[1:a,]
  data.c2.ward=data.ward.sort[(a+1):38,]
  xlc1bar.ward=matrix(c(rep(mean(data.c1.ward$data.X3),a),rep(mean(dat
a.c1.ward$data.X5),a),rep(mean(data.c1.ward$data.X7),a)), ncol=3)
  xlc2bar.ward=matrix(c(rep(mean(data.c2.ward$data.X3),b),rep(mean(dat
a.c2.ward$data.X5),b),rep(mean(data.c2.ward$data.X7),b)), ncol=3)
  xil.c1.ward=matrix(c(data.c1.ward$data.X3,data.c1.ward$data.X5,data.c1
.ward$data.X7),ncol=3)
  xil.c2.ward=matrix(c(data.c2.ward$data.X3,data.c2.ward$data.X5,data.c2
.ward$data.X7),ncol=3)
  x1.ward=sum((xil.c1.ward-xlc1bar.ward)^2)
  x2.ward=sum((xil.c2.ward-xlc2bar.ward)^2)
  SSW.ward.c2=sum(x1.ward,x2.ward)
  SSB.ward.c2=SST-SSW.ward.c2
  Rsquare.ward.c2=SSB.ward.c2/SST
  pseudoF.ward.c2=(Rsquare.ward.c2/(2-1))/((1-Rsquare.ward.c2)/(38-2))
  hasil.ward2=c(SST, SSW.ward.c2, SSB.ward.c2, Rsquare.ward.c2,
pseudoF.ward.c2)
  validasi=data.frame(SST, SSW.ward.c2, SSB.ward.c2, Rsquare.ward.c2,
pseudoF.ward.c2)
  print(validasi)
  print(hasil)}

```

Lampiran 3 Syntax Pengelompokan Data Numerik (Lanjutan)

```
#Metode Ward 3 Kelompok
klastertinga=function(data){
  datanumerik=data.frame(data$X3, data$X5, data$X7)
  #Jarak dan metode yang digunakan
  d = dist(datanumerik, method = "euclidean") # distance matrix
  fit = hclust(d, method="ward")
  plot(fit) # display dendrogram
  rect.hclust(fit, k=3, border="red") #Jumlah klaster
  ward = cutree(fit, k=3)
  hasil = ward
  y=c((data$X3-mean(data$X3))^2,(data$X5-
mean(data$X5))^2,(data$X7-mean(data$X7))^2)
  SST=sum(y)
  #Menghitung Kinerja pengelompokan
  kelompok.ward=hasil
  data.ward=data.frame(ward, data$X3, data$X5, data$X7)
  data.ward.sort=data.ward[order(data.ward$ward),]
  a=length(data.ward.sort$ward[data.ward.sort$ward==1])
  b=length(data.ward.sort$ward[data.ward.sort$ward==2])
  c=length(data.ward.sort$ward[data.ward.sort$ward==3])
  data.c1.ward=data.ward.sort[1:a,]
  data.c2.ward=data.ward.sort[(a+1):(a+b),]
  data.c3.ward=data.ward.sort[(a+b+1):38,]
  xlc1bar.ward=matrix(c(rep(mean(data.c1.ward$data.X3),a),rep(mean(dat
a.c1.ward$data.X5),a),rep(mean(data.c1.ward$data.X7),a)), ncol=3)
  xlc2bar.ward=matrix(c(rep(mean(data.c2.ward$data.X3),b),rep(mean(dat
a.c2.ward$data.X5),b),rep(mean(data.c2.ward$data.X7),b)), ncol=3)
  xlc3bar.ward=matrix(c(rep(mean(data.c3.ward$data.X3),c),rep(mean(dat
a.c3.ward$data.X5),c),rep(mean(data.c3.ward$data.X7),c)), ncol=3)
  xil.c1.ward=matrix(c(data.c1.ward$data.X3,data.c1.ward$data.X5,data.c1
.ward$data.X7),ncol=3)
  xil.c2.ward=matrix(c(data.c2.ward$data.X3,data.c2.ward$data.X5,data.c2
.ward$data.X7),ncol=3)
  xil.c3.ward=matrix(c(data.c3.ward$data.X3,data.c3.ward$data.X5,data.c3
.ward$data.X7),ncol=3)
  x1.ward=sum((xil.c1.ward-xlc1bar.ward)^2)
  x2.ward=sum((xil.c2.ward-xlc2bar.ward)^2)
  x3.ward=sum((xil.c3.ward-xlc3bar.ward)^2)
  SSW.ward.c3=sum(x1.ward,x2.ward,x3.ward)
  SSB.ward.c3=SST-SSW.ward.c3
  Rsquare.ward.c3=SSB.ward.c3/SST
  pseudoF.ward.c3=(Rsquare.ward.c3/(3-1))/((1-Rsquare.ward.c3)/(38-3))
```

Lampiran 3 Syntax Pengelompokan Data Numerik (Lanjutan)

```

hasil.ward3=c(SST, SSW.ward.c3, SSB.ward.c3, Rsquare.ward.c3,
pseudoF.ward.c3)
validasi=data.frame(SST, SSW.ward.c3, SSB.ward.c3, Rsquare.ward.c3,
pseudoF.ward.c3)
print("Validasi Pengelompokan")
print(validasi)
print("Hasil Pengelompokan")
print(hasil)
}

#Metode Ward 4 Kelompok
klasterempat=function(data){
  datanumerik=data.frame(data$X3, data$X5, data$X7)
  d = dist(datanumerik, method = "euclidean") # jarak yang digunakan
  fit = hclust(d, method="ward") #metode yang digunakan
  plot(fit) # display dendrogram
  rect.hclust(fit, k=4, border="red")
  ward = cutree(fit, k=4) # cut tree into 4 clusters
  hasil = ward
  y=c((data$X3-mean(data$X3))^2,(data$X5-
mean(data$X5))^2,(data$X7-mean(data$X7))^2)
  SST=sum(y)
#Menghitung kinerja pengelompokan
kelompok.ward=hasil
data.ward=data.frame(ward, data$X3, data$X5, data$X7)
data.ward.sort=data.ward[order(data.ward$ward),]
a=length(data.ward.sort$ward[data.ward.sort$ward==1])
b=length(data.ward.sort$ward[data.ward.sort$ward==2])
c=length(data.ward.sort$ward[data.ward.sort$ward==3])
d=length(data.ward.sort$ward[data.ward.sort$ward==4])
data.c1.ward=data.ward.sort[1:a,]
data.c2.ward=data.ward.sort[(a+1):(a+b),]
data.c3.ward=data.ward.sort[(a+b+1):(a+b+c),]
data.c4.ward=data.ward.sort[(a+b+c+1):38,]
xlc1bar.ward=matrix(c(rep(mean(data.c1.ward$data.X3),a),rep(mean(dat
a.c1.ward$data.X5),a),rep(mean(data.c1.ward$data.X7),a)), ncol=3)
xlc2bar.ward=matrix(c(rep(mean(data.c2.ward$data.X3),b),rep(mean(dat
a.c2.ward$data.X5),b),rep(mean(data.c2.ward$data.X7),b)), ncol=3)
xlc3bar.ward=matrix(c(rep(mean(data.c3.ward$data.X3),c),rep(mean(dat
a.c3.ward$data.X5),c),rep(mean(data.c3.ward$data.X7),c)), ncol=3)
xlc4bar.ward=matrix(c(rep(mean(data.c4.ward$data.X3),d),rep(mean(dat
a.c4.ward$data.X5),d),rep(mean(data.c4.ward$data.X7),d)), ncol=3)

```

Lampiran 3 Syntax Pengelompokan Data Numerik (Lanjutan)

