



TUGAS AKHIR – SS141501

**PENGELOMPOKAN SEKOLAH MENENGAH ATAS
(SMA) DI SIDOARJO MENGGUNAKAN *SIMILARITY
WEIGHT AND FILTER METHOD* (SWFM)**

**ARYO JOKO PRAKOSO
NRP 1313100059**

**Dosen Pembimbing
Dr. Dra. Ismaini Zain, M.Si
Erma Oktania Permatasari, S.Si, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



TUGAS AKHIR - SS 141501

**PENGELOMPOKAN SEKOLAH MENENGAH
ATAS (SMA) DI SIDOARJO MENGGUNAKAN
SIMILARITY WEIGHT AND FILTER METHOD
(SWFM)**

**ARYO JOKO PRAKOSO
NRP 1313100059**

**Dosen Pembimbing
Dr. Dra. Ismaini Zain, M.Si
Erma Oktania Permatasari, S.Si, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**

(halaman ini sengaja dikosongkan)



FINAL PROJECT - SS 141501

**CLUSTERING NUMERICAL AND CATEGORICAL
DATA SENIOR HIGH SCHOOL (SMA) AT
SIDOARJO USING *SIMILARITY WEIGHT AND
FILTER METHOD (SWFM)***

**ARYO JOKO PRAKOSO
NRP 1313100059**

**Supervisor
Dr. Dra. Ismaini Zain, M.Si
Erma Oktania Permatasari, S.Si, M.Si**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**

(halaman ini sengaja dikosongkan)

LEMBAR PENGESAHAN

PENGELOMPOKAN SEKOLAH MENENGAH ATAS
(SMA) DI SIDOARJO MENGGUNAKAN *SIMILARITY
WEIGHT AND FILTER METHOD (SWFM)*

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :
Aryo Joko Prakoso
NRP. 1313 100 059

Disetujui oleh Pembimbing:
Dr. Dra. Ismaini Zain, M.Si
NIP. 19600525 198803 2 001
Erma Oktania P. S. S., M.Si
NIP. 19881007 200903 2 002



Mengetahui,
Kepala Departemen



DEPARTEMEN
STATISTIKA

Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, JULI 2017

(halaman ini sengaja dikosongkan)

PENGELOMPOKAN SEKOLAH MENENGAH ATAS (SMA) DI SIDOARJO MENGGUNAKAN *SIMILARITY WEIGHT AND FILTER METHOD* (SWFM)

Nama Mahasiswa : Aryo Joko Prakoso
NRP : 1313 100 059
Jurusan : Statistika
Dosen Pembimbing 1 : Dr. Dra. Ismaini Zain, M.Si
Dosen Pembimbing 2 : Erma Oktania P., S.Si., M.Si

Abstrak

Pendidikan merupakan faktor yang sangat penting dalam kemajuan suatu negara. Sektor pendidikan ini harus menjadi fokus utama dalam pembangunan. Salah satu tahapan dalam pendidikan formal adalah Sekolah Menengah Atas (SMA). Pada jenjang SMA inilah anak akan mulai mempersiapkan diri dalam bidang dunia kerja maupun lingkungan yang lebih luas, sehingga kompetensi SMA perlu mendapat perhatian guna mencetak lulusan yang terbaik. SMA yang difokuskan pada penelitian ini adalah SMA di kabupaten Sidoarjo. Kabupaten Sidoarjo memiliki jumlah SMA terbanyak ketiga setelah kabupaten Lamongan, dan Malang. Kabupaten Sidoarjo menduduki peringkat pertama pada rata-rata nilai UN di Jawa Timur untuk jurusan IPA maupun IPS. Aspek pendidikan SMA yang diteliti adalah rasio siswa per rombel, jumlah guru tetap, jumlah guru non-tetap, jumlah lulusan per kelas, luas lahan, daya listrik, status sekolah, akreditasi, dan waktu penyelenggaraan. Penelitian ini dilakukan guna mengelompokkan SMA-SMA di kabupaten Sidoarjo dengan jumlah SMA sebanyak 51 menggunakan similarity weight and filter method (SWFM). Metode ini dapat menggabungkan kelompok data yang berbeda (nominal, ordinal, interval, dan rasio). SWFM mempunyai keunggulan sebagai metode pada campuran variabel numerik dan kategorik dengan tidak mentransformasi ke variabel numerik, pemilihan metode disesuaikan dengan karakteristik data yang berukuran kecil dengan melibatkan seluruh karakteristik. Pada analisis kluster data numerik, diperoleh empat jumlah kelompok optimum

menggunakan metode hierarki single linkage dengan nilai R-square 92,48% dan jumlah masing-masing kelompok berturut-turut sebesar 17 SMA, 31 SMA, 2 SMA, dan 1 SMA. Sedangkan analisis klaster data ketegorik menggunakan metode non-hierarki k-modes dihasilkan dua kelompok dengan kelompok pertama sebanyak 46 SMA dan kelompok kedua 5 SMA. Metode ensemble SWFM menghasilkan nilai rasio terkecil pada jumlah kelompok sebanyak tiga sebesar 0,3329.

Kata kunci: Analisis Klaster, K-Modes, Sekolah Menengah Atas (SMA), Similarity Weight and Filter Method

CLUSTERING DATA SENIOR HIGH SCHOOL IN SIDOARJO USING SIMILARITY WEIGHT AND FILTER METHOD (SWFM)

Name : Aryo Joko Prakoso
NRP : 1313 100 059
Department : Statistics
Supervisor 1 : Dr. Dra. Ismaini Zain, M.Si
Supervisor 2 : Erma Oktania P., S.Si., M.Si

Abstract

Education is one of important factor growth in a country. Main focus of development should be in education sector. One of the stages in formal education is Senior High School (SMA). At this stage of senior high school, student will begin to prepare themselves in the jobs environment, so the quality of senior high school needs attention in order to make the best graduates. Object that focused on this research is SMA in Sidoarjo district. The third highest number of schools in East Java is Sidoarjo after Lamongan, and Malang. And Sidoarjo has the highest on the average of UN in East Java for science and social science majors. Variable of senior high school are ratio of students per study group, number of permanent teachers, number of non-permanent teachers, number of graduates per class, land area, electric power, school status, accreditation, and time of operation. In this research was conducted to classify SMA in Sidoarjo with 51 SMA amount using similarity weight and filter method (SWFM). This method can combine different data groups (nominal, ordinal, interval, and ratio). SWFM has the advantage of a method on a mix of numerical and categorical variables without transforming into numerical variables, the method is adjusted to the characteristics of small data by involving all the characteristics. In the numerical data cluster analysis, there were four optimum groups using a single linkage hierarchy method with R-square 92.48% and the number of each group was 17 SMA, 31 SMA, 2 SMA, and 1 SMA

respectively. While the categorical data cluster analysis using non-hierarchy k-modes method produced two groups with the first group of 46 SMA and the second group of 5 SMA. The ensemble SWFM method yields the smallest value of the ratio in the number of groups of three by 0.3329.

Keywords : Cluster Analysis, K-Modes, Senior High School, Similarity Weight and Filter Method.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT, atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul **“Pengelompokan Sekolah Menengah Atas (SMA) Di Sidoarjo Menggunakan *Similarity Weight and Filter Method (SWFM)*”** dengan baik. Penyusunan Tugas Akhir ini tidak luput dari bantuan serta dukungan dari berbagai pihak. Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada :

1. Ibu Wiji Asri, terima kasih atas doa, nasihat, dan kasih sayang yang sangat besar yang telah diberikan untuk penulis sehingga dapat menjadi motivasi dan penyemangat bagi penulis.
2. Bapak Dr. Suhartono selaku Kepala Departemen Statistika ITS dan Bapak Dr. Sutikno, M.Si selaku Ketua Program Studi Sarjana Jurusan Statistika ITS yang telah menyediakan fasilitas guna kelancaran pengerjaan Tugas Akhir ini.
3. Ibu Dr. Dra. Ismaini Zain, M.Si dan Ibu Erma Oktania Permatasari, S.Si, M.Si selaku dosen pembimbing yang telah memberikan ilmu, dan nasihat yang berharga bagi penulis serta kesabaran dan kebaikan untuk membimbing dan selalu memberikan masukan dalam penyelesaian Tugas Akhir.
4. Bapak R. Mohamad Atok, S.Si., M.Si dan Ibu Dr. Vita Ratnasari, S.Si., M.Si selaku dosen penguji yang telah memberikan kritikan dan saran demi kesempurnaan Tugas Akhir.
5. Teman-teman pejuang wisuda 116 atas semangat yang diberikan kepada penulis dan teman-teman angkatan 2013 atas segala motivasi dan semangatnya

Semoga Tugas Akhir ini dapat bermanfaat bagi penelitian selanjutnya.

Surabaya, September 2017

Penulis

(halaman ini sengaja dikosongkan)

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
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 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	
2.1 Data Numerik dan Kategorik.....	5
2.2 Analisis Multivariat.....	5
2.3 Analisis Klaster.....	5
2.4 Pengelompokan Data Numerik.....	7
2.5 Pengelompokan Data Kategorik.....	9
2.6 Pengelompokan Data Campuran.....	11
2.7 Pengelompokan SWFM.....	12
2.8 Kinerja Hasil Pengelompokan.....	14
2.9 Sekolah Menengah Atas.....	15
BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data.....	17
3.2 Variabel Penelitian.....	17
3.3 Struktur Data.....	18
3.4 Langkah Analisis.....	18
BAB IV ANALISIS DAN PEMBAHASAN	
4.1 Analisis Statistika Deskriptif.....	21
4.1.1 Data Numerik.....	21
4.1.2 Data Kategorik.....	23
4.2 Hasil Pengelompokan dengan Metode SWFM.....	27

4.2.1	Pengelompokan pada Data Numerik	27	
4.2.2	Pengelompokan pada Data Kategorik	31	
4.2.3	Pengelompokan pada Data Campuran.....	33	
BAB V KESIMPULAN DAN SARAN			
5.1	Kesimpulan.....	39	
5.2	Saran.....	40	
DAFTAR PUSTAKA			41
LAMPIRAN.....			43

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Kerangka Analisis Algoritma	11
Gambar 2.2 Kerangka Analisis Berdasarkan <i>Similarity Weight and Filter Method</i>	12
Gambar 4.1 Karakteristik Data Variabel Numerik	23
Gambar 4.2 Histogram Variabel Akreditasi	23
Gambar 4.3 Histogram Variabel Status Sekolah	24
Gambar 4.4 Histogram Variabel Waktu Penyelenggaraan	24
Gambar 4.5 Nilai $R_{\text{-square}}$ dari Metode <i>Single Linkage, Complete Linkage</i> , dan <i>Average Linkage</i>	28
Gambar 4.6 Karakteristik Pengelompokan Data Numerik pada Kelompok I.....	31
Gambar 4.7 Plot Nilai Rasio SWFM.....	34
Gambar 4.8 Karakteristik Hasil Pengelompokan SWFM pada Data Numerik Kelompok I	37
Gambar 4.9 Karakteristik Hasil Pengelompokan SWFM pada Data Numerik Kelompok II	37

(halaman ini sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1	Tabel Kontingensi Data Biner 6
Tabel 2.2	Contoh <i>Matriks Similarity</i> 13
Tabel 3.1	Variabel Penelitian 17
Tabel 3.2	Struktur Data..... 18
Tabel 4.1	Statistika Deskriptif Variabel Numerik 21
Tabel 4.2	Crosstab Akreditasi dengan Status Sekolah 25
Tabel 4.3	Crosstab Akreditasi dengan Penyelenggaraan 26
Tabel 4.4	Crosstab Status Sekolah dengan Waktu Penyelenggaraan 26
Tabel 4.5	Rasio Hasil Pengelompokan pada 4 Kelompok..... 29
Tabel 4.6	Karakteristik Hasil Pengelompokan Data Numerik30
Tabel 4.7	Karakteristik Variabel Hasil Pengelompokkan Data Kategorik 33
Tabel 4.8	Nilai Rasio Pengelompokan Metode SWFM 34
Tabel 4.9	Hasil Pengelompokan Data Numerik Rata-Rata Variabel pada Metode SWFM..... 29
Tabel 4.10	Karakteristik Variabel Kategorik Hasil Pengelompokkan pada Metode SWFM..... 29

(halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Pengamatan Variabel Numerik	43
Lampiran 2. Data Pengamatan Variabel Kategorik.....	44
Lampiran 3. <i>Syntax</i> Analisis Kelompok Data Numerik dengan Jumlah Kelompok 2	45
Lampiran 4. <i>Syntax</i> Analisis Kelompok Data Numerik dengan Jumlah Kelompok 3	49
Lampiran 5. <i>Syntax</i> Analisis Kelompok Data Numerik dengan Jumlah Kelompok 4	54
Lampiran 6. <i>Syntax</i> Analisis Kelompok Data Numerik dengan Jumlah Kelompok 5	59
Lampiran 7. <i>Syntax</i> Analisis Kelompok Data Numerik dengan Jumlah Kelompok 6	65
Lampiran 8. <i>Syntax</i> Nilai Rasio pada Analisis Kelompok Data Numerik dengan Jumlah Kelompok 4.....	72
Lampiran 9. <i>Syntax</i> Analisis Pengelompokan Data Kategorik dengan <i>k-modes</i>	75
Lampiran 10. <i>Syntax</i> Analisis Pengelompokan Data Campuran dengan Ensemble SWFM	76
Lampiran 11. <i>Output</i> untuk Analisis Kelompok pada Data Numerik dengan Jumlah Kelompok 2.....	81
Lampiran 12. <i>Output</i> untuk Analisis Kelompok pada Data Numerik dengan Jumlah Kelompok 3.....	83
Lampiran 13. <i>Output</i> untuk Analisis Kelompok pada Data Numerik dengan Jumlah Kelompok 3.....	85
Lampiran 14. <i>Output</i> untuk Analisis Kelompok pada Data Numerik dengan Jumlah Kelompok 4.....	87
Lampiran 15. <i>Output</i> untuk Analisis Kelompok pada Data Numerik dengan Jumlah Kelompok 5.....	89

Lampiran 16. <i>Output</i> untuk Analisis Kelompok pada Data Kategorik dengan Jumlah Kelompok 2	91
Lampiran 17. Hasil Pengelompokan Terbaik dari Data Numerik dan Kategorik	92
Lampiran 18. <i>Output</i> untuk Analisis Kelompok pada Data Campuran dengan Ensemble SWFM	93

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan hal wajib yang seharusnya diperoleh semua orang. Hakikat pendidikan berfungsi untuk meningkatkan pengetahuan dan kompetensi setiap individu. Berdasarkan KBBI, pendidikan yaitu sebuah proses pembelajaran bagi setiap individu untuk mencapai pengetahuan dan pemahaman yang lebih tinggi mengenai suatu objek dan spesifikasi tertentu. Pendidikan merupakan upaya secara terencana dalam mewujudkan proses pembelajaran supaya peserta didik dapat mengembangkan potensi diri dalam bidang keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak, dan ketrampilan (UU Nomor 20 thn 2003). Kementerian Pendidikan dan Kebudayaan (Kemendikbud) mewajibkan pendidikan di Indonesia selama 12 tahun dari SD, SMP, dan SMA. SMA merupakan salah satu Pendidikan untuk meningkatkan kemampuan individu peserta didik sesuai tujuan sehingga diharapkan mutu dari SMA harus sebaik mungkin dalam menghasilkan lulusan yang terbaik.