```

xil.c1.ward=matrix(c(data.c1.ward$data.X3,data.c1.ward$data.X5,data.c1
.ward$data.X7),ncol=3)
xil.c2.ward=matrix(c(data.c2.ward$data.X3,data.c2.ward$data.X5,data.c2
.ward$data.X7),ncol=3)
xil.c3.ward=matrix(c(data.c3.ward$data.X3,data.c3.ward$data.X5,data.c3
.ward$data.X7),ncol=3)
xil.c4.ward=matrix(c(data.c4.ward$data.X3,data.c4.ward$data.X5,data.c4
.ward$data.X7),ncol=3)
x1.ward=sum((xil.c1.ward-xlc1bar.ward)^2)
x2.ward=sum((xil.c2.ward-xlc2bar.ward)^2)
x3.ward=sum((xil.c3.ward-xlc3bar.ward)^2)
x4.ward=sum((xil.c4.ward-xlc4bar.ward)^2)
#Menghitung kinerja pengelompokan
SSW.ward.c4=sum(x1.ward,x2.ward,x3.ward,x4.ward)
SSB.ward.c4=SST-SSW.ward.c4
Rsquare.ward.c4=SSB.ward.c4/SST
pseudoF.ward.c4=(Rsquare.ward.c4/(4-1))/((1-Rsquare.ward.c4)/(38-4))
hasil.ward4=c(SST, SSW.ward.c4, SSB.ward.c4, Rsquare.ward.c4,
pseudoF.ward.c4)
validasi=data.frame(SST, SSW.ward.c4, SSB.ward.c4, Rsquare.ward.c4,
pseudoF.ward.c4)
print(validasi)
print("Hasil Pengelompokan")
print(hasil)}
#Metode Ward 5 Kelompok
klasterlima=function(data){
datanumerik=data.frame(data$X3, data$X5, data$X7)
d = dist(datanumerik, method = "euclidean") # distance matrix
fit = hclust(d, method="ward")
plot(fit) # display dendrogram
rect.hclust(fit, k=5, border="red")
ward = cutree(fit, k=5) # cut tree into 5 clusters
hasil = ward
y=c((data$X3-mean(data$X3))^2,(data$X5-
mean(data$X5))^2,(data$X7-mean(data$X7))^2)
#Menghitung kinerja pengelompokan
SST=sum(y)
kelompok.ward=hasil
data.ward=data.frame(ward, data$X3, data$X5, data$X7)
data.ward.sort=data.ward[order(data.ward$ward),]
a=length(data.ward.sort$ward[data.ward.sort$ward==1])
b=length(data.ward.sort$ward[data.ward.sort$ward==2])

```

Lampiran 3 Syntax Pengelompokan Data Numerik (Lanjutan)

```

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,]
xlc1bar.ward=matrix(c(rep(mean(data.c1.ward$data.X3),a),rep(mean(data.c1.ward$data.X5),a),rep(mean(data.c1.ward$data.X7),a)), ncol=3)
xlc2bar.ward=matrix(c(rep(mean(data.c2.ward$data.X3),b),rep(mean(data.c2.ward$data.X5),b),rep(mean(data.c2.ward$data.X7),b)), ncol=3)
xlc3bar.ward=matrix(c(rep(mean(data.c3.ward$data.X3),c),rep(mean(data.c3.ward$data.X5),c),rep(mean(data.c3.ward$data.X7),c)), ncol=3)
xlc4bar.ward=matrix(c(rep(mean(data.c4.ward$data.X3),d),rep(mean(data.c4.ward$data.X5),d),rep(mean(data.c4.ward$data.X7),d)), ncol=3)
xlc5bar.ward=matrix(c(rep(mean(data.c5.ward$data.X3),e),rep(mean(data.c5.ward$data.X5),e),rep(mean(data.c5.ward$data.X7),e)), ncol=3)
xil.c1.ward=matrix(c(data.c1.ward$data.X3,data.c1.ward$data.X5,data.c1.ward$data.X7),ncol=3)
xil.c2.ward=matrix(c(data.c2.ward$data.X3,data.c2.ward$data.X5,data.c2.ward$data.X7),ncol=3)
xil.c3.ward=matrix(c(data.c3.ward$data.X3,data.c3.ward$data.X5,data.c3.ward$data.X7),ncol=3)
xil.c4.ward=matrix(c(data.c4.ward$data.X3,data.c4.ward$data.X5,data.c4.ward$data.X7),ncol=3)
xil.c5.ward=matrix(c(data.c5.ward$data.X3,data.c5.ward$data.X5,data.c5.ward$data.X7),ncol=3)
x1.ward=sum((xil.c1.ward-xlc1bar.ward)^2)
x2.ward=sum((xil.c2.ward-xlc2bar.ward)^2)
x3.ward=sum((xil.c3.ward-xlc3bar.ward)^2)
x4.ward=sum((xil.c4.ward-xlc4bar.ward)^2)
x5.ward=sum((xil.c5.ward-xlc5bar.ward)^2)
#Menghitung kinerja pengelompokan
SSW.ward.c5=sum(x1.ward,x2.ward,x3.ward,x4.ward,x5.ward)
SSB.ward.c5=SST-SSW.ward.c5
Rsquare.ward.c5=SSB.ward.c5/SST
pseudoF.ward.c5=(Rsquare.ward.c5/(5-1))/((1- Rsquare.ward.c5)/(38-5))
hasil.ward5=c(SST, SSW.ward.c5, SSB.ward.c5, Rsquare.ward.c5,
pseudoF.ward.c5)

```

Lampiran 3 Syntax Pengelompokan Data Numerik (Lanjutan)

```
validasi=data.frame(SST, SSW.ward.c5, SSB.ward.c5,  
Rsquare.ward.c5, pseudoF.ward.c5)  
print("Validasi Pengelompokan")  
print("Hasil Pengelompokan")  
print(hasil)  
}
```

Lampiran 4 Syntax Pengelompokan Data Kategorik

```

data=read.csv("d:/Kuliah/Bismillah Tugas
Akhir/datarun.csv",header=TRUE)
rock=function(data){
  datakategorik1 = c(data$X11, data$X12, data$X13) #data kategorik
  datakategorik2 = matrix (datakategorik1,38,3)
  x = as.dummy(datakategorik2)
  rc.1 = rockCluster(x, n=5, theta=0.05, debug=FALSE) #threshold
  rc.2= rockCluster(x, n=5, theta=0.1, debug=FALSE)
  rc.3= rockCluster(x, n=5, theta=0.15, debug=FALSE)
  rc.4 = rockCluster(x, n=5, theta=0.2, debug=FALSE)
  rc.5 = rockCluster(x, n=5, theta=0.25, debug=FALSE)
  rc.6 = rockCluster(x, n=5, theta=0.3, debug=FALSE)
  rc.7= rockCluster(x, n=5, theta=0.35, debug=FALSE)
  rc.8= rockCluster(x, n=5, theta=0.4, debug=FALSE)
  rc.9 = rockCluster(x, n=5, theta=0.45, debug=FALSE)
  rc.10 = rockCluster(x, n=5, theta=0.5, debug=FALSE)
  rf.1 = fitted(rc.1)
  rf.2 = fitted(rc.2)
  rf.3 = fitted(rc.3)
  rf.4 = fitted(rc.4)
  rf.5 = fitted(rc.5)
  rf.6 = fitted(rc.6)
  rf.7 = fitted(rc.7)
  rf.8 = fitted(rc.8)
  rf.9 = fitted(rc.9)
  rf.10 = fitted(rc.10)
  theta.1 = rf.1$cl
  theta.2 = rf.2$cl
  theta.3 = rf.3$cl
  theta.4 = rf.4$cl
  theta.5 = rf.5$cl
  theta.6 = rf.6$cl
  theta.7 = rf.7$cl
  theta.8 = rf.8$cl
  theta.9 = rf.9$cl
  theta.10 = rf.10$cl
  Kelompok = data.frame( theta.1, theta.2, theta.3, theta.4, theta.5, theta.6,
  theta.7, theta.8, theta.9, theta.10)
  Hasil = data.frame(data$Kab_Kota, Kelompok)
  jumlah.1 = count(Hasil,"theta.1")
  jumlah.2 = count(Hasil,"theta.2")
  jumlah.3 = count(Hasil,"theta.3")
}

```

Lampiran 4 Syntax Pengelompokan Data Kategorik (Lanjutan)