Pada tahun 2016, Jawa Timur merupakan provinsi dengan jumlah SMA terbanyak kedua di Indonesia setelah provinsi Jawa Barat dengan jumlah total 1.446 SMA dengan rincian SMA negeri sebanyak 420 dan SMA swasta sebanyak 1026 (Kemendikbud, 2013). Provinsi Jawa Timur memiliki 29 kabupaten dan 9 kota. Berdasarkan Dinas Pendidikan Provinsi Jawa Timur tahun 2015, Surabaya merupakan kota dengan jumlah SMA terbanyak yaitu 168 SMA, sedangkan untuk kabupaten dengan SMA terbanyak adalah kabupaten Lamongan, Malang, dan Sidoarjo dengan jumlah masing-masing 67, 64, dan 64 SMA. Pencapaian rata-rata hasil UN secara nasional mengalami penurunan, termasuk pada provinsi Jawa Timur. Faktor yang mempengaruhi turunnya nilai UN adalah bobot soal yang semakin tinggi dan banyak ujian nasional berbasis komputer (Kemendikbud, 2016). Meskipun perolehan nilai UN di Jawa Timur mengalami penurunan, Kabupaten Sidoarjo menduduki peringkat pertama pada rata-rata nilai UN di Jawa Timur untuk jurusan IPA maupun IPS dengan nilai masing-masing

50,66 dan 47,61. Bahkan pada tahun 2014 kabupaten Sidoarjo memiliki tingkat kelulusan pada UN SMA 100%.

Kabupaten Sidoarjo memiliki 64 SMA yang tersebar pada 18 kecamatan. Penduduk usia 16-18 tahun di kabupaten Sidoarjo sebanyak 97.286 orang dengan kepadatan penduduk sebesar 2.869/km (Dispendik Sidoarjo, 2016). Angka Partisipasi Murni (APM) dari anak usia SMA pada kabupaten Sidoarjo pada tahun 2014 sekitar 61,34% yang berarti perbandingan jumlah anak usia 16 sampai 18 tahun yang bersekolah di SMA dengan jumlah anak usia 16 sampai 18 tahun hanya 61,34%.

Pada penelitian ini dilakukan pengelompokan SMA di kabupaten Sidoarjo. Langkah yang dapat dilakukan untuk dapat mengelompokkan SMA adalah dengan mengetahui faktor-faktor dan mengklasifikasikan faktor-faktor dengan menggunakan algoritma tertentu. Metode yang akan digunakan dalam penelitian ini menggunakan menggunakan *Similarity Weight and Filter Method* (SWFM). Metode ini dapat menggabungkan kelompok data yang berbeda (nominal, ordinal, interval, dan rasio). Terdapat beberapa metode untuk mengelompokkan data campuran antara lain menggunakan *k-prototype*, ensemble ROCK, dan SWFM. SWFM mempunyai keunggulan sebagai metode pada campuran variabel numerik dan kategorik dengan tidak mentransformasi ke variabel numerik, pemilihan metode disesuaikan dengan karakteristik data yang berukuran kecil dengan melibatkan seluruh karakteristik. Variabel yang digunakan pada penelitian ini adalah variabel yang bersifat kategorik dan numerik. Variabel yang bersifat kategorik antara lain akreditasi SMA, status SMA, dan waktu penyelenggaraan sedangkan variabel yang bersifat numerik antara lain rasio siswa/rombel, jumlah guru tetap, jumlah guru non-tetap, jumlah lulusan/kelas, luas lahan, dan daya listrik. Penelitian mengenai pendidikan di kabupaten Sidoarjo pernah dilakukan oleh Henlita dan Handayani (2013) untuk menganalisis tingkat pelayanan pendidikan SMA namun penelitian yang dilakukan hanya sebatas analisa statistika deskriptif dengan teknik analisis daya tampung dan analisis distribusi frekuensi relatif. Prayoga (2015) mengelompokkan kecamatan di Sidoarjo berdasarkan indikator mutu pendidikan jenjang pendidikan dasar. Sedangkan

untuk teknik pengelompokan dengan metode SWFM pernah dilakukan oleh Alvionita (2017) mengelompokkan akses jeruk hasil fusi protoplasma menggunakan metode ensemble ROCK dan ensemble SWFM. Reddy dan Kavitha (2012) mengelompokkan data campuran numerik dan kategorik menggunakan SWFM.

1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan yang telah diuraikan, dirumuskan permasalahan penelitian sebagai berikut.

1. Bagaimana karakteristik data faktor-faktor sarana prasarana dan SDM pada SMA di Kabupaten Sidoarjo?
2. Bagaimana pengelompokan menggunakan *Similarity Weight and Filter Method* (SWFM) pada faktor-faktor sarana prasarana dan SDM pada SMA di Kabupaten Sidoarjo?

1.3 Tujuan

Tujuan yang ingin dicapai dari penelitian, antara lain sebagai berikut.

1. Mendeskripsikan karakteristik data faktor-faktor sarana prasarana dan SDM pada SMA di Kabupaten Sidoarjo.
2. Mengelompokkan faktor-faktor sarana prasarana dan SDM pada SMA di Kabupaten Sidoarjo menggunakan *Similarity Weight and Filter Method* (SWFM).

1.4 Manfaat

Manfaat yang didapatkan dari penelitian ini adalah mampu memberi saran kepada Dinas Pendidikan Provinsi Jawa Timur wilayah kabupaten Sidoarjo maupun SMA dalam kebijakan dibidang sarana prasarana dan SDM. Selain itu, diharapkan masyarakat memahami kondisi SMA yang ada di Kabupaten Sidoarjo saat ini. Dari hasil analisis dalam penelitian dapat dijadikan acuan bagi penelitian selanjutnya yang sejenis yang berhubungan dengan faktor-faktor sarana prasarana dan SDM SMA di Kabupaten Sidoarjo.

1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah data SMA di kabupaten Sidoarjo pada tahun ajaran 2015/2016 dengan metode pengelompokan hierarki menggunakan *single linkage*, *complete linkage*, dan *average linkage* pada data yang

bersifat numerik. Pada data kategorik analisis yang digunakan menggunakan nilai k sebesar 2. Pada 13 SMA terdapat data missing sehingga tidak dapat dimasukkan dalam analisis. 13 SMA yang dimaksud antara lain SMA Progresif Bumi Shalawat, SMAS R Rahmat, SMA Hang Tuah 5, SMA Al Muslim, SMAS Dharma Wanita 6, SMAN 2 Sidoarjo, SMAS Al Islam Krian, SMA Pembangunan Jaya 2, SMAS YPM 2 Panjunan, SMAN 3 Sidoarjo, SMAS Plus Dharma Siswa, SMAS Hang Tuah 2 Gedangan, dan SMAS Darussalamah.

BAB II

TINJAUAN PUSTAKA

2.1 Data Numerik dan Kategorik

Data merupakan segala fakta dan angka yang dapat dijadikan bahan dalam menyusun suatu informasi untuk keperluan tertentu (Arikunto, 2002). Data berdasarkan jenis variabelnya dapat diklasifikasikan menjadi dua, yaitu data kategorik dan data numerik (Anderson & Sclove, 1974). Data numerik merupakan data yang bernilai angka sebenarnya, hasil dari pengukuran maupun pencacahan. Sedangkan data kategorik merupakan data yang bersifat kualitatif dari hasil penggolongan (Agresi, 2007).

2.2 Analisis Multivariat

Analisis multivariat merupakan teknik statistik yang dilakukan pada data dengan jumlah variabel lebih dari dua yang dianalisa secara simultan pada masing-masing pengamatan (Johnson & Wichern, 2007). Data pada analisis multivariat berupa beberapa variabel terikat dan beberapa variabel bebas dengan memperlakukan variabel yang saling berkorelasi sebagai suatu sistem dengan mempertimbangkan korelasi antar variabel-variabel (Suryanto, 1988).

2.3 Analisis Klaster

Metode klasifikasi merupakan metode penganalisis data yang bertujuan untuk mengelompokkan data dengan karakteristik yang sama ke suatu kelompok yang sama dengan karakteristik yang berbeda dengan kelompok lain (Rokach, 2010). Ada beberapa pendekatan dalam metode klasifikasi antara lain dengan pendekatan partisi dan dengan pendekatan hirarki. Klasifikasi dengan pendekatan partisi atau sering disebut dengan *partition-based clustering* mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam klaster-klaster yang ada. Klasifikasi dengan pendekatan hirarki atau sering disebut dengan *hierarchical clustering* mengelompokkan data dengan membuat suatu hirarki berupa *dendogram* dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan (Oliveira, 2007).

Metode *Clustering* merupakan salah satu teknik untuk mengelompokkan data. Metode ini bertujuan untuk mengelompokkan pengamatan-pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimiliki sehingga pengamatan yang paling memiliki kemiripan dengan pengamatan lain berdekatan (Johnson & Wichern, 2007). Hasil metode *clustering* dipengaruhi oleh pengamatan yang dikelompokkan, variabel yang digunakan, ukuran kemiripan, skala ukur, dan metode pengelompokan yang digunakan.

Algoritma pengelompokan menggunakan ukuran kemiripan untuk menggabungkan objek dari suatu data. Ukuran kemiripan biasanya digunakan untuk data yang bersifat kategorik, sedangkan ukuran ketidakmiripan digunakan untuk data yang bersifat numerik. Ukuran ketidakmiripan antara objek ke- i dengan objek ke- j (d_{ij}), memiliki sifat sebagaimana pada (2.1)

$$d_{ij} \geq 0, d_{ii} = 0, d_{ij} = d_{ji}, \text{ dan } d_{ik} + d_{jk} \geq d_{ij} \quad (2.1)$$

Untuk setiap i, j, k . Semakin besar ukuran ketidakmiripan antara dua objek, maka semakin besar pula perbedaan antara kedua objek tersebut, sehingga tidak berada dalam kelompok yang sama (Johnson & Wichern, 2007). Salah satu faktor yang berpengaruh terhadap hasil dari kelompok yang dibentuk adalah jarak antar objek pengamatan (Sharma, 1996). Metode yang digunakan untuk mengukur jarak antara objek ke- i (x_i) dengan objek ke- j (x_j) berdasarkan karakteristik variabel antara lain sebagai berikut.

a. Tabel kontingensi

Tabel kontingensi digunakan ketika variabel yang diamati berupa variabel biner yang memiliki dua karakter berbeda (0,1). Perhitungan ukuran jarak antara variabel x_i dan x_j sebagai berikut

Tabel 2.1 Tabel kontingensi data biner

Kategori x_i	Kategori x_j		Total
	0	1	
0	a	b	$a + b$
1	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

b. Jarak Euclidean

Metode pengukuran jarak untuk variabel numerik antara lain *euclidean*, *manhattan*, dan *minkowski*. Pada penelitian ini menggunakan jarak *Euclidean* dikarenakan metode ini relatif mudah dimengerti dan dapat digunakan pada data yang memiliki jumlah variabel lebih dari dua. Jarak *euclidean* digunakan pada data yang berskala numerik dengan variabel berdimensi m yaitu $X_i = |x_1, x_2, x_3, \dots, x_m|^T$ dan $X_j = |x_1, x_2, x_3, \dots, x_m|^T$. Rumus jarak Euclidean antara X_i dan X_j sebagaimana persamaan (2.2)

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2} \quad (2.2)$$

$$i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}) \quad \text{dan} \quad j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jm})$$

dengan jumlah variabel sebanyak m . Semakin kecil jarak antara kedua objek, mengindikasikan semakin besar kedekatan antara kedua objek tersebut.

2.4 Pengelompokan Data Numerik

Pengelompokan data numerik dilakukan berdasarkan ukuran ketidakmiripan berdasarkan data numerik. Hasil pengelompokan dapat berupa dendogram yang menyajikan objek-objek pengamatan lebih mudah. Teknik pengelompokan meliputi metode hierarki dan non-hierarki.

Pengelompokan hierarki dilakukan pada dua atau lebih objek yang memiliki kesamaan terdekat, selanjutnya proses dilakukan pada objek lain hingga kelompok membentuk semacam pohon dimana tingkatan antar objek lebih jelas dari yang terdekat hingga terjauh (Hair, *et al*, 2009). Teknik pengelompokan pada metode hierarki menggunakan teknik penggabungan (*agglomerative*) dan teknik pembagian (*divisive*). Teknik pengelompokan antar kelompok adalah sebagai berikut (Johnson, 2007).

a. Pautan tunggal (*Single Linkage*)

Teknik ini didasarkan pada jarak terkecil atau terdekat. Jika dua objek dipisahkan oleh jarak yang terpendek, maka dua objek tersebut digabungkan menjadi satu kelompok. Rumus yang digunakan sebagaimana persamaan (2.3) berikut

$$d_{w(i,j)} = \min\{d_{wi}, d_{wj}\} \quad (2.3)$$

b. Pautan Lengkap (*Complete Linkage*)

Teknik ini didasarkan pada jarak maksimum atau terjauh (Johnson, 2007). Rumus yang digunakan sebagaimana persamaan (2.4)

$$d_{w(i,j)} = \max\{d_{wi}, d_{wj}\} \quad (2.4)$$

c. Pautan Rataan (*Average Linkage*)

Teknik ini menggunakan jarak rata-rata sebagai alat pengelompokan dengan menggabungkan pasangan kelompok dengan nilai varian terkecil. Rumus yang digunakan sebagaimana persamaan (2.5)

$$d_{w(i,j)} = \frac{n_i}{n_i + n_j} d_{wi} + \frac{n_j}{n_i + n_j} d_{wj} \quad (2.5)$$

$d_{w(i,j)}$ merupakan ukuran kemiripan antara kelompok ke- w dengan kelompok (i, j) dengan penggabungan antara kelompok ke- i dan ke- j . n_i dan n_j merupakan jumlah pengamatan dalam kelompok ke- i dan ke- j .