```

jumlah.4 = count(Hasil,"theta.4")
jumlah.5 = count(Hasil,"theta.5")
jumlah.6 = count(Hasil,"theta.6")
jumlah.7 = count(Hasil,"theta.7")
jumlah.8 = count(Hasil,"theta.8")
jumlah.9 = count(Hasil,"theta.9")
jumlah.10 = count(Hasil,"theta.10")
X11 = factor (data$X11)
X12 = factor (data$X12)
X13 = factor (data$X13)
numtheta.1=as.numeric(as.character(theta.1))
numtheta.2=as.numeric(as.character(theta.2))
numtheta.3=as.numeric(as.character(theta.3))
numtheta.4=as.numeric(as.character(theta.4))
numtheta.5=as.numeric(as.character(theta.5))
numtheta.6=as.numeric(as.character(theta.6))
numtheta.7=as.numeric(as.character(theta.7))
numtheta.8=as.numeric(as.character(theta.8))
numtheta.9=as.numeric(as.character(theta.9))
numtheta.10=as.numeric(as.character(theta.10))
p.1= numtheta.1~X11+X12+X13
p.2= numtheta.2~X11+X12+X13
p.3= numtheta.3~X11+X12+X13
p.4= numtheta.4~X11+X12+X13
p.5= numtheta.5~X11+X12+X13
p.6= numtheta.6~X11+X12+X13
p.7= numtheta.7~X11+X12+X13
p.8= numtheta.8~X11+X12+X13
p.9= numtheta.9~X11+X12+X13
p.10= numtheta.10~X11+X12+X13
mylogit.1 = aov( p.1)
mylogit.2 = aov( p.2)
mylogit.3 = aov( p.3)
mylogit.4 = aov( p.4)
mylogit.5 = aov( p.5)
mylogit.6 = aov( p.6)
mylogit.7 = aov( p.7)
mylogit.8 = aov( p.8)
mylogit.9 = aov( p.9)
mylogit.10 = aov( p.10)
SSW.1 = sum((mylogit.1$residuals)^2)
SSW.2 = sum((mylogit.2$residuals)^2)

```

Lampiran 4 Syntax Pengelompokan Data Kategorik (Lanjutan)

```
SSW.3 = sum((mylogit.3$residuals)^2)
SSW.4 = sum((mylogit.4$residuals)^2)
SSW.5 = sum((mylogit.5$residuals)^2)
SSW.6 = sum((mylogit.6$residuals)^2)
SSW.7 = sum((mylogit.7$residuals)^2)
SSW.8 = sum((mylogit.8$residuals)^2)
SSW.9 = sum((mylogit.9$residuals)^2)
SSW.10 = sum((mylogit.10$residuals)^2)
h.1=unlist(summary(mylogit.1))
h.2=unlist(summary(mylogit.2))
h.3=unlist(summary(mylogit.3))
h.4=unlist(summary(mylogit.4))
h.5=unlist(summary(mylogit.5))
h.6=unlist(summary(mylogit.6))
h.7=unlist(summary(mylogit.7))
h.8=unlist(summary(mylogit.8))
h.9=unlist(summary(mylogit.9))
h.10=unlist(summary(mylogit.10))
summary.1 = matrix(h.1, 4,5 )
summary.2 = matrix(h.2, 4,5 )
summary.3 = matrix(h.3, 4,5 )
summary.4 = matrix(h.4, 4,5 )
summary.5 = matrix(h.5, 4,5 )
summary.6 = matrix(h.6, 4,5 )
summary.7 = matrix(h.7, 4,5 )
summary.8 = matrix(h.8, 4,5 )
summary.9 = matrix(h.9, 4,5 )
summary.10 = matrix(h.10, 4,5 )
a=sum(rf.1$size>0, na.rm=TRUE)
b=sum(rf.2$size>0, na.rm=TRUE)
c=sum(rf.3$size>0, na.rm=TRUE)
d=sum(rf.4$size>0, na.rm=TRUE)
e=sum(rf.5$size>0, na.rm=TRUE)
f=sum(rf.6$size>0, na.rm=TRUE)
g=sum(rf.7$size>0, na.rm=TRUE)
h=sum(rf.8$size>0, na.rm=TRUE)
i=sum(rf.9$size>0, na.rm=TRUE)
j=sum(rf.10$size>0, na.rm=TRUE)
SSB.1 = sum(summary.1[1:3,2])
SSB.2 = sum(summary.2[1:3,2])
SSB.3 = sum(summary.3[1:3,2])
SSB.4 = sum(summary.4[1:3,2])
```

Lampiran 4 Syntax Pengelompokan Data Kategorik (Lanjutan)

```

SSB.5 = sum(summary.5[1:3,2])
SSB.6 = sum(summary.6[1:3,2])
SSB.7 = sum(summary.7[1:3,2])
SSB.8 = sum(summary.8[1:3,2])
SSB.9 = sum(summary.9[1:3,2])
SSB.10 = sum(summary.10[1:3,2])
SW.1 = sqrt(SSW.1/(38-a))
SW.2 = sqrt(SSW.2/(38-b))
SW.3 = sqrt(SSW.3/(38-c))
SW.4 = sqrt(SSW.4/(38-d))
SW.5 = sqrt(SSW.5/(38-e))
SW.6 = sqrt(SSW.6/(38-f))
SW.7 = sqrt(SSW.7/(38-g))
SW.8 = sqrt(SSW.8/(38-h))
SW.9 = sqrt(SSW.9/(38-i))
SW.10 = sqrt(SSW.10/(38-j))
SB.1 = sqrt(SSB.1/(a-1))
SB.2 = sqrt(SSB.2/(b-1))
SB.3 = sqrt(SSB.3/(c-1))
SB.4 = sqrt(SSB.4/(d-1))
SB.5 = sqrt(SSB.5/(e-1))
SB.6 = sqrt(SSB.6/(f-1))
SB.7 = sqrt(SSB.7/(g-1))
SB.8 = sqrt(SSB.8/(h-1))
SB.9 = sqrt(SSB.9/(i-1))
SB.10 = sqrt(SSB.10/(j-1))
Ratio.1 = (SW.1/SB.1)
Ratio.2 = (SW.2/SB.2)
Ratio.3 = (SW.3/SB.3)
Ratio.4 = (SW.4/SB.4)
Ratio.5 = (SW.5/SB.5)
Ratio.6 = (SW.6/SB.6)
Ratio.7 = (SW.7/SB.7)
Ratio.8 = (SW.8/SB.8)
Ratio.9 = (SW.9/SB.9)
Ratio.10 = (SW.10/SB.10)
Ratio.ROCK = rbind(Ratio.1, Ratio.2, Ratio.3, Ratio.4, Ratio.5, Ratio.6,
Ratio.7, Ratio.8, Ratio.9, Ratio.10)
print("Hasil Pengelompokan ROCK")
print(Hasil)
print(jumlah.1)
print(jumlah.2)

```

Lampiran 4 Syntax Pengelompokan Data Kategorik (Lanjutan)

```
print(jumlah.3)
print(jumlah.4)
print(jumlah.5)
print(jumlah.6)
print(jumlah.7)
print(jumlah.8)
print(jumlah.9)
print(jumlah.10)
print("-----")
print("Nilai Ratio yang Terbentuk")
print(Ratio.ROCK)
}
```

Lampiran 5 Syntax Pengelompokan Ensemble ROCK