Pada ketiga metode pengelompokan hierarki menggunakan perhitungan jarak euclidean. Metode hierarki mempercepat pengolahan data karena hasil yang terbentuk berupa tingkatan (pohon) sehingga mudah dalam memahami namun metode ini sangat lemah pada data yang terdapat *outlier* dan data yang memiliki variabel yang tidak sesuai. Validasi pengelompokan sesuai jumlah kelompok optimum yang ditentukan (Halkidi, et. al., 2001). Salah satu cara yang digunakan untuk menentukan jumlah kelompok optimum adalah menggunakan *r-squared* (Sharma, 1996). *R-squared* menghitung keragaman total dan keragaman antar kelompok. Adapun rumus yang digunakan sebagaimana persamaan berikut

Sum of Square Between Group (SSB):

$$SSB = SST - SSW \quad (2.6)$$

Sum of Square Total (SST):

$$SST = \sum_{l=1}^m \sum_{i=1}^n (x_{il} - \bar{x}_l)^2 \quad (2.7)$$

Sum of Square Within Group (SSW):

$$SSW = \sum_{c=1}^C \sum_{l=1}^m \sum_{i=1}^{n_c} (x_{ilc} - \bar{x}_{lc})^2 \quad (2.8)$$

Dimana,

m : Jumlah variabel dalam pengamatan

C : Jumlah kelompok yang dibentuk dalam pengamatan

n : Total objek pengamatan

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

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

Nilai $r\text{-squared}$ dapat digunakan sebagai ukuran untuk membedakan antar kelompok, dengan nilai $r\text{-squared}$ berkisar antara 0 sampai 1. Ketika nilai $r\text{-squared}$ bernilai 0 maka tidak terdapat perbedaan yang cukup signifikan antar kelompok, sedangkan ketika nilai $r\text{-squared}$ bernilai 1 maka menunjukkan perbedaan yang signifikan antar kelompok yang terbentuk. Rumus $r\text{-squared}$ sebagaimana pada persamaan (2.10)

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

2.5 Pengelompokan Data Kategorik

Pengelompokan data bersifat kategorik dapat dilakukan dengan ukuran kemiripan data dengan menggunakan metode hierarki dan non hierarki. $K\text{-modes}$ merupakan kluster algoritma dengan data bersifat kategorik. $K\text{-modes}$ tepat digunakan pada kelompok data yang bersifat kategorik. Berdasarkan Kaufman dan Rousseeuw (1990), andaikan X dan Y adalah dua data dengan fitur bertipe kategorikal. Ukuran ketidakmiripan di antara X dan Y dapat diukur dengan jumlah ketidakcocokan nilai dari fitur yang berkorespondensi dari dua data. Semakin kecil nilai ketidakcocokan, maka semakin mirip dua data tersebut. Berikut

merupakan rumus yang digunakan sebagaimana pada persamaan (2.10)

$$d(X, Y) = \sum_{j=1}^r \delta(x_j, y_j) \quad (2.10)$$

Dimana r adalah jumlah variabel, sedangkan (x_j, y_j) merupakan nilai pencocokan sebagaimana pada persamaan (2.12)

$$\delta(x_j, y_j) = \begin{cases} 0 & (x_j = y_j) \\ 1 & (x_j \neq y_j) \end{cases} \quad (2.12)$$

Misalkan X merupakan kelompok data dengan variabel bersifat kategorik $A_1, A_2, A_3, \dots, A_n$ maka modus dari $X = (X_1, X_2, X_3, \dots, X_n)$ adalah data $Q = (q_1, q_2, q_3, \dots, q_n)$ dengan nilai minimal sebagaimana pada persamaan (2.13)

$$D(X, Y) = \sum_{i=1}^n d(x_i, Q) \quad (2.13)$$

Vektor Q merupakan vektor diluar bagian dari X . Misal $n_{c_{kj}}$ adalah jumlah objek yang dimiliki oleh kategori c_{kj} ke- k pada atribut A_j dan $f(A_j = c_{kj}|X) = \frac{n_{c_{kj}}}{n}$ adalah frekuensi relatif kategori $n_{c_{kj}}$ dalam X sehingga fungsi $D(X, Q)$ akan minimal jika $f(A_j = q_j|X) \geq f(A_j = c_{kj}|X)$ untuk $q_j \neq c_{kj}$ untuk semua $j = 1, 2, 3, \dots, r$. Fungsi objektif yang digunakan dalam K -modes seperti pada persamaan (2.14)

$$J = \sum_{l=1}^k \sum_{i=1}^n \sum_{j=1}^r w_{i,l}(x_{i,j}, q_{l,j}) \quad (2.14)$$

$(x_{i,j}, q_{l,j})$ merupakan nilai pencocokan antara vektor dengan modus kluster yang diikuti, sedangkan $w_{i,l}$ merupakan nilai keanggotaan data setiap kluster. $w_{i,l}$ memiliki nilai antara 0 sampai 1 yang diperoleh dari persamaan (2.15)

$$w_{i,l} = \begin{cases} 1 & \text{jika } d(X_i, Q_l) < d(X_i, Q_t) \\ 0 & \text{untuk } t \neq l \end{cases} \quad (2.15)$$

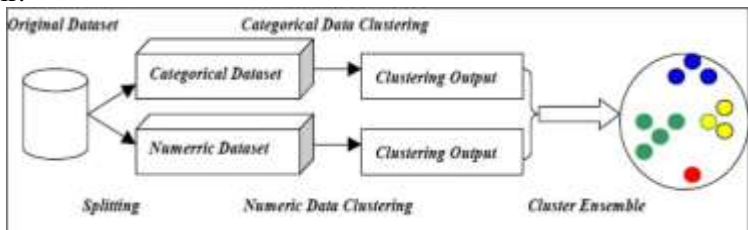
Dimana $l = 1, 2, 3, \dots, k$ dan $t = 1, 2, 3, \dots, k$. k adalah jumlah kluster sedangkan n adalah jumlah data dalam kluster. Algoritma k -modes yang digunakan (Prasetyo, 2014) adalah sebagai berikut.

1. Menentukan k data sebagai modus untuk setiap kelompok.
2. Mengalokasikan data ke kelompok dengan modulusnya terdekat.
3. Memperbarui modus sebagai centroid dari setiap kelompok dengan nilai kategori yang sering muncul pada setiap kelompok.
4. Setelah semua data telah terkelompokkan, uji data dengan nilai modus baru dan memperbarui kelompok.
5. Mengulangi langkah ke (2) dan (3) selama masih ada data yang berpindah kelompok atau ada perubahan nilai fungsi objektif berada dibawah ambang batas yang ditentukan.

Metode *K-modes* dapat diterapkan pada data kategorik menghasilkan kelompok yang lebih rinci, waktu pembentukan kelompok lebih singkat, dan unggul dalam pengelompokan pada data berdimensi besar.

2.6 Pengelompokan Data Campuran Numerik dan Kategorik

Pengelompokan data campuran dilakukan dengan membagi data menjadi sub-data numerik dan kategorik. Pengelompokan dilakukan secara terpisah dengan algoritma tertentu. Misal jumlah variabel numerik adalah n dan jumlah variabel kategorik adalah m , sehingga diperoleh penjumlahan dari $n + m$. Hasil pengelompokan tersebut digabung menggunakan metode pengelompokan berdasarkan SWFM sehingga diperoleh kelompok akhir.

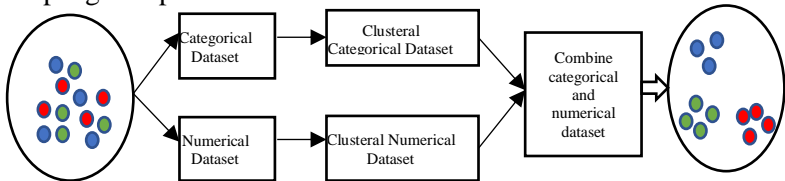


Sumber: Reddy dan Kavitha (2012)

Gambar 2.1 Kerangka Analisis Algoritma

2.7 Pengelompokan SWFM

Pengelompokan SWFM memiliki konsep analisis yang sama dengan pengelompokan lainnya, yang menjadi pembeda adalah pada tahap pembentukan klaster akhir. Pada SWFM menggunakan algoritma yang berdasarkan dengan *similarity weight and filter method*. Gambar 2.2 menjelaskan pengelompokan dari SWFM



Sumber: Reddy dan Kavitha (2012)

Gambar 2.2 Kerangka Analisis Berdasarkan *Similarity Weight and Filter Method*

Pada tahap metode *similarity weight* menggunakan ukuran kemiripan dengan memasukkan faktor-faktor bobot pada rumus ukuran kemiripan. Pembobot yang diberikan berdasarkan jumlah anggota pengamatan (n_i) atau (n_j). Rumus yang akan digunakan dalam menghitung ukuran kemiripan anggota antar pasangan objek ke- i dan objek ke- j adalah sebagaimana pada persamaan 2.16 (Reddy dan Kavitha, 2012).

$$\text{sim}(X_i, X_j) = \sum_{i \leq n_i, j \leq n_j} \frac{S_{ij}}{\max(n_i, n_j)}, i \neq j \quad (2.16)$$

Dengan

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

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

X_i : Himpunan pengamatan ke- i dengan $X_i = \{x_{1i}, x_{2i}, x_{3i}, \dots, x_{ki}\}$

X_j : Himpunan pengamatan ke- j dengan $X_j = \{x_{1j}, x_{2j}, x_{3j}, \dots, x_{kj}\}$

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

- m_k : Jumlah variabel kategorik dalam pengamatan
 n : Total jumlah objek pengamatan
 n_i : Jumlah anggota dalam kelompok ke- i
 n_j : Jumlah anggota dalam kelompok ke- j

Untuk mengelompokkan data campuran menggunakan algoritma metode *filter* berdasarkan *similarity weight*. Pertama data dibagi menjadi dua-sub data yaitu data kategorik murni dan data numerik murni. Pada masing-masing data dilakukan pengelompokan berdasarkan jenis data untuk mendapatkan kelompok yang sesuai. Selanjutnya hasil pengelompokan data numerik dan kategorik digabungkan untuk memperoleh pengelompokan akhir menggunakan metode *filter* (Reddy & Kavitha, 2012). Persamaan yang digunakan pada metode *filter* sebagaimana pada persamaan (2.17)

$$F(X_i, X_j) = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w_{ij} d(X_i, X_j) \quad (2.17)$$

Dimana $d(X_i, X_j) = 1 - sim(X_i, X_j)$ dan w_{ij} merupakan bobot antara kelompok ke- i dan ke- j dan pada umumnya bernilai 0,5 dan $0 < sim(X_i, X_j) \leq 1$.

Misal X_1 dan X_2 merupakan hasil pengelompokan pada data numerik dengan hasil kelompok $X_1 = \{1,3,5\}$ dan $X_2 = \{2,4,6\}$. X_3 , X_4 , dan X_5 merupakan hasil pengelompokan pada data kategorik dengan hasil kelompok $X_3 = \{1,2\}$; $X_4 = \{3,4\}$; $X_5 = \{5,6\}$ jadi $m = 2$, $n = 2$ dan *similarity* antar kelompok diberikan pada Tabel 2.2

Tabel 2.2 Contoh Matriks Similarity

Klaster	X_3	X_4	X_5
X_1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
X_2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Pada sel $X_1 X_3$, $p = |X_1 \cap X_3| = |\{1\}| = 1$, dan $q = |X_1 \cup X_3| = |\{1,2,3,5\}| = 4$ oleh karena itu $X_1 X_3 = \frac{p}{q} = \frac{1}{4}$. Similarity pada sel lainnya juga dihitung sehingga diperoleh

similarity antara kelompok data X_i dan kelompok data X_j adalah $sim(X_i, X_j) = \frac{(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4})}{3} = 0.5$.

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

2.8 Kinerja Hasil Pengelompokan

Hasil kinerja pengelompokan digunakan untuk melihat validitas dari suatu hasil pengelompokan. Homogenitas yang tinggi antar anggota dalam satu kelompok dan heterogenitas yang tinggi antar kelompok mengindikasikan kelompok yang baik (Hair, *et. al.*, 2009). Kinerja hasil diketahui dari rasio nilai S_W dan S_B menggunakan nilai rata-rata variabel, simpangan baku dalam kelompok, dan simpangan baku antar kelompok. Berikut adalah rumus yang digunakan sebagai mana persamaan

$$S_W = \frac{1}{C} \sum_{c=1}^C S_c$$

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

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

Dimana \bar{x}_c adalah rata-rata kelompok ke- c dan \bar{x} adalah rata-rata keseluruhan kelompok. Kinerja suatu metode pengelompokan semakin baik, jika semakin kecil nilai rasio antara S_W dan S_B . Hal ini mengindikasikan bahwa terdapat homogenitas yang tinggi dalam kelompok dan heterogenitas maksimum antar kelompok (Bunker & James, 1996).

Pengukuran kinerja dengan rasio S_W dan S_B hanya digunakan untuk data numerik sedangkan untuk data kategorik menggunakan tabel kontingensi berupa ANOVA.

2.9 Sekolah Menengah Atas (SMA)

SMA merupakan kelanjutan pendidikan menengah pertama yang diselenggarakan untuk mempersiapkan peserta didik untuk menjadi anggota masyarakat yang memiliki kemampuan akademis dan professional yang dapat menerapkan, mengembangkan, dan menciptakan ilmu pengetahuan, teknologi, dan kesenian sesuai dengan bidangnya dan berada dibawah pengawasan pemerintah (UU 2 tahun 1989, pasal 16, ayat (1)). SMA memiliki peran yang cukup penting dalam pengembangan suatu negara sehingga kualitas dari SMA juga harus sesuai. Beberapa ukuran dalam sarana prasarana maupun SDM dapat dilihat dari faktor-faktor antara lain rasio siswa per rombel, jumlah guru tetap, jumlah guru non-tetap, jumlah lulusan per kelas, luas lahan, dan daya listrik.

Aspek lain yang digunakan pada penelitian ini yaitu akreditasi yang menjadi salah satu tolak ukur dari SMA, semakin tinggi akreditasi maka semakin baik pula standar mutu serta penilaian terhadap SMA tersebut dari pihak luar. Berdasarkan badan akreditasi nasional, nilai akreditasi ditentukan secara sistematis melalui evaluasi diri dan evaluasi eksternal.

Status pengelolaan SMA juga menjadi acuan kualitas. Status pengelolaan SMA dibagi menjadi dua antara lain, sekolah negeri dan swasta. Sedangkan waktu penyelenggaraan SMA menjadi karakteristik tambahan dimana pada penelitian ini dibagi menjadi pagi, siang, dan kombinasi.

(halaman ini sengaja dikosongkan)

BAB III

METODE PENELITIAN

3.1 Sumber Data

Sumber data yang digunakan dalam penelitian adalah data sekunder yang diambil dari Dinas Pendidikan provinsi Jawa Timur wilayah kabupaten Sidoarjo tentang deskripsi SMA. Unit penelitian berupa SMA negeri maupun swasta yang ada di wilayah kabupaten Sidoarjo. Data yang diambil merupakan data SMA pada tahun angkatan 2015-2016.

3.2 Variabel Penelitian

Terdapat tiga variabel penelitian berskala kategorik dan empat variabel berskala numerik yang nantinya akan digunakan dalam pengelompokan SMA berdasarkan karakteristik SMA. Variabel-variabel penelitian yang digunakan sebagai mana pada Tabel 3.1

Tabel 3.1. Variabel Penelitian

Variabel	Keterangan	Kategori	Skala Pengukuran
Z ₁	Akreditasi SMA	0: Akreditasi A	Ordinal
		1: Akreditasi B	
		2: Akreditasi C	
Z ₂	Waktu Penyelenggaraan	0: Pagi	Nominal
		1: Siang	
		2: Kombinasi	
Z ₃	Status SMA	0: Negeri	Nominal
		1: Swasta	
X ₁	Rasio siswa per rombel	-	Rasio
X ₂	Jumlah guru tetap	-	Rasio
X ₃	Jumlah tenaga non-guru	-	Rasio
X ₄	Jumlah lulusan per kelas	-	Rasio
X ₅	Luas Lahan	-	Rasio
X ₆	Daya listrik	-	Rasio

3.3 Struktur Data

Berikut ini merupakan struktur data dari variabel yang digunakan dalam penelitian.