```
Data=read.csv("d:/Kuliah/Bismillah Tugas  
Akhir/datarun.csv",header=TRUE)  
ensemblerock=function(data){  
#Input data hasil pengelompokan numerik dan kategorik  
datakategorik1=c(data$numerik1, data$kategorik)  
datakategorik2=matrix(datakategorik1,38,2)  
x = as.dummy(datakategorik2)  
rc.1 = rockCluster(x, n=5, theta=0.05, debug=FALSE)  
rc.2= rockCluster(x, n=5, theta=0.1, debug=FALSE)  
rc.3= rockCluster(x, n=5, theta=0.15, debug=FALSE)  
rc.4 = rockCluster(x, n=5, theta=0.2, debug=FALSE)  
rc.5 = rockCluster(x, n=5, theta=0.25, debug=FALSE)  
rc.6 = rockCluster(x, n=5, theta=0.3, debug=FALSE)  
rc.7= rockCluster(x, n=5, theta=0.35, debug=FALSE)  
rc.8= rockCluster(x, n=5, theta=0.4, debug=FALSE)  
rc.9 = rockCluster(x, n=5, theta=0.45, debug=FALSE)  
rc.10 = rockCluster(x, n=5, theta=0.5, debug=FALSE)  
rf.1 = fitted(rc.1)  
rf.2 = fitted(rc.2)  
rf.3 = fitted(rc.3)  
rf.4 = fitted(rc.4)  
rf.5 = fitted(rc.5)  
rf.6 = fitted(rc.6)  
rf.7 = fitted(rc.7)  
rf.8 = fitted(rc.8)  
rf.9 = fitted(rc.9)  
rf.10 = fitted(rc.10)  
theta.1 = rf.1$cl  
theta.2 = rf.2$cl  
theta.3 = rf.3$cl  
theta.4 = rf.4$cl  
theta.5 = rf.5$cl  
theta.6 = rf.6$cl  
theta.7 = rf.7$cl  
theta.8 = rf.8$cl  
theta.9 = rf.9$cl  
theta.10 = rf.10$cl  
Kelompok = data.frame( theta.1, theta.2, theta.3, theta.4, theta.5, theta.6,  
theta.7, theta.8, theta.9, theta.10)  
Hasil = data.frame(data$Kab_Kota, Kelompok)  
jumlah.1 = count(Hasil,"theta.1")  
jumlah.2 = count(Hasil,"theta.2")
```

Lampiran 5 Syntax Pengelompokan *Ensemble* ROCK (Lanjutan)

```

jumlah.3 = count(Hasil,"theta.3")
jumlah.4 = count(Hasil,"theta.4")
jumlah.5 = count(Hasil,"theta.5")
jumlah.6 = count(Hasil,"theta.6")
jumlah.7 = count(Hasil,"theta.7")
jumlah.8 = count(Hasil,"theta.8")
jumlah.9 = count(Hasil,"theta.9")
jumlah.10 = count(Hasil,"theta.10")
numerik = factor(data$numerik)
kategorik = factor(data$kategorik)
numtheta.1=as.numeric(as.character(theta.1))
numtheta.2=as.numeric(as.character(theta.2))
numtheta.3=as.numeric(as.character(theta.3))
numtheta.4=as.numeric(as.character(theta.4))
numtheta.5=as.numeric(as.character(theta.5))
numtheta.6=as.numeric(as.character(theta.6))
numtheta.7=as.numeric(as.character(theta.7))
numtheta.8=as.numeric(as.character(theta.8))
numtheta.9=as.numeric(as.character(theta.9))
numtheta.10=as.numeric(as.character(theta.10))
p.1= numtheta.1~numerik+kategorik
p.2= numtheta.2~numerik+kategorik
p.3= numtheta.3~numerik+kategorik
p.4= numtheta.4~numerik+kategorik
p.5= numtheta.5~numerik+kategorik
p.6= numtheta.6~numerik+kategorik
p.7= numtheta.7~numerik+kategorik
p.8= numtheta.8~numerik+kategorik
p.9= numtheta.9~numerik+kategorik
p.10= numtheta.10~numerik+kategorik
mylogit.1 = aov( p.1)
mylogit.2 = aov( p.2)
mylogit.3 = aov( p.3)
mylogit.4 = aov( p.4)
mylogit.5 = aov( p.5)
mylogit.6 = aov( p.6)
mylogit.7 = aov( p.7)
mylogit.8 = aov( p.8)
mylogit.9 = aov( p.9)
mylogit.10 = aov( p.10)
SSW.1 = sum((mylogit.1$residuals)^2)
SSW.2 = sum((mylogit.2$residuals)^2)

```

Lampiran 5 Syntax Pengelompokan *Ensemble* ROCK (Lanjutan)

```

SSW.3 = sum((mylogit.3$residuals)^2)
SSW.4 = sum((mylogit.4$residuals)^2)
SSW.5 = sum((mylogit.5$residuals)^2)
SSW.6 = sum((mylogit.6$residuals)^2)
SSW.7 = sum((mylogit.7$residuals)^2)
SSW.8 = sum((mylogit.8$residuals)^2)
SSW.9 = sum((mylogit.9$residuals)^2)
SSW.10 = sum((mylogit.10$residuals)^2)
h.1=unlist(summary(mylogit.1))
h.2=unlist(summary(mylogit.2))
h.3=unlist(summary(mylogit.3))
h.4=unlist(summary(mylogit.4))
h.5=unlist(summary(mylogit.5))
h.6=unlist(summary(mylogit.6))
h.7=unlist(summary(mylogit.7))
h.8=unlist(summary(mylogit.8))
h.9=unlist(summary(mylogit.9))
h.10=unlist(summary(mylogit.10))
summary.1 = matrix(h.1, 3,5 )
summary.2 = matrix(h.2, 3,5 )
summary.3 = matrix(h.3, 3,5 )
summary.4 = matrix(h.4, 3,5 )
summary.5 = matrix(h.5, 3,5 )
summary.6 = matrix(h.6, 3,5 )
summary.7 = matrix(h.7, 3,5 )
summary.8 = matrix(h.8, 3,5 )
summary.9 = matrix(h.9, 3,5 )
summary.10 = matrix(h.10, 3,5 )
a=sum(rf.1$size>0, na.rm=TRUE)
b=sum(rf.2$size>0, na.rm=TRUE)
c=sum(rf.3$size>0, na.rm=TRUE)
d=sum(rf.4$size>0, na.rm=TRUE)
e=sum(rf.5$size>0, na.rm=TRUE)
f=sum(rf.6$size>0, na.rm=TRUE)
g=sum(rf.7$size>0, na.rm=TRUE)
h=sum(rf.8$size>0, na.rm=TRUE)
i=sum(rf.9$size>0, na.rm=TRUE)
j=sum(rf.10$size>0, na.rm=TRUE)
SSB.1 = sum(summary.1[1:2,2])
SSB.2 = sum(summary.2[1:2,2])
SSB.3 = sum(summary.3[1:2,2])
SSB.4 = sum(summary.4[1:2,2])

```

Lampiran 5 Syntax Pengelompokan Ensemble ROCK (Lanjutan)