Tabel 3.2 Struktur Data

SMA	X_1	X_2	X_3	X_4	X_5	X_6	Z_1	Z_2	Z_3
1	$X_{1.1}$	$X_{2.1}$	$X_{3.1}$	$X_{4.1}$	$X_{5.1}$	$X_{6.1}$	$Z_{1.1}$	$Z_{2.1}$	$Z_{3.1}$
2	$X_{1.2}$	$X_{2.2}$	$X_{3.2}$	$X_{4.2}$	$X_{5.2}$	$X_{6.2}$	$Z_{1.2}$	$Z_{2.2}$	$Z_{3.2}$
3	$X_{1.3}$	$X_{2.3}$	$X_{3.3}$	$X_{4.3}$	$X_{5.3}$	$X_{6.3}$	$Z_{1.3}$	$Z_{2.3}$	$Z_{3.3}$
.
.
51	$X_{1.51}$	$X_{2.51}$	$X_{3.51}$	$X_{4.51}$	$X_{5.51}$	$X_{6.51}$	$Z_{1.51}$	$Z_{2.51}$	$Z_{3.51}$

3.4 Langkah Analisis

Berdasarkan hasil penentuan variabel yang digunakan dalam penelitian, langkah-langkah yang dilakukan untuk mengelompokkan data kategorik dan numerik pada SMA berdasarkan sarana prasarana dan SDM menggunakan *Similarity Weight and Filter Method* adalah sebagai berikut.

1. Mengklasifikasikan data yang berskala nominal dan ordinal kedalam kategorik tertentu.
2. Mendeskripsikan variabel-variabel yang bersifat kategorik dan numerik pada data SMA. Untuk menjawab tujuan pertama dilakukan pengolahan data secara deskriptif untuk mengetahui karakteristik pada setiap variabel dengan menggunakan grafik batang ataupun grafik lingkaran dan mean pada data numerik dan modus pada data kategorik untuk dapat dibaca dari berbagai sudut pandang.
3. Memisahkan data menjadi dua bagian. Satu untuk data numerik dan yang lain untuk data kategorik.
4. Mengelompokkan SMA berdasarkan sarana prasarana dan SDM dengan langkah-langkah sebagai berikut.
 - a. Melakukan penerapan metode *k-modes* pada data kategorik.
 - b. Melakukan pengelompokan menggunakan metode *hierarki agglomerative* yaitu *single linkage*,

- complete linkage*, dan *average linkage* pada data numerik.
- c. Menggabungkan hasil pengelompokan pada kedua bagian.
 - d. Melakukan pengolahan data menggunakan *Similarity Weight Method*
 - e. Mengelompokan data menggunakan *Similarity Analysis*.
 - f. Menentukan *Filter Algorithm* untuk data kategorik dan numerik.
 - g. Mengelompokan data menggunakan *Similarity Weight and Filter Method*.
5. Menganalisis hasil pengelompokan.

(halaman ini sengaja dikosongkan)

BAB IV ANALISIS DAN PEMBAHASAN

Pada penelitian ini dijelaskan mengenai analisis dan pembahasan untuk menjawab tujuan penelitian yang telah dijelaskan sebelumnya.

4.1 Analisis Statistika Deskriptif

Pada subbab ini, dilakukan analisis deskriptif terhadap variabel numerik dan variabel kategorik. Statistika deskriptif untuk variabel numerik akan disajikan dalam ukuran pemusatan data dan *boxplot* sedangkan untuk variabel kategorik akan disajikan dalam bentuk histogram.

4.1.1 Data Numerik

Variabel numerik pada penelitian ini antara lain rasio siswa/rombel, jumlah guru tetap, jumlah guru non-tetap, jumlah lulusan/kelas, luas lahan sekolah, dan daya listrik yang digunakan. Statistika deskriptif data numerik ditampilkan pada Tabel 4.1,

Tabel 4.1 Statistika Deskriptif Variabel Numerik

Variabel	<i>Mean</i>	<i>StDev</i>	<i>Variance</i>	<i>Min</i>	<i>Median</i>	<i>Max</i>
Rasio siswa/rombel	30,02	8,27	69,46	10,67	31,23	50.03
Jumlah guru tetap	21,27	17,43	303,68	3	12	59
Jumlah guru non-tetap	4,902	6,874	47,25	0	2	31
Jumlah lulusan/kelas	27,99	9,74	94,93	10	28,92	48,2
Luas lahan (m ²)	5854	7615	57981304	23	3000	38000
Daya listrik (watt)	38777	88810	7887181531	1200	12000	600000

Variabel rasio siswa per rombel, memiliki rata-rata sebesar 30 siswa per rombel dengan rasio siswa per rombel terkecil 11 siswa per rombel pada SMA Terpadu Al Mubarakah dan terbesar sebesar 50 siswa per rombel pada SMA Bhayangkari 3. Rasio

siswa/rombel tiap pengamatan relatif sama karena nilai mean dan median tidak terpaut jauh.

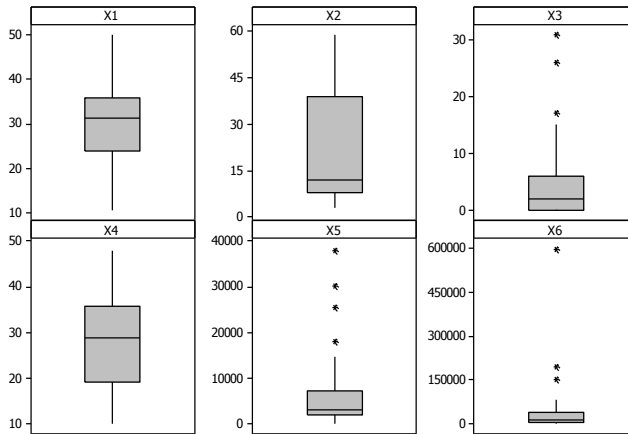
Tabel 4.1 juga menyajikan rata-rata jumlah guru tetap sebesar 22 guru dengan jumlah guru tetap terkecil sebesar 3 guru pada SMA Panca Bhakti dan terbesar sebesar 59 guru pada SMA 1 Krian. Pada variabel jumlah guru tetap nilai varian sebesar 303 menunjukkan jumlah guru antar SMA di Sidoarjo memiliki keberagaman yang cukup besar.

Rata-rata jumlah guru non-tetap sebesar 5 guru dengan jumlah guru non-tetap terkecil sebesar 0 guru dan terbesar sebesar 31 guru pada SMAS 2 Muhammadiyah. Nilai 0 diketahui bahwa terdapat pada 19 SMA yang disebabkan karena tenaga pengajar yang digunakan berasal dari guru tetap.

Tabel 4.1 juga menjelaskan rata-rata jumlah lulusan per kelas sebesar 28 lulusan dengan jumlah lulusan per kelas terkecil sebesar 10 lulusan pada SMAS Yayasan Taman dan terbesar sebesar 48 lulusan pada SMA Bhayangkari 3. Nilai variabel jumlah lulusan per kelas hampir sama karena rata-rata dan median bernilai 28 dan 29. Hal ini menunjukkan bahwa jumlah lulusan/kelas dari SMA di Sidoarjo lebih kecil dari rasio siswa/rombel sehingga terdapat siswa yang tidak lulus pada beberapa SMA.

Berdasarkan pada Tabel 4.1, rata-rata luas lahan sebesar 5854 m² dengan luas lahan terkecil sebesar 23 m² pada SMAN 1 Krembung dan terbesar sebesar 38000 m² SMAN Olah Raga Sidoarjo. Luas lahan relatif beragam karena terlihat dari nilai varian yang tinggi. Sedangkan variabel daya listrik yang digunakan memiliki rata-rata sebesar 38777 watt dengan daya listrik terkecil sebesar 1200 watt pada SMAS Dharma Wanita 4 dan terbesar sebesar 600000 watt pada SMAS Bhayangkari 3.

Selain memperlihatkan secara deskriptif, dapat disajikan suatu boxplot dalam rangka memperlihatkan penyebaran data. Boxplot diperlihatkan pada Gambar 4.1.

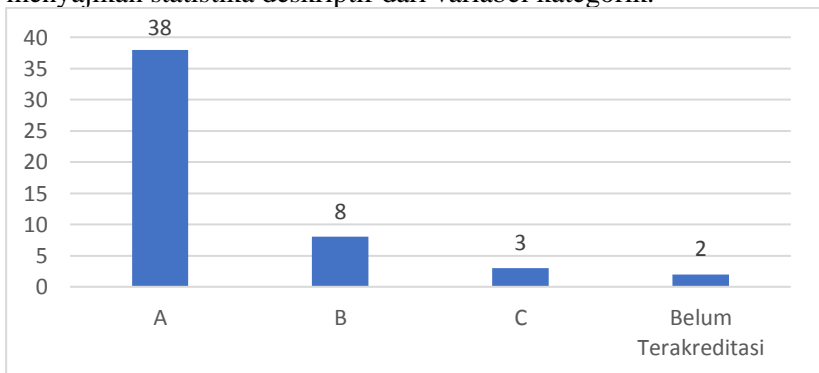


Gambar 4.1 Karakteristik Data Variabel Numerik

Dari enam variabel yang berskala numerik, terdapat tiga variabel yang sangat menonjol dalam aspek outlier yaitu variabel jumlah guru non-tetap (x3), luas lahan (x5), dan daya listrik (x6) sedangkan variabel rasio siswa per rombel (x1), jumlah guru tetap (x2), dan jumlah lulusan (x4) tidak memiliki outlier.

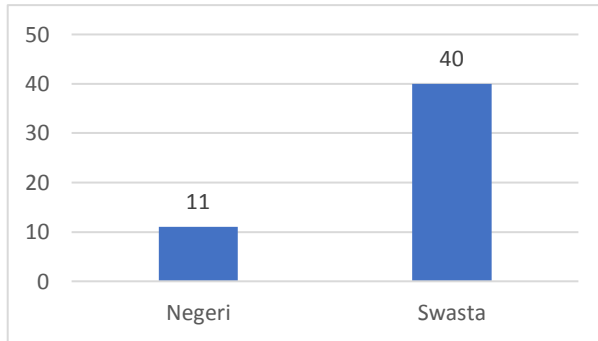
4.1.2 Data Kategorik

Variabel kategorik pada penelitian ini adalah akreditasi sekolah, status SMA, dan waktu penyelenggaraan. Gambar 4.2 menyajikan statistika deskriptif dari variabel kategorik.



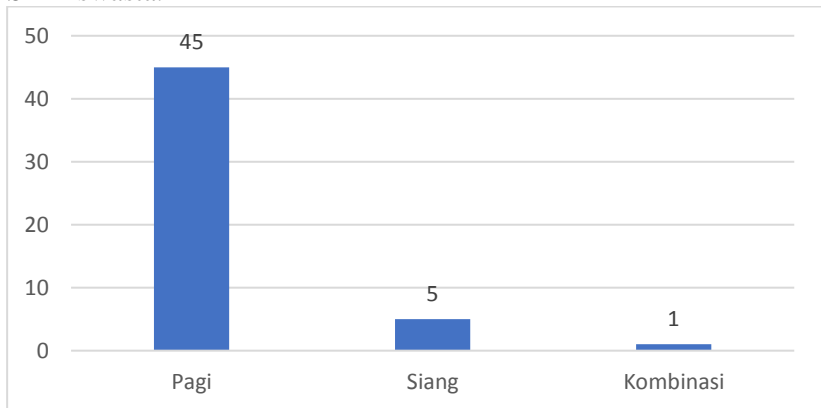
Gambar 4.2 Histogram Variabel Akreditasi

Karakteristik pada Gambar 4.2 menyajikan akreditasi terbanyak pada kategori akreditasi A sebesar 38 SMA (74,5%) diikuti akreditasi B sebanyak 8 SMA (15,7%), akreditasi C sebanyak 3 SMA (5,8%), dan paling sedikit adalah SMA yang belum terakreditasi sebanyak 2 SMA (4%). Hal ini menunjukkan sebagian besar SMA di Sidoarjo memiliki akreditasi A.



Gambar 4.3 Histogram Variabel Status SMA

Variabel status SMA pada kabupaten Sidoarjo seperti Gambar 4.3, yang berstatus swasta sebanyak 40 SMA dari 51 SMA yang ada (78,5%) sedangkan SMA negeri hanya sebanyak 11 (21,5%). Hal ini menunjukkan SMA di Sidoarjo didominasi oleh SMA swasta.



Gambar 4.4 Histogram Variabel Waktu Penyelenggaraan

Karakteristik pada Gambar 4.4 adalah variabel waktu penyelenggaraan. Waktu penyelenggaraan terbanyak pada pagi hari sebesar 45 SMA (88,2%) sedangkan sisanya dilakukan pada siang hari dan kombinasi masing-masing sebanyak 5 SMA (9,8%) dan 1 SMA (2%). Hal ini menunjukkan SMA di Sidoarjo mayoritas diselenggarakan pada pagi hari.

Selain secara grafik, data kategorik akan disajikan dalam bentuk *crosstab*. *Crosstab* digunakan untuk mengetahui hubungan kecenderungan antara satu variabel dengan variabel lain. Terdapat tiga *crosstab* antara lain, akreditasi dengan status sekolah, akreditasi dengan waktu penyelenggaraan, dan status sekolah dengan waktu penyelenggaraan. *Crosstab* antara akreditasi dengan status sekolah disajikan pada Tabel 4.2.

Tabel 4.2 Crosstab antara Akreditasi dengan Status Sekolah

		Akreditasi			
			A	Selain A	Total
Status Sekolah	Negeri	Jumlah	10	1	11
		%	90.9%	9.1%	100.0%
	Swasta	Jumlah	28	12	40
		%	70%	30%	100.0%
Total		Jumlah	38	13	51
		%	74.5%	25.5%	100.0%

Berdasarkan Tabel 4.2, diperoleh nilai odds ratio dari variabel SMA berstatus Negeri dan swasta sebesar 4,286 yang berarti SMA yang memiliki berstatus negeri memiliki kecenderungan untuk memiliki akreditasi A sebesar 4,286 kali dibanding dengan SMA yang berakreditasi swasta. Sedangkan nilai relative risk dari SMA yang berstatus negeri dengan swasta sebesar 1,299 yang artinya SMA dengan status negeri memiliki peluang berakreditasi A 1,299 kali lebih besar dibanding dengan SMA swasta.

Tabel 4.3 Crosstab antara Akreditasi dengan Waktu Penyelenggaraan

		Akreditasi			
		A	Selain A	Total	
Waktu Penyelenggaraan	Pagi	Jumlah	37	8	45
		%	82.2%	17.8%	100.0%
	Siang	Jumlah	1	5	6
		%	16.7%	83.3%	100.0%
Total	Jumlah	38	13	51	
	%	74.5%	25.5%	100.0%	

Dari Tabel 4.3, nilai odds ratio dari variabel waktu penyelenggaraan pada pagi hari dan siang hari sebesar 23,125 yang berarti SMA yang dilaksanakan pada pagi hari memiliki kecenderungan berakreditasi A sebesar 23,125 kali dibanding dengan SMA yang diselenggarakan pada siang hari. Sedangkan nilai relative risk dari SMA yang diselenggarakan pada pagi hari dengan waktu penyelenggaraan pagi siang hari bernilai 4.933 yang artinya SMA dengan waktu penyelenggaraan pagi hari memiliki peluang berakreditasi A 1,582 kali lebih besar dibanding dengan SMA yang dilaksanakan pada siang hari.