```

SSB.5 = sum(summary.5[1:2,2])
SSB.6 = sum(summary.6[1:2,2])
SSB.7 = sum(summary.7[1:2,2])
SSB.8 = sum(summary.8[1:2,2])
SSB.9 = sum(summary.9[1:2,2])
SSB.10 = sum(summary.10[1:2,2])
SW.1 = sqrt(SSW.1/(38-a))
SW.2 = sqrt(SSW.2/(38-b))
SW.3 = sqrt(SSW.3/(38-c))
SW.4 = sqrt(SSW.4/(38-d))
SW.5 = sqrt(SSW.5/(38-e))
SW.6 = sqrt(SSW.6/(38-f))
SW.7 = sqrt(SSW.7/(38-g))
SW.8 = sqrt(SSW.8/(38-h))
SW.9 = sqrt(SSW.9/(38-i))
SW.10 = sqrt(SSW.10/(38-j))
SB.1 = sqrt(SSB.1/(a-1))
SB.2 = sqrt(SSB.2/(b-1))
SB.3 = sqrt(SSB.3/(c-1))
SB.4 = sqrt(SSB.4/(d-1))
SB.5 = sqrt(SSB.5/(e-1))
SB.6 = sqrt(SSB.6/(f-1))
SB.7 = sqrt(SSB.7/(g-1))
SB.8 = sqrt(SSB.8/(h-1))
SB.9 = sqrt(SSB.9/(i-1))
SB.10 = sqrt(SSB.10/(j-1))
Ratio.1 = (SW.1/SB.1)
Ratio.2 = (SW.2/SB.2)
Ratio.3 = (SW.3/SB.3)
Ratio.4 = (SW.4/SB.4)
Ratio.5 = (SW.5/SB.5)
Ratio.6 = (SW.6/SB.6)
Ratio.7 = (SW.7/SB.7)
Ratio.8 = (SW.8/SB.8)
Ratio.9 = (SW.9/SB.9)
Ratio.10 = (SW.10/SB.10)
Ratio.EnsembleROCK = rbind(Ratio.1, Ratio.2, Ratio.3, Ratio.4, Ratio.5,
Ratio.6, Ratio.7, Ratio.8, Ratio.9, Ratio.10)
print("Hasil Pengelompokan Ensemble ROCK")
print(Hasil)
print(jumlah.1)
print(jumlah.2)

```

Lampiran 5 Syntax Pengelompokan *Ensemble* ROCK (Lanjutan)

```
print(jumlah.3)
print(jumlah.4)
print(jumlah.5)
print(jumlah.6)
print(jumlah.7)
print(jumlah.8)
print(jumlah.9)
print(jumlah.10)
print("-----")
print("Nilai Ratio yang Terbentuk")
print(Ratio.EnsembleROCK)
}
```

Lampiran 6 Syntax Metode SWFM

```

data= read.csv("d/Kuliah/BismillahTugasAkhir/datarun.csv",header=TRUE)
Ensembl.SWFM = function(data, k.numerik, k.kategorik,
theta.kategorik){
  datanumerik = data.frame(data$X1, data$X2, data$X3, data$X4,
data$X5, data$X6, data$X7, data$X8, data$X9, data$X10)
  datakategorik = data.frame(data$X11, data$X12, data$X13)
# Pengelompokan Numerik
d = dist(datanumerik, method = "euclidean")
fit.ward = hclust(d, method = "ward")
k=2
k=k.numerik
ward = cutree(fit.ward, k=2)
ward = cutree(fit.ward, k=k)
# Pengelompokan Kategorik
datakategorik1 = c(data$X11, data$X12, data$X13)
datakategorik2 = matrix (datakategorik1,38,3)
d = as.dummy(datakategorik2)
n=5
n = k.kategorik
t=0.35
t = theta.kategorik
rc = rockCluster(d,n=5,theta=0.35,debug=TRUE)
rc = rockCluster(d,n=n,theta=t,debug=TRUE)
rf.hasil = fitted(rc)
u = rf.hasil$c1
# Pengelompokan Ensembl SWFM
#Menghitung jarak (Similarity Weight)
sij = function(x,y){
  sij = length(intersect(x,y))/length(union(x,y))
  p = length(unique(x))
  q = length(unique(y))
  m = max (p,q)
  sm = sum (sij/m)
return(sm)
}
data1 = c(ward[1],u[1]) #menggabungkan hasil pengelompokan ward dan
ROCK
data2 = c(ward[2],u[2])
data3 = c(ward[3],u[3])
data4 = c(ward[4],u[4])
data5 = c(ward[5],u[5])
data6 = c(ward[6],u[6])

```

Lampiran 6 Syntax Metode SWFM (Lanjutan)

```

data7 = c(ward[7],u[7])
data8 = c(ward[8],u[8])
data9 = c(ward[9],u[9])
data10 = c(ward[10],u[10])
data11 = c(ward[11],u[11])
data12 = c(ward[12],u[12])
data13 = c(ward[13],u[13])
data14 = c(ward[14],u[14])
data15 = c(ward[15],u[15])
data16 = c(ward[16],u[16])
data17 = c(ward[17],u[17])
data18 = c(ward[18],u[18])
data19 = c(ward[19],u[19])
data20 = c(ward[20],u[20])
data21 = c(ward[21],u[21])
data22 = c(ward[22],u[22])
data23 = c(ward[23],u[23])
data24 = c(ward[24],u[24])
data25 = c(ward[25],u[25])
data26 = c(ward[26],u[26])
data27 = c(ward[27],u[27])
data28 = c(ward[28],u[28])
data29 = c(ward[29],u[29])
data30 = c(ward[30],u[30])
data31 = c(ward[31],u[31])
data32 = c(ward[32],u[32])
data33 = c(ward[33],u[33])
data34 = c(ward[34],u[34])
data35 = c(ward[35],u[35])
data36 = c(ward[36],u[36])
data37 = c(ward[37],u[37])
data38 = c(ward[38],u[38])
datax=
rbind(data1,data2,data3,data4,data5,data6,data7,data8,data9,data10,data1
1,data12,data13,data14,data15,data16,data17,data18,data19,data20,data21
,data22,data23,data24,data25,data26,data27,data28,data29,data30,data31
,data32,data33,data34,data35,data36,data37,data38)
sim = matrix (1, nrow = 38, ncol = 38)
rownames(sim) = c(1:38)
colnames(sim) = c(1:38)
for (i in 1:(nrow(sim)-1)){
  for (j in (i+1):nrow(sim)){
    
```

Lampiran 6 Syntax Metode SWFM (Lanjutan)

```

x = data[i,]
y = data[j,]
sim [i,j] = sij(x,y)
sim [j,i] = sim[i,j]
}
}
wi = 0.5
F = wi*(1-sim)
jarak = as.dist(F)
fit.sin = hclust(jarak,method = "single")
hasil.k2 = cutree(fit.sin, k=2)
hasil.k3 = cutree(fit.sin, k=3)
hasil.k4 = cutree(fit.sin, k=4)
hasil.k5 = cutree(fit.sin, k=5)
hasil.k6 = cutree(fit.sin, k=6)
hasil.kelompok
= data.frame(hasil.k2,hasil.k3,hasil.k4,hasil.k5,hasil.k6)
SWFM = data.frame(data$Kab_Kota, datax, hasil.kelompok, ward,u)
write.csv(SWFM, "d:/Kuliah/data1.csv")
SWFM1 = read.csv("d:/Kuliah/data1.csv", header=TRUE, sep=",")
SWFM1 = read.csv("Data Hasil Ensemel SWFM.csv", header=TRUE,
sep=",")
X1=SWFM1$X1
X2=SWFM1$X2
p.2 = hasil.k2~X1+X2
p.3 = hasil.k3~X1+X2
p.4 = hasil.k4~X1+X2
p.5 = hasil.k5~X1+X2
p.6 = hasil.k6~X1+X2
model.2 = aov(p.2)
model.3 = aov(p.3)
model.4 = aov(p.4)
model.5 = aov(p.5)
model.6 = aov(p.6)
SSW.2 = sum((model.2$residuals)^2)
SSW.3 = sum((model.3$residuals)^2)
SSW.4 = sum((model.4$residuals)^2)
SSW.5 = sum((model.5$residuals)^2)
SSW.6 = sum((model.6$residuals)^2)
summary.2 = c(summary(model.2))
summary.3 = c(summary(model.3))
summary.4 = c(summary(model.4))

```

Lampiran 6 Syntax Metode SWFM (Lanjutan)