Tabel 4.4 Crosstab antara Status Sekolah dengan Waktu Penyelenggaraan

		Waktu Penyelenggaraan			
		Pagi	Siang	Total	
Status Sekolah	Negeri	Jumlah	11	0	11
		%	100.0%	0.0%	100.0%
	Swasta	Jumlah	34	6	40
		%	85.0%	15.0%	100.0%
Total	Jumlah	45	6	51	
	%	88.2%	11.8%	100.0%	

Tabel 4.4, menunjukkan nilai relative risk dari SMA yang berstatus negeri dengan waktu penyelenggaraan pagi hari bernilai 1,176 yang artinya SMA dengan status sekolah negeri memiliki peluang 1,176 kali lebih besar dibanding dengan SMA yang

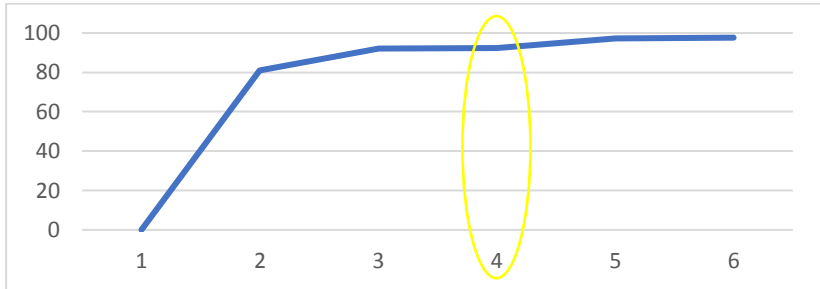
berstatus swasta atau dengan kata kalin hampir sama. Nilai odds ratio pada crosstab status sekolah dengan waktu penyelenggaraan tidak ada Karena terdapatnya nilai 0 pada kolom status sekolah negeri dan waktu penyelenggaraan siang.

4.2 Hasil Pengelompokan dengan Metode SWFM

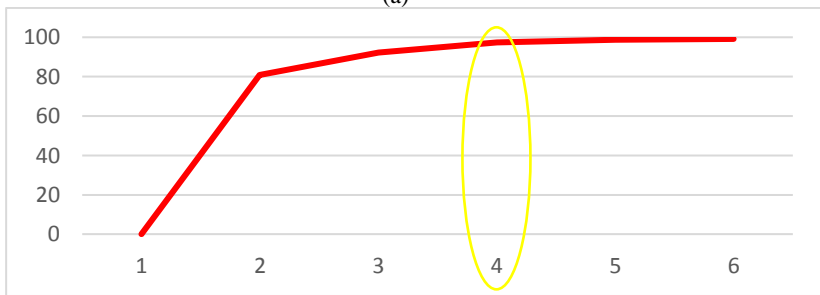
Pada subbab ini dijelaskan mengenai analisis yang dilakukan untuk mengelompokan kualitas SMA di Kabupaten Sidoarjo. Analisis ini dilakukan menggunakan metode *hierarki agglomerative* pada data numerik sedangkan *k-modes* digunakan pada data kategorik dan metode SWFM pada data campuran.

4.2.1 Data Numerik

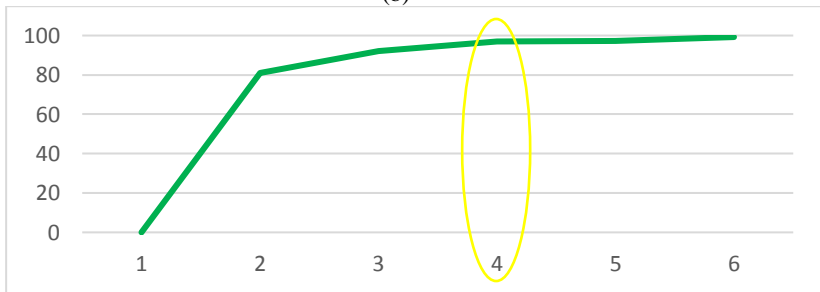
Pengelompokan untuk data numerik dilakukan menggunakan teknik *single linkage*, *complete linkage*, dan *average linkage* dengan setiap objek pengamatan sebagai suatu kelompok tunggal dengan anggota tunggal. Tahapan pada analisis ini dengan membentuk matrik jarak antar objek pengamatan. Jarak yang digunakan pada penelitian ini adalah jarak Euclidean yang diperoleh dari 51 objek pengamatan. Jumlah kelompok yang mungkin terbentuk antara dua sampai enam kelompok yang dilakukan menggunakan *software R*. Setelah didapatkan hasil pengelompokan dari dua sampai enam kelompok, kemudian didapatkan nilai R-square dari lampiran 11 - 15. Nilai R-square dari jumlah kelompok 2 sampai 6 dibentuk plot sehingga terlihat jumlah kelompok yang meningkat secara tajam dan cenderung stabil pada jumlah kelompok selanjutnya. Jumlah kelompok maksimal sebanyak enam dikarenakan jumlah variabel pada variabel numerik hanya sebanyak enam. Gambar 4.6 merupakan hasil pengelompokan data numerik berdasarkan nilai *R-Square*.



(a)



(b)



(c)

Gambar 4.5 Nilai R-Square dari Metode Single Linkage (a), Complete Linkage (b), dan Average Linkage (c)

Berdasarkan Gambar 4.5, jumlah kelompok optimum berada pada 4 kelompok untuk metode single linkage, complete linkage, dan average linkage. Jumlah kelompok optimum didasarkan pada nilai R-square yang meningkat tajam dan cenderung stabil pada jumlah kelompok selanjutnya. Perubahan

nilai R-square pada 3 kelompok ke 4 kelompok masih cukup tajam namun pada 4 kelompok ke 5 kelompok dan 5 kelompok ke kelompok 6 cenderung stabil. Nilai R-square pada 4 kelompok menggunakan metode single linkage, complete linkage, dan average linkage masing-masing sebesar 92,48%; 97,26%; 96,85%.

Tabel 4.5 Rasio hasil pengelompokan pada 4 kelompok

Metode	Sw	Sb	Rasio
Single	1158.584	57638.05	0.0201
Complete	2503.502	55738.5	0.0449
Average	2437.85	55791.33	0.0436

Berdasarkan Tabel 4.5, pada 4 kelompok diperoleh nilai rasio terbaik pada metode single linkage yaitu sebesar 0,0201 karena semakin kecil rasio antara Sw dengan Sb mengindikasikan homogenitas yang lebih tinggi dalam kelompok dan heterogenitas yang lebih maksimum antar kelompok. Hal ini menunjukkan bahwa pengelompokan data numerik dengan 4 kelompok lebih tepat menggunakan metode single linkage.

Pembagian jumlah anggota kelompok pada metode single linkage dengan 4 kelompok adalah

Kelompok 1 : 48 SMA

Kelompok 2 : 1 SMA

Kelompok 3 : 1 SMA

Kelompok 4 : 1 SMA

Sedangkan anggota kelompok hasil pengelompokan adalah sebagai berikut,

a. Kelompok 1

Kelompok 1 berjumlah 48 objek yang beranggotakan SMAS Ulul Al Abab, SMAS Antartika, SMAN Olah Raga Sidoarjo, SMAS Muhammadiyah 3, SMAN 1 Porong, SMAN 1 Krian, SMAN 4 Sidoarjo, SMAS PGRI 1 Sidoarjo, SMAS Muhammadiyah 1 Taman, SMAS Senopati Sedati, SMAN 1 Taman, SMAN 1 Krembung, SMAN 1 Wonoayu, SMAN 1 Sidoarjo, SMAN 1

Gedangan, SMAN 1 Tarik, SMAS Cendekia, SMAS Islam Sidoarjo, SMAS Al Fatah, SMAS Al Ahmad, SMAS TPI Porong, SMAS Dharma Wanita 3, SMAS Persatuan Tulangan, SMAS Katholik Untung Surapati Sidoarjo, SMAS Kemala Bhayangkara 4, SMAS Al Islamiyah, SMAS Dharma Wanita 1, SMAS Katolik Untung Suropati Krian, SMAS Budi Utomo Prambon, SMAS TPI Gedangan, SMAS Wachid Hasyim 2, SMAS Putra Bangsa, SMAS Wachid Hasyim 1, SMAS Unggala, SMAS PGRI 5 Sidoarjo, SMAS Tribakti, SMAS Kristen Petra 4, SMAS Dharma Wanita 4, SMAS Panca Bhakti, SMAS Yayasan Taman, SMAS Terpadu Al Mubarakah, SMAS Islam Perlaungan, SMAS Wachid Hasyim 4, SMAS Muhammadiyah 4, SMAS Avisena Jabon, SMAS Nurul Huda Porong, SMAS Dharma Wanita 2, dan SMAS Wachid Hasyim 3.

b. Kelompok 2

Kelompok 2 beranggotakan satu objek yaitu SMA 1 Waru.

c. Kelompok 3

Kelompok 3 beranggotakan satu objek yaitu SMAS Kemala Bhayangkari 3.

d. Kelompok 4

Kelompok 4 beranggotakan satu objek yaitu SMAS Muhammadiyah 2.

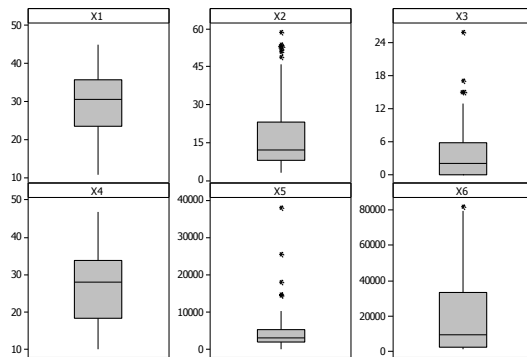
Karakteristik dari anggota hasil pengelompokan pada data numerik masing-masing kelompok disajikan pada tabel 4.6 berikut.

Tabel 4.6 Karakteristik Hasil Pengelompokan Data Numerik

Variabel	Kel 1	Kel 2	Kel 3	Kel 4
Rasio siswa/rombel	29,381	36	50,032	34,65
Jumlah guru tetap	19,875	50	39	42
Jumlah guru non-tetap	4,354	5	5	31
Jumlah lulusan/kelas	27,202	36,4	48,2	37,083
Luas lahan (m ²)	5319,667	10051	2944	30230
Daya listrik (watt)	21471,875	150000	600000	197000

Tabel 4.6 menunjukkan nilai rata-rata dari anggota kelompok masing-masing. Pada kelompok 3 terlihat memiliki rata-

rata tertinggi pada variabel rasio siswa per rombongan, jumlah lulusan per kelas, dan daya listrik yang digunakan dibandingkan dengan rata-rata pada kelompok lain. Rata-rata terbesar pada variabel jumlah guru non-tetap dan luas lahan berada pada kelompok 4 sedangkan pada kelompok 2 memiliki rata-rata terbesar pada variabel jumlah guru tetap. Kelompok 1 terlihat memiliki rata-rata yang lebih rendah dari kelompok-kelompok lain pada ke enam variabel numerik.



Gambar 4.6 Karakteristik Hasil Pengelompokan Data Numerik pada Kelompok 1

Gambar 4.6 menunjukkan boxplot dari anggota kelompok pertama. Pada kelompok 1 terlihat memiliki outlier pada variabel jumlah guru non tetap, jumlah guru tetap, luas lahan, dan daya listrik.

4.2.2 Pengelompokan pada Data Kategorik

Pengelompokan pada data kategorik dilakukan menggunakan teknik *k-modes* dengan setiap objek pengamatan sebagai suatu kelompok tunggal dengan anggota tunggal. Ukuran kemiripan antar objek pengamatan diukur dengan nilai modus yang berkorespondensi dari dua objek. Pada penelitian ini program R digunakan untuk melakukan analisis dengan menggunakan *k-modes*. Nilai *k* pada analisis ini ditentukan sebesar 2. Pengelompokan data lebih banyak masuk kedalam kelompok 1

terlihat dari jumlah kelompok 1 sebanyak 46 SMA sedangkan pada kelompok 2 hanya terdapat 5 SMA didalamnya.

Anggota kelompok hasil pengelompokan adalah sebagai berikut,

a. Kelompok 1

Kelompok 1 berjumlah 46 objek yang beranggotakan SMAS Ulul Al Abab, SMAS Antartika, SMAN Olah Raga Sidoarjo, SMAS Muhammadiyah 3, SMAN 1 Porong, SMAN 1 Krian, SMAN 4 Sidoarjo, SMAS PGRI 1 Sidoarjo, SMAS Muhammadiyah 1 Taman, SMAS Senopati Sedati, SMAN 1 Taman, SMAN 1 Krembung, SMAN 1 Wonoayu, SMAN 1 Sidoarjo, SMAN 1 Gedangan, SMAN 1 Tarik, SMAS Cendekia, SMAS Islam Sidoarjo, SMAS Al Fatah, SMAS Al Ahmad, SMAS TPI Porong, SMAS Dharma Wanita 3, SMAS Persatuan Tulangan, SMAS Katholik Untung Surapati Sidoarjo, SMAS Kemala Bhayangkara 4, SMAS Al Islamiyah, SMAS Dharma Wanita 1, SMAS Katolik Untung Suropati Krian, SMAS Budi Utomo Prambon, SMAS TPI Gedangan, SMAS Wachid Hasyim 2, SMAS Putra Bangsa, SMAS Wachid Hasyim 1, SMAS Unggala, SMAS PGRI 5 Sidoarjo, SMAS Tribakti, SMAS Kristen Petra 4, SMAS Dharma Wanita 4, SMAS Panca Bhakti, SMAS Yayasan Taman, SMAS Terpadu Al Mubarakah, SMAS Islam Perlaungan, SMAS Wachid Hasyim 4, SMAS Muhammadiyah 4, SMAS Avisena Jabon, SMAS Nurul Huda Porong, SMAS Dharma Wanita 2, dan SMAS Wachid Hasyim 3, SMA 1 Waru, SMAS Kemala Bhayangkari 3, SMAS Muhammadiyah 2.

b. Kelompok 2

Kelompok 2 berjumlah 5 objek yang beranggotakan SMAS Al Ahmad, SMAS Dharma Wanita 3, SMAS TPI Gedangan, SMAS Wachid Hasyim 1, SMAS Yayasan Taman.

Karakteristik dari anggota hasil pengelompokan pada data kategorik dijelaskan pada tabel 4.7 berikut.