```

summary.5 = c(summary(model.5))
summary.6 = c(summary(model.6))
summary.22=matrix(unlist(summary.2),3,5)
summary.33=matrix(unlist(summary.3),3,5)
summary.44=matrix(unlist(summary.4),3,5)
summary.55=matrix(unlist(summary.5),3,5)
summary.66=matrix(unlist(summary.6),3,5)
SSB.2 = sum(summary.22[1:2,3])
SSB.3 = sum(summary.33[1:2,3])
SSB.4 = sum(summary.44[1:2,3])
SSB.5 = sum(summary.55[1:2,3])
SSB.6 = sum(summary.66[1:2,3])
SW.2 = sqrt(SSW.2/(38-2))
SW.3 = sqrt(SSW.3/(38-3))
SW.4 = sqrt(SSW.4/(38-4))
SW.5 = sqrt(SSW.5/(38-5))
SW.6 = sqrt(SSW.6/(38-6))
SB.2 = sqrt(SSB.2/(2-1))
SB.3 = sqrt(SSB.3/(3-1))
SB.4 = sqrt(SSB.4/(4-1))
SB.5 = sqrt(SSB.5/(5-1))
SB.6 = sqrt(SSB.6/(6-1))
Ratio.2 = SW.2 / SB.2
Ratio.3 = SW.3 / SB.3
Ratio.4 = SW.4 / SB.4
Ratio.5 = SW.5 / SB.5
Ratio.6 = SW.6 / SB.6
Ratio.SWFM = rbind (Ratio.2, Ratio.3, Ratio.4, Ratio.5, Ratio.6)
u = c(2,3,4,5,6)
z = Ratio.SWFM
plot(u, z, main = " Plot Nilai Ratio Ensemel SWFM", xlab = "Jumlah
Kluster", ylab = "Nilai Ratio")
lines(u,z,col = "Blue")
print("Hasil Pengelompokan SWFM")
print(hasil.kelompok)
print("-----")
print("Nilai Ratio untuk Setiap Nilai K")
print(Ratio.SWFM)
}

```

Lampiran 7 Output Statistika Deskriptif**#Data Numerik**

Variable	Mean	Minimum	Maximum
X1	0.9223	0.4608	1.0000
X2	0.005684	0.001000	0.030000
X3	0.002509	0.001383	0.004476
X4	98.953	81.040	100.000
X5	0.03691	0.00000	0.25472
X6	12.540	4.770	27.080
X7	6.719	4.445	13.888
X8	70.728	65.430	73.850
X9	92.301	77.930	98.520
X10	4.225	1.010	7.660

#Data Kategorik

Total	N for				
Variable	Count	Minimum	Maximum	Mode	Mode
X11	38	1.0000	2.0000	2	34
X12	38	1.000	6.000	2	24
X13	38	1.000	6.000	4, 5	12

**Lampiran 8 Output Analisis Faktor
KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.705
Bartlett's Test of Sphericity	173.305
df	45
Sig.	.000

Rotated Component Matrix^a

	Component		
	1	2	3
Aspal	.650	.104	.527
RS	.679	-.095	-.163
SMA_MA	-.149	-.857	-.071
Listrik_PLN	.090	.764	-.065
Lahan_kritis	-.248	-.047	.805
Presentase_penduduk_miskin	-.755	-.436	.054
Pengeluaran_penduduk_perkapita	.852	.106	-.284
AHH	.725	.354	-.031
AMH	.780	.414	-.122
Presentase_pengangguran	.437	.473	-.376

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.^a

a. Rotation converged in 4 iterations.

Lampiran 8 Output Analisis Faktor (Lanjutan)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings			Total Variance Explained
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	
1	4.542	45.420	45.420	4.542	45.420	45.420	3.596	35.960	35.960	
2	1.240	12.397	57.817	1.240	12.397	57.817	2.063	20.627	56.588	
3	1.079	10.794	68.611	1.079	10.794	68.611	1.202	12.023	68.611	
4	.767	7.673	76.284							
5	.711	7.107	83.392							
6	.628	6.280	89.671							
7	.436	4.364	94.035							
8	.291	2.908	96.943							
9	.216	2.165	99.108							
10	.089	.892	100.000							

Extraction Method: Principal Component Analysis.

Lampiran 9 Hasil Pengelompokan Metode Ward

#Dua Kelompok

[1] "Validasi Pengelompokan"

```
SST SSW.ward.c2 SSB.ward.c2 Rsquare.ward.c2 pseudoF.ward.c2
1 167.9427 50.66575 117.277 0.6983153 83.32988
```

[1] "Hasil Pengelompokan"

```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2 2 2 2 2 2 2
2 2
```

#Tiga Kelompok

[1] "Validasi Pengelompokan"

```
SST SSW.ward.c3 SSB.ward.c3 Rsquare.ward.c3 pseudoF.ward.c3
1 167.9427 17.50928 150.4334 0.8957425 150.3537
```

[1] "Hasil Pengelompokan"

```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2 2 3 2 2 2 2
3 2
```

#Empat Kelompok

[1] "Validasi Pengelompokan"

```
SST SSW.ward.c4 SSB.ward.c4 Rsquare.ward.c4 pseudoF.ward.c4
1 167.9427 10.95597 156.9867 0.9347637 162.394
```

[1] "Hasil Pengelompokan"

```
[1] 1 1 2 2 2 2 1 1 2 1 1 2 3 3 2 2 2 2 2 2 3 2 1 1 1 3 3 4 3 3 3 3
4 3
```

#Lima Kelompok

[1] "Validasi Pengelompokan"

```
SST SSW.ward.c5 SSB.ward.c5 Rsquare.ward.c5 pseudoF.ward.c5
1 167.9427 8.071463 159.8712 0.9519392 163.4075
```

[1] "Hasil Pengelompokan"