Tabel 4.7 Hasil Pengelompokan Data Kategorik Rata-Rata Variabel

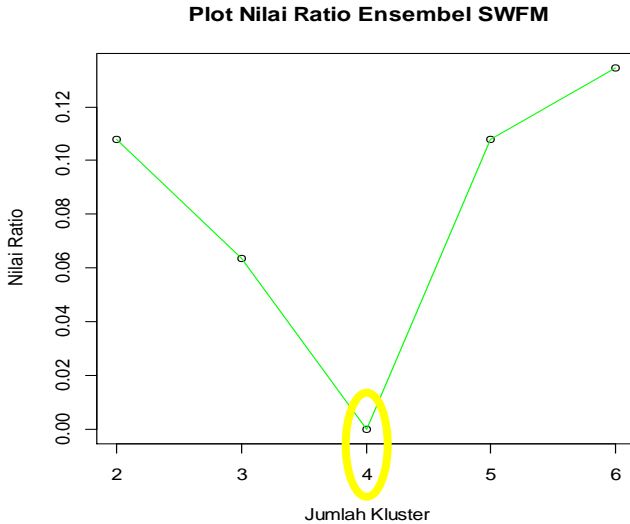
Variabel	Kel 1	Kel 2
Akreditasi	A (82,6%)	B (60%)
Status sekolah	Swasta (76,1%)	Swasta (100%)
Waktu penyelenggaraan	Pagi (97,8%)	Siang (80%)

Pada Tabel 4.7 menyajikan persentase dari kategori yang paling dominan dari variabel-variabel yang ada dalam kelompok. Pada kelompok 1 terlihat sebagian besar SMA memiliki status sekolah swasta (76,1%), diselenggarakan pada pagi hari (97,8%), dan berakreditasi A (82,6%). Kelompok 2 memiliki karakteristik akreditasi B sebesar 60%, status sekolah 100% swasta, dan waktu penyelenggaraan pada siang hari sebesar 80%.

4.2.3 Pengelompokan pada Data Campuran

Pada pengelompokan data campuran menggunakan teknik SWFM dengan melakukan pengelompokan masing-masing pada subbab sebelumnya. Tahap pertama dalam melakukan analisis SWFM untuk data campuran adalah dengan melakukan pengelompokan masing-masing jenis data menggunakan metode masing-masing. Hasil pengelompokan untuk data numerik diperoleh dari analisis pada subbab 4.2.1 dan hasil pengelompokan untuk data kategorik diperoleh dari analisis pada subbab 4.2.2. Hasil pengelompokan yang telah diperoleh dinyatakan sebagai variabel kategorik yang nantinya akan diolah dengan metode SWFM.

Tahap selanjutnya membentuk matrik jarak antar objek pengamatan menggunakan jarak similarity weight, kemudian dibentuk nilai jarak F antar objek pengamatan. Setelah nilai jarak F terbentuk, dilakukan penggabungan kelompok dengan jarak terdekat dan pembaruan matriks jarak. Jumlah kelompok yang dibentuk antara dua sampai enam kelompok yang dilakukan dengan program R. Hasil pengelompokan terbaik ditentukan dari nilai rasio Sw dan Sb terkecil. Berikut ini merupakan plot nilai rasio pada Gambar 4.7.



Gambar 4.7 Plot Nilai Rasio SWFM

Hasil pengelompokan terbaik yang disajikan pada Gambar 4.9 menunjukkan nilai rasio Sw dan Sb terkecil pada pengelompokan dengan jumlah kelompok 4 sebesar $6.745682e-17$. Nilai sebesar $6.745682e-17$ menunjukkan bahwa simpangan baku dalam kelompok bernilai $6.745682e-17$ kali dari simpangan baku kelompok lain sehingga variansi data dalam kelompok juga lebih kecil dibanding variansi kelompok lain. Tabel 4.4 menjelaskan nilai rasio pada kelompok dua sampai enam.

Tabel 4.8 Nilai Rasio Pengelompokan dengan Metode SWFM

Jumlah Kelompok	Rasio Sw dan Sb
2	$1.076942e-01$
3	$6.339664e-02$
4	$6.745682e-17$
5	$1.078036e-01$
6	$1.344235e-01$

Terlihat dari Tabel 4.8 nilai rasio terkecil berada pada 4 kelompok. Jumlah anggota pada kelompok pertama berjumlah 43 SMA sedangkan pada kelompok kedua berisi 6 anggota, kelompok

ketiga dan kelompok empat berisi masing-masing 1 SMA. Anggota setiap kelompok pada pengelompokan dengan metode SWFM adalah sebagai berikut

a. Kelompok 1

Kelompok 1 berjumlah 43 objek yang beranggotakan SMAS Ulul Al Abab, SMAS Islam Sidoarjo, SMAS Al Fatah, SMAS Antartika, SMAS TPI Porong, SMAS Dharma Wanita 3, SMAS Persatuan Tulangan, SMAS Katholik Untung Surapati Sidoarjo, SMAS Kemala Bhayangkara 4, SMAS Al Islamiyah, SMAS Dharma Wanita 1, SMAS Katolik Untung Suropati Krian, SMAS Muhammadiyah 3, SMAS Budi Utomo Prambon, SMAS Wachid Hasyim 2, SMAS Putra Bangsa, SMAS PGRI 1 Sidoarjo, SMAS Unggala, SMAS PGRI 5 Sidoarjo, SMAS Muhammadiyah 1 Taman, SMAS Senopati Sedati, SMAS Tribakti, SMAS Kristen Petra 4, SMAS Dharma Wanita 4, SMAS Panca Bhakti, SMAS Terpadu Al Mubarakah, SMAS Islam Perlaungan, SMAS Wachid Hasyim 4, SMAS Cendekia, SMAS Muhammadiyah 4, SMAS Avisena Jabon, SMAS Nurul Huda Porong, SMAS Dharma Wanita 2, SMAS Wachid Hasyim 3, SMAN Olah Raga Sidoarjo, SMAN 1 Porong, SMAN 1 Krian, SMAN 4 Sidoarjo, SMAN 1 Taman, SMAN 1 Krembung, SMAN 1 Wonoayu, SMAN 1 Sidoarjo, SMAN 1 Gedangan, dan SMAN 1 Tarik.

b. Kelompok 2

Kelompok 2 berjumlah 6 objek yaitu SMAS AL Ahmad, SMAS Yayasan Taman, SMAS Wachid Hasyim 1, SMAN 1 Waru, SMAS TPI Gedangan, dan SMAS Dharma Wanita 3.

c. Kelompok 3

Kelompok 3 beranggotakan 1 objek yaitu SMAS Kemala Bhayangkari 3.

d. Kelompok 4

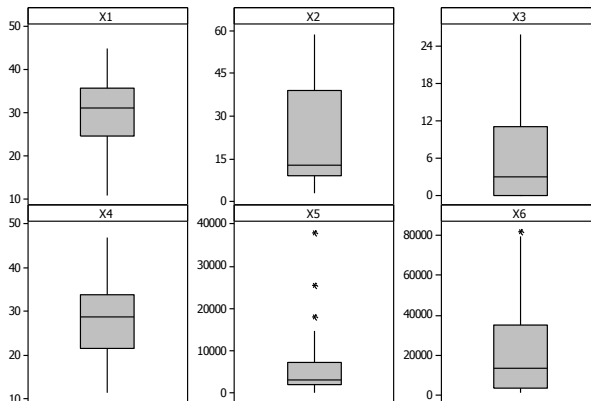
Kelompok 3 berisi 1 objek yaitu SMAS Muhammadiyah 2.

Karakteristik variabel numerik dari anggota hasil pengelompokan pada analisis SWFM pada masing-masing kelompok disajikan pada Tabel 4.9 berikut.

Tabel 4.9 Hasil Pengelompokan Data Numerik Rata-Rata Variabel Pada Metode SWFM

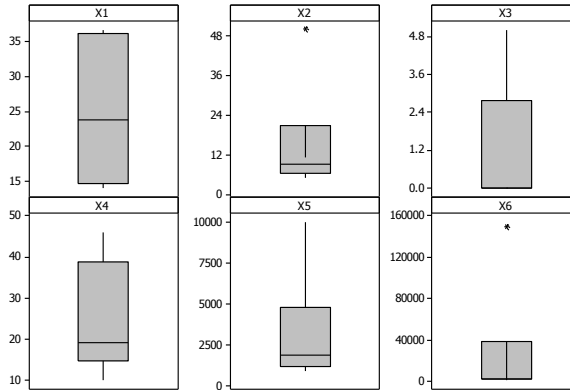
Variabel	Kel 1	Kel 2	Kel 3	Kel 4
Rasio siswa/rombel	30,173	24,806	50,03	34,65
Jumlah guru tetap	21,233	15,167	39	42
Jumlah guru non-tetap	44,814	1,167	5	31
Jumlah lulusan/kelas	27,784	24,567	48,2	37,08
Luas lahan (m ²)	5726,605	3191,833	2944	30230
Daya listrik (watt)	23731,395	26700	600000	197000

Tabel 4.9 menunjukkan nilai rata-rata dari anggota kelompok masing-masing. Pada kelompok 3 terlihat memiliki rata-rata tertinggi pada tiga variabel yaitu rasio siswa per rombel, jumlah lulusan, dan daya listrik dibandingkan dengan rata-rata pada kelompok lain. Sedangkan rata-rata terbesar pada variabel jumlah guru tetap, jumlah guru non-tetap, dan luas lahan berada pada kelompok 4.



Gambar 4.8 Karakteristik Hasil Pengelompokan SWFM pada Data Numerik Kelompok 1

Gambar 4.8 menunjukkan boxplot dari anggota kelompok pertama. Pada kelompok 1 terlihat memiliki outlier pada variabel luas lahan.



Gambar 4.9 Karakteristik Hasil Pengelompokan SWFM pada Data Numerik Kelompok 2

Gambar 4.9 menunjukkan boxplot dari anggota kelompok pertama. Pada kelompok 1 terlihat memiliki outlier pada variabel jumlah guru tetap dan daya listrik yang digunakan.

Sedangkan karakteristik variabel kategorik dari anggota hasil pengelompokan pada analisis SWFM disajikan pada Tabel 4.10 berikut.

Tabel 4.10 Karakteristik Variabel Kategorik Hasil Pada Metode SWFM

Variabel	Kel 1	Kel 2	Kel 3	Kel 4
Akreditasi	A (81,39%)	B (50%)	A (100%)	A (100%)
Status Sekolah	Swasta (76,74%)	Swasta (83,33%)	Swasta (100%)	Swasta (100%)
Waktu	Pagi	Siang	Pagi	Pagi
Penyelenggaraan	(97,67%)	(66,67%)	(100%)	(100%)

Tabel 4.10 menunjukkan persentase dari kategori yang paling dominan dari variabel-variabel yang ada dalam kelompok. Pada kelompok 1 terlihat sebesar 81,39% SMA memiliki akreditasi A, 76,74% SMA berstatus sekolah swasta, dan diselenggarakan pada pagi hari sebesar 97,67%, kelompok 2 berkarateristik sebagian

besar berakreditasi B sebanyak 50%, SMA berstatus sekolah swasta sebanyak 83,33%, dan diselenggarakan pada siang hari sebesar 66,67% sedangkan pada kelompok 3 dan 4 memiliki karakteristik yang sama yaitu berakreditasi A, berstatus sekolah swasta dan diselenggarakan pada pagi hari.

BAB V

KESIMPULAN DAN SARAN

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

5.1 Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, maka diperoleh beberapa kesimpulan yaitu,

1. Variabel numerik pada penelitian ini sebanyak 6 dengan jumlah objek pengamatan 51. Rata-rata rasio siswa per rombel (x_1) sebesar 30 siswa per rombel. Variabel jumlah guru tetap (x_2) memiliki rata-rata 21 guru sedangkan rata-rata variabel jumlah guru non-tetap (x_3) sebesar 5 guru. Jumlah lulusan per kelas (x_4) memiliki rata-rata 28 siswa per kelas. Luas lahan (x_5) memiliki rata-rata 5854m^2 dan rata-rata daya listrik sebesar 38777 watt. Pada data kategorik, mayoritas SMA di kabupaten Sidoarjo memiliki karakteristik berakreditasi A (74,5%), berstatus SMA swasta sebesar 78,5%, dan waktu penyelenggaraan pada pagi hari sebanyak 45 dari 51 SMA. Nilai odds ratio dari variabel akreditasi A dan selain A sebesar 4,286. Nilai odds ratio dari variabel akreditasi A dan selain A sebesar 23,125 yang berarti SMA yang memiliki akreditasi A memiliki kecenderungan untuk waktu penyelenggaraan dilaksanakan pada pagi hari sebesar 23,125 kali dibanding dengan SMA yang tidak berakreditasi A. Nilai relative risk dari SMA yang berstatus negeri dengan waktu penyelenggaraan pagi hari bernilai 1,176 yang artinya SMA dengan status sekolah negeri memiliki peluang 1,176 kali lebih besar dibanding dengan SMA yang berstatus swasta

2. Pada pengelompokan data numerik diperoleh jumlah kelompok optimum sebesar 4 menggunakan metode single linkage dengan nilai rasio S_w dan S_b 0,0201 dan R-square sebesar 92,48%. Pada pengelompokan data kategorik diperoleh jumlah kelompok 2 dan anggota masing-masing kelompok yaitu 46 dan 5 SMA. Pada pengelompokan menggunakan SWFM didapatkan jumlah kelompok terbaik pada jumlah kelompok 3 dengan nilai rasio S_w dan S_b sebesar 0,33.

5.2 Saran

Saran yang dapat diambil dari penelitian ini adalah perlu mempertimbangkan adanya outlier dari variabel jumlah guru non-tetap, luas lahan, dan daya listrik sehingga metode pengelompokan variabel numerik menghasilkan anggota kelompok yang lebih merata.

DAFTAR PUSTAKA

- Agresti, A. (2007). *Categorical Data Analysis (2nd ed.)* John Wiley & Sons, INC, New York.
- Ahmad, A., & Dey, L. (2007). *A K-mean clustering algorithm for mixed numeric and categorical data. Data and Knowledge Engineering Elsevier Publication vol 63*, 503-527.
- Alvionita. (2017). *Metode Ensemble ROCK dan SWFM untuk Pengelompokan Data Campuran Numerik dan Kategorik pada Kasus Akses Jeruk*. Institut Teknologi Sepuluh Nopember. Surabaya.
- Anderson, T. W., & Sclove, S. L. (1974). *Introductory Statistical Analysis*. Houghton Mifflin, Boston.
- Arikunto, Suharsimi. (2002). *Metodologi Penelitian*. PT. Rineka Cipta. Jakarta.
- Barnet, G. (1992). *Portofolio Us in Educational Leadership Preparation Program from Teory to Practise*. New York: Human Sciences Press.
- BPS. (2015). *Statistik Pendidikan 2015*. Pusat Statistik. Jakarta
- Deng S, He Z, & Xu X. (2005). *Clustering mixed numeric and categorical data: A cluster ensemble approach*. Arxiv preprint cs/0509011.
- Depdiknas. (2003). *Undang-Undang No. 20 tahun 2003 tentang Sistem Pendidikan Nasional*. Jakarta: Depdiknas.
- Hair, J. F., Black, W. C., Babin, J. B., dan Anderson, E. R., (2009). *Multivariate Data Analysis (7 ed.)*. New Jersey: Prentice Hall, Inc.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). *On Clustering Validation Techniques*. Journal of Intelligent Information Systems.
- Henlita, Sisca., & Handayani. (2013). *Tingkat Pelayanan Fasilitas Pendidikan Sekolah Menengah Tingkat Atas di Kabupaten Sidoarjo*. Institut Teknologi Sepuluh Nopember. Surabaya.
- Huang Z, Ng MK. (2005). *A fuzzy k-modes algorithm for clustering categorical data*. Manage Inf. Principles Ltd. Melbourne, Vic. vol. 7 pp. 446-452.

- Johnson, R.A., & Winchern, D. W. (2007). *Applied Multivariate Statistical Analysis (7 ed.)*. New Jersey: Pearson Education Inc.
- Kemendikbud. (2013). *Permendikbud No 64 tentang Standar Kompetensi Lulusan Pendidikan Dasar dan Menengah*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Kemendikbud. (2013). *Permendikbud No 66 tentang Standar Penilaian Pendidikan*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Maimon, Oded., & Rokach, Lior. (2010). *Data mining and knowledge Discovery Handbook*.
- Reddy, MVJ., & Kavitha B. (2012). *Clustering the mixed Numerical and Categorical Dataset using Similarity Weight and Filter Method*. vol 5, no 1.
- Rokach, L. (2010). *A survey of Clustering Algorithms*. In O. Maimon & L. Rokach (Eds), *Data Mining and Knowledge Discovery Handbook (2nd ed)*. Springer Science Business Media, LLC.
- Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Willey and Sons, Inc.
- Suryanto. (1988). *Metode Statistika Multivariat*. Yogyakarta: IKIP Yogyakarta.
- Wang X, & Wilkes DM. (2009). *A Divide-and-Conquer Approach for Minimum Spanning Tree-Based Clustering*. *IEEE Knowledge and Data Engineering Transactions*, vol. 21.

Lampiran 1. Data Pengamatan Variabel Numerik

No	Nama Sekolah	Rasio Siswa/Rombel (X1)	Jumlah Guru Tetap (X2)	Jumlah Guru non-tetap (X3)	Jumlah lulusan /kelas (X4)	Luas lahan (X5)	Daya listrik (X6)
1	SMAS ULUL AL ABAB	21.100	21	0	36.000	2020	46000
2	SMAS ISLAM SIDOARJO	25.833	7	6	21.500	2034	7700
3	SMAS AL FATAH	22.333	8	4	16.000	200	2200
4	SMAS AL AHMAD	14.833	10	0	16.500	1255	2200
5	SMAS ANTARTIKA	45.026	53	1	35.692	25438	32000
6	SMAS TPI PORONG	14.333	6	0	15.000	1001	2400
7	SMAN OLAH RAGA SIDOARJO	22.500	12	1	15.667	38000	41000
8	SMAS DHARMA WANITA 3	36.667	5	2	46.000	3100	2200
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50	SMAS DHARMA WANITA 2	24.000	12	0	40.000	3000	1300
51	SMAS WACHID HASYIM 3	30.000	9	0	18.000	4025	2200

Lampiran 2. Data Pengamatan Variabel Kategorik

No	Nama Sekolah	Akreditasi (X7)	Status Sekolah (X8)	Waktu Penyelenggaraan (X9)
1	SMAS ULUL AL ABAB	0	1	0
2	SMAS ISLAM SIDOARJO	0	1	0
3	SMAS AL FATAH	1	1	0
4	SMAS AL AHMAD	1	1	1
5	SMAS ANTARTIKA	0	1	0
6	SMAS TPI PORONG	2	1	0
7	SMAN OLAH RAGA SIDOARJO	3	0	0
8	SMAS DHARMA WANITA 3	2	1	1
9	SMAS PERSATUAN TULANGAN	0	1	0
⋮	⋮	⋮	⋮	⋮
51	SMAS WACHID HASYIM 3	0	1	0

Lampiran 3. *Syntax* Analisis Kelompok Data Numerik dengan Jumlah Kelompok 2

```

datanumerik = data.frame(data$X1, data$X2, data$X3, data$X4,
data$X5, data$X6)
d=dist(datanumerik, method="euclidean")
fit.sing=hclust(d,method="single")
fit.comp=hclust(d, method="complete")
fit.aver=hclust(d,method="average")
single=cutree(fit.sing,k=2)
complete=cutree(fit.comp,k=2)
average=cutree(fit.aver,k=2)
hasil=data.frame(single, complete, average)
y=c((data$X1-mean(data$X1))^2,(data$X2-mean(data$X2))^2, (data$X3-
mean(data$X3))^2, (data$X4-mean(data$X4))^2, (data$X5-mean(data$X5))^2,
(data$X6-mean(data$X6))^2)
SST=sum(y)
#single
kelompok.single=hasil$single
data.single=data.frame(single, data$X1,data$X2, data$X3, data$X4, data$X5,
data$X6)
data.single.sort=data.single[order(data.single$single),]
a=length(data.single.sort$single[data.single.sort$single==1])
b=length(data.single.sort$single[data.single.sort$single==2])
data.c1.single=data.single.sort[1:a,]
data.c2.single=data.single.sort[(a+1):51,]
xlc1bar.single=matrix(c(rep(mean(data.c1.single$data.X1), a),
rep(mean(data.c1.single$data.X2),a),
rep(mean(data.c1.single$data.X3),a),rep(mean(data.c1.single$data.X4),a),
rep(mean(data.c1.single$data.X5),a), rep(mean(data.c1.single$data.X6),a)),
ncol=6)
xlc2bar.single=matrix(c(rep(mean(data.c2.single$data.X1),b),
rep(mean(data.c2.single$data.X2),b),
rep(mean(data.c2.single$data.X3),b),rep(mean(data.c2.single$data.X4),b),
rep(mean(data.c2.single$data.X5),b), rep(mean(data.c2.single$data.X6),b)),
ncol=6)
xil.c1.single=matrix(c(data.c1.single$data.X1,data.c1.single$data.X2,data.c1.single
e$data.X3,data.c1.single$data.X4, data.c1.single$data.X5,
data.c1.single$data.X6),ncol=6)

```

```

xil.c2.single=matrix(c(data.c2.single$data.X1,data.c2.single$data.X2,data.c2.single
gle$data.X3,data.c2.single$data.X4,data.c2.single$data.X5,data.c2.single$data.
X6),ncol=6)
x1.single=sum((xil.c1.single-xlc1bar.single)^2)
x2.single=sum((xil.c2.single-xlc2bar.single)^2)
SSW.single.c2=sum(x1.single,x2.single)
SSB.single.c2=SST-SSW.single.c2
Rsquare.single.c2=SSB.single.c2/SST
hasil.single2=c(SST, SSW.single.c2, SSB.single.c2, Rsquare.single.c2)
#complete
kelompok.complete=hasil$complete
data.complete=data.frame(complete, data$X1, data$X2, data$X3, data$X4,
data$X5, data$X6)
data.complete.sort=data.complete[order(data.complete$complete),]
a=length(data.complete.sort$complete[data.complete.sort$complete==1])
b=length(data.complete.sort$complete[data.complete.sort$complete==2])
data.c1.complete=data.complete.sort[1:a,]
data.c2.complete=data.complete.sort[(a+1):51,]
xlc1bar.complete=matrix(c(rep(mean(data.c1.complete$data.X1), a),
rep(mean(data.c1.complete$data.X2),a),
rep(mean(data.c1.complete$data.X3),a),rep(mean(data.c1.complete$data.X4),a),
rep(mean(data.c1.complete$data.X5),a),
rep(mean(data.c1.complete$data.X6),a)), ncol=6)
xlc2bar.complete=matrix(c(rep(mean(data.c2.complete$data.X1), b),
rep(mean(data.c2.complete$data.X2),b),
rep(mean(data.c2.complete$data.X3),b),rep(mean(data.c2.complete$data.X4),b)
, rep(mean(data.c2.complete$data.X5),b),
rep(mean(data.c2.complete$data.X6),b)), ncol=6)
xil.c1.complete=matrix(c(data.c1.complete$data.X1,data.c1.complete$data.X2,d
ata.c1.complete$data.X3,data.c1.complete$data.X4,data.c1.complete$data.X5,d
ata.c1.complete$data.X6 ),ncol=6)
xil.c2.complete=matrix(c(data.c2.complete$data.X1,data.c2.complete$data.X2,d
ata.c2.complete$data.X3,data.c2.complete$data.X4,data.c2.complete$data.X5,d
ata.c2.complete$data.X6),ncol=6)

```

```

xil.c1.complete=matrix(c(data.c1.complete$data.X1,data.c1.complete$data.X2,
data.c1.complete$data.X3,data.c1.complete$data.X4,data.c1.complete$data.X5,
data.c1.complete$data.X6 ),ncol=6)
xil.c2.complete=matrix(c(data.c2.complete$data.X1,data.c2.complete$data.X2,
data.c2.complete$data.X3,data.c2.complete$data.X4,data.c2.complete$data.X5,
data.c2.complete$data.X6),ncol=6)
x1.complete=sum((xil.c1.complete-x1c1bar.complete)^2)
x2.complete=sum((xil.c2.complete-x1c2bar.complete)^2)
SSW.complete.c2=sum(x1.complete,x2.complete)
SSB.complete.c2=SST-SSW.complete.c2
Rsquare.complete.c2=SSB.complete.c2/SST
hasil.complete2=c(SST, SSW.complete.c2, SSB.complete.c2,
Rsquare.complete.c2)
#average
kelompok.average=hasil$average
data.average=data.frame(average, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.average.sort=data.average[order(data.average$average),]
a=length(data.average.sort$average[data.average.sort$average==1])
b=length(data.average.sort$average[data.average.sort$average==2])
data.c1.average=data.average.sort[1:a,]
data.c2.average=data.average.sort[(a+1):51,]
x1c1bar.average=matrix(c(rep(mean(data.c1.average$data.X1), a),
rep(mean(data.c1.average$data.X2),a),
rep(mean(data.c1.average$data.X3),a),rep(mean(data.c1.average$data.X4),a),re
p(mean(data.c1.average$data.X5),a),rep(mean(data.c1.average$data.X6),a)),
ncol=6)
x1c2bar.average=matrix(c(rep(mean(data.c2.average$data.X1), b),
rep(mean(data.c2.average$data.X2),b),
rep(mean(data.c2.average$data.X3),b),rep(mean(data.c2.average$data.X4),b),re
p(mean(data.c2.average$data.X5),b),rep(mean(data.c2.average$data.X6),b)),
ncol=6)
xil.c1.average=matrix(c(data.c1.average$data.X1,data.c1.average$data.X2,data.
c1.average$data.X3,data.c1.average$data.X4,data.c1.average$data.X5,data.c1.a
verage$data.X6),ncol=6)

```

```
xil.c2.average=matrix(c(data.c2.average$data.X1,data.c2.average$data.X2,data.c2.average$data.X3,data.c2.average$data.X4,data.c2.average$data.X5,data.c2.average$data.X6),ncol=6)
x1.average=sum((xil.c1.average-xlc1bar.average)^2)
x2.average=sum((xil.c2.average-xlc2bar.average)^2)
SSW.average.c2=sum(x1.average,x2.average)
SSB.average.c2=SST-SSW.average.c2
Rsquare.average.c2=SSB.average.c2/SST
hasil.average2=c(SST, SSW.average.c2, SSB.average.c2, Rsquare.average.c2)
#ringkasan
hasil2=rbind(hasil.single2,hasil.complete2,hasil.average2)
colnames(hasil2)=c("SST", "SSW", "SSB", "R-Square")
print ("Hasil Pengelompokkan")
print (hasil)
print ("Hasil R-Square")
print (hasil2)
```

Lampiran 4. *Syntax* Analisis Kelompok Data Numerik dengan Jumlah Kelompok 3

```

datanumerik = data.frame(data$X1, data$X2, data$X3, data$X4,
data$X5, data$X6)
d=dist(datanumerik, method="euclidean")
fit.sing=hclust(d,method="single")
fit.comp=hclust(d, method="complete")
fit.aver=hclust(d,method="average")
single=cutree(fit.sing,k=3)
complete=cutree(fit.comp,k=3)
average=cutree(fit.aver,k=3)
hasil=data.frame(single, complete, average)
y=c((data$X1-mean(data$X1))^2,(data$X2-mean(data$X2))^2, (data$X3-
mean(data$X3))^2, (data$X4-mean(data$X4))^2, (data$X5-mean(data$X5))^2,
(data$X6-mean(data$X6))^2)
SST=sum(y)
#single
kelompok.single=hasil$single
data.single=data.frame(single, data$X1,data$X2, data$X3, data$X4, data$X5,
data$X6)
data.single.sort=data.single[order(data.single$single),]
a=length(data.single.sort$single[data.single.sort$single==1])
b=length(data.single.sort$single[data.single.sort$single==2])
c=length(data.single.sort$single[data.single.sort$single==3])
data.c1.single=data.single.sort[1:a,]
data.c2.single=data.single.sort[(a+1):(a+b),]
data.c3.single=data.single.sort[(a+b+1):51,]
xlc1bar.single=matrix(c(rep(mean(data.c1.single$data.X1), a),
rep(mean(data.c1.single$data.X2),a),
rep(mean(data.c1.single$data.X3),a),rep(mean(data.c1.single$data.X4),a),
rep(mean(data.c1.single$data.X5),a), rep(mean(data.c1.single$data.X6),a)),
ncol=6)
xlc2bar.single=matrix(c(rep(mean(data.c2.single$data.X1), b),
rep(mean(data.c2.single$data.X2),b),
rep(mean(data.c2.single$data.X3),b),rep(mean(data.c2.single$data.X4),b),
rep(mean(data.c2.single$data.X5),b), rep(mean(data.c2.single$data.X6),b)),
ncol=6)

```

```

xlc3bar.single=matrix(c(rep(mean(data.c3.single$data.X1), c),
rep(mean(data.c3.single$data.X2),c),
rep(mean(data.c3.single$data.X3),c),rep(mean(data.c3.single$data.X4),c),
rep(mean(data.c3.single$data.X5),c), rep(mean(data.c3.single$data.X6),c)),
ncol=6)

xil.c1.single=matrix(c(data.c1.single$data.X1,data.c1.single$data.X2,data.c1.single$data.X3,data.c1.single$data.X4, data.c1.single$data.X5,
data.c1.single$data.X6),ncol=6)

xil.c2.single=matrix(c(data.c2.single$data.X1,data.c2.single$data.X2,data.c2.single$data.X3,data.c2.single$data.X4, data.c2.single$data.X5,
data.c2.single$data.X6),ncol=6)

xil.c3.single=matrix(c(data.c3.single$data.X1,data.c3.single$data.X2,data.c3.single$data.X3,data.c3.single$data.X4, data.c3.single$data.X5,
data.c3.single$data.X6),ncol=6)
x1.single=sum((xil.c1.single-xlc1bar.single)^2)
x2.single=sum((xil.c2.single-xlc2bar.single)^2)
x3.single=sum((xil.c3.single-xlc3bar.single)^2)
SSW.single.c3=sum(x1.single,x2.single,x3.single)
SSB.single.c3=SST-SSW.single.c3
Rsquare.single.c3=SSB.single.c3/SST
hasil.single3=c(SST, SSW.single.c3, SSB.single.c3, Rsquare.single.c3)
#complete
kelompok.complete=hasil$complete
data.complete=data.frame(complete, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.complete.sort=data.complete[order(data.complete$complete),]
a=length(data.complete.sort$complete[data.complete.sort$complete==1])
b=length(data.complete.sort$complete[data.complete.sort$complete==2])
c=length(data.complete.sort$complete[data.complete.sort$complete==3])