```
[1] 1 1 2 2 2 2 1 1 3 1 1 1 2 4 4 3 2 2 3 2 2 2 3 4 2 1 1 1 4 4 5 4 4 4
5 4
```

Lampiran 10 Anggota Kelompok Optimum Metode Ward

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

Lampiran 11 Anggota Kelompok Optimum *K-means*

Kab/Kota	Kelompok	Kab/Kota	Kelompok
Pacitan	3	Sumenep	3
Ponorogo	3	Kota Kediri	1
Trenggalek	3	Kota Blitar	4
Tulungagung	3	Kota Malang	5
Blitar	3	Kota Probolinggo	4
Kediri	3	Kota Pasuruan	1
Malang	3	Kota Mojokerto	4
Lumajang	3	Kota Madiun	4
Jember	3	Kota Surabaya	2
Banyuwangi	1	Kota Batu	4
Bondowoso	3		
Situbondo	3		
Probolinggo	3		
Pasuruan	3		
Sidoarjo	4		
Mojokerto	1		
Jombang	1		
Nganjuk	3		
Madiun	3		
Magetan	1		
Ngawi	3		
Bojonegoro	3		
Tuban	3		
Lamongan	1		
Gresik	1		

Lampiran 12 Output Pengelompokan Data Kategorik

	theta.	theta.	theta.	theta.	theta.
Kab/Kota	1	2	3	4	5
Kabupaten Pacitan	3	3	1	2	1
Kabupaten Ponorogo	1	3	3	3	2
Kabupaten Trenggalek	3	3	1	2	1
Kabupaten Tulungagung	1	3	3	3	1
Kabupaten Blitar	3	3	1	2	1
Kabupaten Kediri	3	3	3	3	2
Kabupaten Malang	3	3	3	3	1
Kabupaten Lumajang	2	3	3	3	1
Kabupaten Jember	2	3	3	3	1
Kabupaten Banyuwangi	1	3	1	3	1
Kabupaten Bondowoso	1	3	1	2	1
Kabupaten Situbondo	3	3	1	3	1
Kabupaten Probolinggo	1	3	3	3	2
Kabupaten Pasuruan	4	3	3	3	2
Kabupaten Sidoarjo	4	3	1	2	3
Kabupaten Mojokerto	4	3	3	3	2
Kabupaten Jombang	1	3	3	3	2
Kabupaten Nganjuk	3	3	3	3	2
Kabupaten Madiun	1	3	3	3	1
Kabupaten Magetan	2	3	1	2	1
Kabupaten Ngawi	2	3	1	3	1
Kabupaten Bojonegoro	3	3	3	3	2
Kabupaten Tuban	4	3	1	3	3

Lampiran 12 Output Pengelompokan Data Kategorik (Lanjutan)

Kabupaten Lamongan	4	3	3	3	3
Kabupaten Gresik	4	3	3	3	3
Kabupaten Bangkalan	1	3	1	2	1
Kabupaten Sampang	1	3	1	2	1
Kabupaten Pamekasan	1	3	3	3	1
Kabupaten Sumenep	2	3	1	2	1
Kota Kediri	2	3	3	3	2
Kota Blitar	2	3	1	3	1
Kota Malang	4	3	1	2	3
Kota Probolinggo	1	3	3	3	2
Kota Pasuruan	4	3	3	3	3
Kota Mojokerto	4	3	1	3	3
Kota Madiun	4	3	3	3	3
Kota Surabaya	4	3	3	3	<NA>
Kota Batu	2	3	3	3	1
	theta.6	theta.7	theta.8	theta.9	theta.10
1	1	1	1	1	1
2	1	2	1	1	1
3	1	1	1	1	1
4	1	2	1	1	1
5	1	1	1	1	1
6	1	2	1	1	1
7	1	1	1	1	1
8	1	2	1	1	1
9	1	2	1	1	1
10	1	2	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1

Lampiran 12 Output Pengelompokan Data Kategorik (Lanjutan)

13	1	2	1	1	1
14	1	2	1	1	1
15	1	3	1	1	1
16	1	2	1	1	1
17	1	2	1	1	1
18	1	2	1	1	1
19	1	1	1	1	1
20	1	1	1	1	1
21	1	2	1	1	1
22	1	2	1	1	1
23	1	3	1	1	1
24	1	3	1	1	1
25	<NA>	3	<NA>	<NA>	<NA>
26	1	1	1	1	1
27	1	1	1	1	1
28	1	2	1	1	1
29	1	1	1	1	1
30	1	2	1	1	1
31	1	2	1	1	1
32	1	3	1	1	1
33	1	2	1	1	1
34	1	3	1	1	1
35	1	3	1	1	1
36	1	3	1	1	1
37	<NA>	<NA>	<NA>	<NA>	<NA>
38	1	1	1	1	1

Lampiran 13 Output Pengelompokan Ensemble ROCK

data.Kab_Kota	theta.1	theta.2	theta.3	theta.4	theta.5
Kabupaten Pacitan	2	1	1	1	1
Kabupaten Ponorogo	2	1	1	1	1
Kabupaten Trenggalek	2	3	3	3	2
Kabupaten Tulungagung	1	3	3	2	1
Kabupaten Blitar	1	3	3	2	1
Kabupaten Kediri	1	3	3	2	1
Kabupaten Malang	1	3	3	2	1
Kabupaten Lumajang	2	1	1	1	1
Kabupaten Jember	2	1	1	1	1
Kabupaten Banyuwangi	1	1	1	1	1
Kabupaten Bondowoso	2	1	1	1	1
Kabupaten Situbondo	2	1	1	1	1
Kabupaten Probolinggo	2	1	1	1	1
Kabupaten Pasuruan	3	3	3	2	3
Kabupaten Sidoarjo	3	4	4	4	3
Kabupaten Mojokerto	3	4	4	4	3
Kabupaten Jombang	2	2	2	3	2
Kabupaten Nganjuk	2	3	3	3	2
Kabupaten Madiun	1	3	3	2	1
Kabupaten Magetan	2	2	2	3	2
Kabupaten Ngawi	1	3	3	2	1
Kabupaten Bojonegoro	1	3	3	2	1
Kabupaten Tuban	3	3	3	2	3

Lampiran 13 Output Pengelompokan Ensemble ROCK (Lanjutan)

Kabupaten Lamongan	3	4	4	4	3
Kabupaten Gresik	3	4	4	4	3
Kabupaten Bangkalan	1	3	3	2	1
Kabupaten Sampang	2	1	1	1	1
Kabupaten Pamekasan	2	1	1	1	1
Kabupaten Sumenep	2	1	1	1	1
Kota Kediri	2	2	2	3	2
Kota Blitar	1	1	1	1	1
Kota Malang	3	4	4	4	3
Kota Probolinggo	2	2	2	3	2
Kota Pasuruan	3	4	4	4	3
Kota Mojokerto	3	4	4	4	3
Kota Madiun	3	4	4	4	3
Kota Surabaya	3	4	4	4	3
Kota Batu	2	2	2	3	2
	theta.6	theta.7	theta.8	theta.9	theta.10
1	1	<NA>	<NA>	<NA>	<NA>
2	1	2	2	2	2
3	3	3	3	3	3
4	2	4	4	4	4
5	2	3	3	3	3
6	2	3	3	3	3
7	1	3	3	3	3
8	1	5	5	5	5
9	1	5	5	5	5
10	1	<NA>	<NA>	<NA>	<NA>
11	1	2	2	2	2
12	1	<NA>	<NA>	<NA>	<NA>
13	1	2	2	2	2

Lampiran 13 Output Pengelompokan Ensemble ROCK (lanjutan)

14	4	<NA>	<NA>	<NA>	<NA>
15	4	9	9	9	9
16	4	9	9	9	9
17	3	<NA>	<NA>	<NA>	<NA>
18	3	3	3	3	3
19	1	4	4	4	4
20	3	<NA>	<NA>	<NA>	<NA>
21	2	<NA>	<NA>	<NA>	<NA>
22	1	3	3	3	3
23	4	<NA>	<NA>	<NA>	<NA>
24	4	<NA>	<NA>	<NA>	<NA>
25	4	9	9	9	9
26	2	4	4	4	4
27	1	2	2	2	2
28	1	2	2	2	2
29	1	5	5	5	5
30	3	15	15	15	15
31	1	15	15	15	15
32	4	<NA>	<NA>	<NA>	<NA>
33	3	<NA>	<NA>	<NA>	<NA>
34	4	9	9	9	9
35	4	9	9	9	9
36	4	9	9	9	9
37	4	<NA>	<NA>	<NA>	<NA>
38	3	15	15	15	15

Lampiran 13 Output Pengelompokan Ensemble ROCK (Lanjutan)

theta.1 freq			theta.8 freq		
1	1	10	1	1	3
2	4	17	2	2	3
3	5	11	3	4	5
theta.2 freq			4	5	3
1	1	12	5	6	4
2	2	5	6	9	6
3	3	12	7	15	3
4	5	9	8	<NA>	11
theta.3 freq			theta.9 freq		
1	1	12	1	1	3
2	2	5	2	2	3
3	3	12	3	4	5
4	5	9	4	5	3
theta.4 freq			5	6	4
1	1	12	6	9	6
2	2	10	7	15	3
3	3	7	8	<NA>	11
4	5	9	theta.10 freq		
theta.5 freq			1	1	3
1	1	20	2	2	3
2	3	7	3	4	5
3	5	11	4	5	3
theta.6 freq			5	6	4
1	1	15	6	9	6
2	2	5	7	15	3
3	3	7	8	<NA>	11
4	5	11	theta.7 freq		
theta.7 freq			1	3	
1	1	3	2	3	
2	2	3	3	5	
3	4	5	4	3	
4	5	3	5	4	
5	6	4	6	4	
6	9	6	7	3	
7	15	3	8	<NA>	
8	<NA>	11			

Lampiran 14 Output Pengelompokan SWFM

Kab_Kota	hasil.k 2	hasil.k 3	hasil.k 4	hasil.k 5
Pacitan	1	1	1	1
Ponorogo	1	1	1	1
Trenggalek	1	1	1	1
Tulungagung	1	1	1	1
Blitar	1	1	1	1
Kediri	1	1	1	1
Malang	1	1	1	1
Lumajang	1	1	1	1
Jember	1	1	1	1
Banyuwangi	1	1	1	1
Bondowoso	1	1	1	1
Situbondo	1	1	1	1
Probolinggo	1	1	1	1
Pasuruan	1	1	1	1
Sidoarjo	1	1	1	1
Mojokerto	1	1	1	1
Jombang	1	1	1	2
Nganjuk	1	1	1	1
Madiun	1	1	1	1
Magetan	1	1	1	2
Ngawi	1	1	1	1
Bojonegoro	1	1	1	1
Tuban	1	1	1	1
Lamongan	1	1	2	3
Gresik	1	1	1	1
Bangkalan	1	1	1	1
Sampang	1	1	1	1

Lampiran 14 Output Pengelompokan SWFM (Lanjutan)

Kab_Kota	hasil.k2	hasil.k3	hasil.k4	hasil.k5
Pamekasan	1	1	1	1
Sumenep	1	1	1	1
Kota Kediri	1	1	2	3
Kota Blitar	1	2	3	4
Kota Malang	2	3	4	5
Kota Probolinggo	1	1	2	3
Kota Pasuruan	1	1	1	1
Kota Mojokerto	1	1	1	1
Kota Madiun	1	1	1	1
Kota Surabaya	2	3	4	5
Kota Batu	1	1	2	3

Lampiran 15 Output One-way MANOVA

Box's Test of Equality of Covariance Matrices^a	
Box's M	216.709
F	1.937
df1	55
df2	1562.986
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Klaster2

Multivariate Tests^a								
Effect		Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power ^d
Intercept	Pillai's Trace	1.000	18029.238 ^b	10.000	25.000	.000	180292.381	1.000
	Wilks'Lambda	.000	18029.238 ^b	10.000	25.000	.000	180292.381	1.000
	Hotelling's Trace	7211.655	18029.238 ^b	10.000	25.000	.000	180292.381	1.000
	Roy's Largest Root	7211.655	18029.238 ^b	10.000	25.000	.000	180292.381	1.000
Klaster2	Pillai's Trace	1.222	1.856	30.000	81.000	.015	55.669	.986
	Wilks'Lambda	.148	2.264	30.000	74.056	.002	65.937	.996
	Hotelling's Trace	3.487	2.751	30.000	71.000	.000	82.533	1.000
	Roy's Largest Root	2.786	7.523 ^c	10.000	27.000	.000	75.233	1.000

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Lampiran 16 Output ANOVA dan Uji Kruskal-Wallis**One-way ANOVA: X1 versus Klaster**

Source	DF	SS	MS	F	P
Klaster	3	0.0002318	0.0000773	3.20	0.035
Error	34	0.0008204	0.0000241		
Total	37	0.0010522			

S = 0.004912 R-Sq = 22.03% R-Sq(adj) = 15.15%

One-way ANOVA: X2 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	0.0002318	0.0000773	3.20	0.035
Error	34	0.0008204	0.0000241		
Total	37	0.0010522			

S = 0.004912 R-Sq = 22.03% R-Sq(adj) = 15.15%

One-way ANOVA: X3 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	0.0000034	0.0000011	2.59	0.069
Error	34	0.0000150	0.0000004		
Total	37	0.0000184			

S = 0.0006643 R-Sq = 18.62% R-Sq(adj) = 11.44%

One-way ANOVA: X4 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	35.99	12.00	1.29	0.293
Error	34	315.87	9.29		
Total	37	351.86			

S = 3.048 R-Sq = 10.23% R-Sq(adj) = 2.31%

One-way ANOVA: X5 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	0.01239	0.00413	1.99	0.134
Error	34	0.07068	0.00208		
Total	37	0.08308			

S = 0.04560 R-Sq = 14.92% R-Sq(adj) = 7.41%

One-way ANOVA: X6 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	329.7	109.9	5.54	0.003
Error	34	674.3	19.8		
Total	37	1004.0			

S = 4.453 R-Sq = 32.84% R-Sq(adj) = 26.91

Lampiran 16 Output ANOVA dan Uji Kruskal-Wallis (Lanjutan)**One-way ANOVA: X7 versus Klaster**

Source	DF	SS	MS	F	P
Klaster	3	97.03	32.34	15.52	0.000
Error	34	70.83	2.08		
Total	37	167.86			

S = 1.443 R-Sq = 57.80% R-Sq(adj) = 54.08%

One-way ANOVA: X8 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	72.71	24.24	8.93	0.000
Error	34	92.25	2.71		
Total	37	164.96			

S = 1.647 R-Sq = 44.08% R-Sq(adj) = 39.14%

One-way ANOVA: X9 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	416.0	138.7	10.43	0.000
Error	34	452.1	13.3		
Total	37	868.1			

S = 3.646 R-Sq = 47.92% R-Sq(adj) = 43.33%

One-way ANOVA: X10 versus Klaster

Source	DF	SS	MS	F	P
Klaster	3	21.66	7.22	3.11	0.039
Error	34	79.03	2.32		
Total	37	100.70			

S = 1.525 R-Sq = 21.51% R-Sq(adj) = 14.59%

Test Statistics^{a,b}

	X11	X12	X13
Chi-Square	6.847	11.662	20.237
df	3	3	3
Asymp. Sig.	.077	.009	.000

a. Kruskal Wallis Test

b. Grouping Variable: Klaster

Lampiran 17 Surat Keterangan Data Penelitian



SURAT KETERANGAN

Yang bertanda tangan dibawah ini :

N a m a : Thomas Wunang Tjahjo, M.Sc, M.Eng.

N I P : 19700329 1992 11 1 001

Jabatan : Kepala Bidang Integrasi Pengolahan dan
Diseminasi Statistik

Dengan ini menerangkan bahwa :

N a m a : Dwi Harid Setiadi

Fakultas/Program Studi : Fakultas Matematika, Komputasi dan Sains Data / Statistika

N.R.P : 0621144000066

Alamat Rumah : Dsn. Pandean, Desa Banjarkemantren RT 04 RW 01
Kecamatan Buduran, Kabupaten Sidoarjo

Akademi / Universitas : Institut Teknologi Sepuluh Nopember (ITS)
Telp (031) 594 3352, (031) 599 4251-55
Fax (031) 592 2940

Benar-benar telah mencari data di Kantor Badan Pusat Statistik (BPS) Provinsi Jawa Timur dalam rangka menyusun Tugas Akhir / Skripsi dengan judul :

*"Pemetaan Kabupaten/Kota di jawa Timur Berdasarkan Indikator Daerah Tertinggal
dengan Metode Data Campuran Ensemble ROCK dan SWFM"*

Demikian surat keterangan ini dibuat dan agar dipergunakan sebagaimana mestinya

Surabaya, 2 Mei 2018

An. Kepala BPS Provinsi Jawa Timur

Kepala Bidang IPDS

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



BIODATA PENULIS



Dwi Harid Setiadi atau yang akrab dipanggil Harid, merupakan anak kedua dari Bapak Asmu' id dan Ibu Hariningsih. Penulis lahir di Sidoarjo, 24 Desember 1995. Penulis menempuh Pendidikan formal di SDN Banjarkemantren I Sidoarjo (2002-2008) , SMPN 5 Sidoarjo (2008-2011), SMAN 3 Sidoarjo (2011-2014). Penulis diterima di Departemen Statistika FMKSD ITS pada tahun 2014. Selama masa kuliah, penulis aktif pada organisasi kemahasiswaan tingkat departemen dengan menjadi staff Departemen Kesejahteraan Mahasiswa HIMASTA-ITS 2015/2016. Kemudian berlanjut menjadi Kabiro Akademik dan Apresiasi HIMAsta-ITS 2016/2017. Selain itu, penulis juga berpartisipasi dalam beberapa kepanitiaan yaitu OC Gerigi ITS 2015, Pekan Raya Statistika (2016), Ketua Panitia *Job Preparation Training* (JPT) 2015 dan lain-lain. Dalam bidang akademik, penulis pernah menjadi finalis dalam Pekan Analisis Statistik, Universitas Mulawarman pada tahun 2018. Kritik dan saran terkait Tugas Akhir ini dapat diberikan kepada penulis melalui email hariddwi@gmail.com atau telfon pada 089616729594.