```

```

data.c1.complete=data.complete.sort[1:a,]
data.c2.complete=data.complete.sort[(a+1):(a+b),]
data.c3.complete=data.complete.sort[(a+b+1):51,]
xlc1bar.complete=matrix(c(rep(mean(data.c1.complete$data.X1), a),
rep(mean(data.c1.complete$data.X2),a),
rep(mean(data.c1.complete$data.X3),a),rep(mean(data.c1.complete$data.X4),a)
, rep(mean(data.c1.complete$data.X5),a),
rep(mean(data.c1.complete$data.X6),a)), ncol=6)
xlc2bar.complete=matrix(c(rep(mean(data.c2.complete$data.X1), b),
rep(mean(data.c2.complete$data.X2),b),
rep(mean(data.c2.complete$data.X3),b),rep(mean(data.c2.complete$data.X4),b)
, rep(mean(data.c2.complete$data.X5),b),
rep(mean(data.c2.complete$data.X6),b)), ncol=6)
xlc3bar.complete=matrix(c(rep(mean(data.c3.complete$data.X1), c),
rep(mean(data.c3.complete$data.X2),c),
rep(mean(data.c3.complete$data.X3),c),rep(mean(data.c3.complete$data.X4),c)
, rep(mean(data.c3.complete$data.X5),c),
rep(mean(data.c3.complete$data.X6),c)), ncol=6)

xil.c1.complete=matrix(c(data.c1.complete$data.X1,data.c1.complete$data.X2,
data.c1.complete$data.X3,data.c1.complete$data.X4,data.c1.complete$data.X5,
data.c1.complete$data.X6 ),ncol=6)

xil.c2.complete=matrix(c(data.c2.complete$data.X1,data.c2.complete$data.X2,
data.c2.complete$data.X3,data.c2.complete$data.X4,data.c2.complete$data.X5,
data.c2.complete$data.X6),ncol=6)

xil.c3.complete=matrix(c(data.c3.complete$data.X1,data.c3.complete$data.X2,
data.c3.complete$data.X3,data.c3.complete$data.X4,data.c3.complete$data.X5,
data.c3.complete$data.X6),ncol=6)
x1.complete=sum((xil.c1.complete-xlc1bar.complete)^2)
x2.complete=sum((xil.c2.complete-xlc2bar.complete)^2)
x3.complete=sum((xil.c3.complete-xlc3bar.complete)^2)

```



```

SSW.complete.c3=sum(x1.complete,x2.complete,x3.complete)
SSB.complete.c3=SST-SSW.complete.c3
Rsquare.complete.c3=SSB.complete.c3/SST
hasil.complete3=c(SST, SSW.complete.c3, SSB.complete.c3,
Rsquare.complete.c3)
#average
kelompok.average=hasil$average
data.average=data.frame(average, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.average.sort=data.average[order(data.average$average),]
a=length(data.average.sort$average[data.average.sort$average==1])
b=length(data.average.sort$average[data.average.sort$average==2])
c=length(data.average.sort$average[data.average.sort$average==3])
data.c1.average=data.average.sort[1:a,]
data.c2.average=data.average.sort[(a+1):(a+b),]
data.c3.average=data.average.sort[(a+b+1):51,]
xlc1bar.average=matrix(c(rep(mean(data.c1.average$data.X1), a),
rep(mean(data.c1.average$data.X2),a),
rep(mean(data.c1.average$data.X3),a),rep(mean(data.c1.average$data.X4),a),re
p(mean(data.c1.average$data.X5),a),rep(mean(data.c1.average$data.X6),a)),
ncol=6)
xlc2bar.average=matrix(c(rep(mean(data.c2.average$data.X1), b),
rep(mean(data.c2.average$data.X2),b),
rep(mean(data.c2.average$data.X3),b),rep(mean(data.c2.average$data.X4),b),re
p(mean(data.c2.average$data.X5),b),rep(mean(data.c2.average$data.X6),b)),
ncol=6)
xlc3bar.average=matrix(c(rep(mean(data.c3.average$data.X1), c),
rep(mean(data.c3.average$data.X2),c),
rep(mean(data.c3.average$data.X3),c),rep(mean(data.c3.average$data.X4),c),re
p(mean(data.c3.average$data.X5),c),rep(mean(data.c3.average$data.X6),c)),
ncol=6)
xil.c1.average=matrix(c(data.c1.average$data.X1,data.c1.average$data.X2,data.
c1.average$data.X3,data.c1.average$data.X4,data.c1.average$data.X5,data.c1.a
verage$data.X6),ncol=6)
xil.c2.average=matrix(c(data.c2.average$data.X1,data.c2.average$data.X2,data.
c2.average$data.X3,data.c2.average$data.X4,data.c2.average$data.X5,data.c2.a
verage$data.X6),ncol=6)

```

```
xil.c3.average=matrix(c(data.c3.average$data.X1,data.c3.average$data.X2,data.
c3.average$data.X3,data.c3.average$data.X4,data.c3.average$data.X5,data.c3.a
verage$data.X6),ncol=6)
x1.average=sum((xil.c1.average-xlc1bar.average)^2)
x2.average=sum((xil.c2.average-xlc2bar.average)^2)
x3.average=sum((xil.c3.average-xlc3bar.average)^2)
SSW.average.c3=sum(x1.average,x2.average,x3.average)
SSB.average.c3=SST-SSW.average.c3
Rsquare.average.c3=SSB.average.c3/SST
hasil.average3=c(SST, SSW.average.c3, SSB.average.c3, Rsquare.average.c3)
#ringkasan
hasil3=rbind(hasil.single3,hasil.complete3,hasil.average3)
colnames(hasil3)=c("SST", "SSW", "SSB", "R-Square")
print ("Hasil Pengelompokkan")
print (hasil)
print ("Hasil R-Square")
print (hasil3)
```

Lampiran 5. *Syntax* Analisis Kelompok Data Numerik dengan Jumlah Kelompok 4

```

datanumerik = data.frame(data$X1, data$X2, data$X3, data$X4,
data$X5, data$X6)
d=dist(datanumerik, method="euclidean")
fit.sing=hclust(d,method="single")
fit.comp=hclust(d, method="complete")
fit.aver=hclust(d,method="average")
single=cutree(fit.sing,k=4)
complete=cutree(fit.comp,k=4)
average=cutree(fit.aver,k=4)
hasil=data.frame(single, complete, average)
y=c((data$X1-mean(data$X1))^2,(data$X2-mean(data$X2))^2, (data$X3-
mean(data$X3))^2, (data$X4-mean(data$X4))^2, (data$X5-mean(data$X5))^2,
(data$X6-mean(data$X6))^2)
SST=sum(y)
#single
kelompok.single=hasil$single
data.single=data.frame(single, data$X1,data$X2, data$X3, data$X4, data$X5,
data$X6)
data.single.sort=data.single[order(data.single$single),]
a=length(data.single.sort$single[data.single.sort$single==1])
b=length(data.single.sort$single[data.single.sort$single==2])
c=length(data.single.sort$single[data.single.sort$single==3])
d=length(data.single.sort$single[data.single.sort$single==4])
data.c1.single=data.single.sort[1:a,]
data.c2.single=data.single.sort[(a+1):(a+b),]
data.c3.single=data.single.sort[(a+b+1):(a+b+c),]
data.c4.single=data.single.sort[(a+b+c+1):51,]
xlc1bar.single=matrix(c(rep(mean(data.c1.single$data.X1), a),
rep(mean(data.c1.single$data.X2),a),
rep(mean(data.c1.single$data.X3),a),rep(mean(data.c1.single$data.X4),a),
rep(mean(data.c1.single$data.X5),a), rep(mean(data.c1.single$data.X6),a)),
ncol=6)
xlc2bar.single=matrix(c(rep(mean(data.c2.single$data.X1), b),
rep(mean(data.c2.single$data.X2),b),
rep(mean(data.c2.single$data.X3),b),rep(mean(data.c2.single$data.X4),b),
rep(mean(data.c2.single$data.X5),b), rep(mean(data.c2.single$data.X6),b)),
ncol=6)

```

```

xlc3bar.single=matrix(c(rep(mean(data.c3.single$data.X1), c),
rep(mean(data.c3.single$data.X2),c),
rep(mean(data.c3.single$data.X3),c),rep(mean(data.c3.single$data.X4),c),
rep(mean(data.c3.single$data.X5),c), rep(mean(data.c3.single$data.X6),c)),
ncol=6)
xlc4bar.single=matrix(c(rep(mean(data.c4.single$data.X1), d),
rep(mean(data.c4.single$data.X2),d),
rep(mean(data.c4.single$data.X3),d),rep(mean(data.c4.single$data.X4),d),
rep(mean(data.c4.single$data.X5),d), rep(mean(data.c4.single$data.X6),d)),
ncol=6)
xil.c1.single=matrix(c(data.c1.single$data.X1,data.c1.single$data.X2,data.c1.single
$data.X3,data.c1.single$data.X4, data.c1.single$data.X5,
data.c1.single$data.X6),ncol=6)
xil.c2.single=matrix(c(data.c2.single$data.X1,data.c2.single$data.X2,data.c2.single
$data.X3,data.c2.single$data.X4, data.c2.single$data.X5,
data.c2.single$data.X6),ncol=6)
xil.c3.single=matrix(c(data.c3.single$data.X1,data.c3.single$data.X2,data.c3.single
$data.X3,data.c3.single$data.X4, data.c3.single$data.X5,
data.c3.single$data.X6),ncol=6)
xil.c4.single=matrix(c(data.c4.single$data.X1,data.c4.single$data.X2,data.c4.single
$data.X3,data.c4.single$data.X4, data.c4.single$data.X5,
data.c4.single$data.X6),ncol=6)
x1.single=sum((xil.c1.single-xlc1bar.single)^2)
x2.single=sum((xil.c2.single-xlc2bar.single)^2)
x3.single=sum((xil.c3.single-xlc3bar.single)^2)
x4.single=sum((xil.c4.single-xlc4bar.single)^2)
SSW.single.c4=sum(x1.single,x2.single,x3.single,x4.single)
SSB.single.c4=SST-SSW.single.c4
Rsquare.single.c4=SSB.single.c4/SST
hasil.single4=c(SST, SSW.single.c4, SSB.single.c4, Rsquare.single.c4)
kelompok.complete=hasil$complete
data.complete=data.frame(complete, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.complete.sort=data.complete[order(data.complete$complete),]
a=length(data.complete.sort$complete[data.complete.sort$complete==1])
b=length(data.complete.sort$complete[data.complete.sort$complete==2])
c=length(data.complete.sort$complete[data.complete.sort$complete==3])
d=length(data.complete.sort$complete[data.complete.sort$complete==4])

```

```

data.c1.complete=data.complete.sort[1:a,]
data.c2.complete=data.complete.sort[(a+1):(a+b),]
data.c3.complete=data.complete.sort[(a+b+1):(a+b+c),]
data.c4.complete=data.complete.sort[(a+b+c+1):51,]
xlc1bar.complete=matrix(c(rep(mean(data.c1.complete$data.X1), a),
rep(mean(data.c1.complete$data.X2),a),
rep(mean(data.c1.complete$data.X3),a),rep(mean(data.c1.complete$data.X4),a)
, rep(mean(data.c1.complete$data.X5),a),
rep(mean(data.c1.complete$data.X6),a)), ncol=6)
xlc2bar.complete=matrix(c(rep(mean(data.c2.complete$data.X1), b),
rep(mean(data.c2.complete$data.X2),b),
rep(mean(data.c2.complete$data.X3),b),rep(mean(data.c2.complete$data.X4),b)
, rep(mean(data.c2.complete$data.X5),b),
rep(mean(data.c2.complete$data.X6),b)), ncol=6)
xlc3bar.complete=matrix(c(rep(mean(data.c3.complete$data.X1), c),
rep(mean(data.c3.complete$data.X2),c),
rep(mean(data.c3.complete$data.X3),c),rep(mean(data.c3.complete$data.X4),c)
, rep(mean(data.c3.complete$data.X5),c),
rep(mean(data.c3.complete$data.X6),c)), ncol=6)
xlc4bar.complete=matrix(c(rep(mean(data.c4.complete$data.X1), d),
rep(mean(data.c4.complete$data.X2),d),
rep(mean(data.c4.complete$data.X3),d),rep(mean(data.c4.complete$data.X4),d)
, rep(mean(data.c4.complete$data.X5),d),
rep(mean(data.c4.complete$data.X6),d)), ncol=6)
xil.c1.complete=matrix(c(data.c1.complete$data.X1,data.c1.complete$data.X2,
data.c1.complete$data.X3,data.c1.complete$data.X4,data.c1.complete$data.X5,
data.c1.complete$data.X6 ),ncol=6)
xil.c2.complete=matrix(c(data.c2.complete$data.X1,data.c2.complete$data.X2,
data.c2.complete$data.X3,data.c2.complete$data.X4,data.c2.complete$data.X5,
data.c2.complete$data.X6),ncol=6)
xil.c3.complete=matrix(c(data.c3.complete$data.X1,data.c3.complete$data.X2,
data.c3.complete$data.X3,data.c3.complete$data.X4,data.c3.complete$data.X5,
data.c3.complete$data.X6),ncol=6)
xil.c4.complete=matrix(c(data.c4.complete$data.X1,data.c4.complete$data.X2,
data.c4.complete$data.X3,data.c4.complete$data.X4,data.c4.complete$data.X5,
data.c4.complete$data.X6),ncol=6)

```

```

x1.complete=sum((xil.c1.complete-xlc1bar.complete)^2)
x2.complete=sum((xil.c2.complete-xlc2bar.complete)^2)
x3.complete=sum((xil.c3.complete-xlc3bar.complete)^2)
x4.complete=sum((xil.c4.complete-xlc4bar.complete)^2)
SSW.complete.c4=sum(x1.complete,x2.complete,x3.complete,x4.complete)
SSB.complete.c4=SST-SSW.complete.c4
Rsquare.complete.c4=SSB.complete.c4/SST
hasil.complete4=c(SST, SSW.complete.c4, SSB.complete.c4,
Rsquare.complete.c4)
#average
kelompok.average=hasil$average
data.average=data.frame(average, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.average.sort=data.average[order(data.average$average),]
a=length(data.average.sort$average[data.average.sort$average==1])
b=length(data.average.sort$average[data.average.sort$average==2])
c=length(data.average.sort$average[data.average.sort$average==3])
d=length(data.average.sort$average[data.average.sort$average==4])
data.c1.average=data.average.sort[1:a,]
data.c2.average=data.average.sort[(a+1):(a+b),]
data.c3.average=data.average.sort[(a+b+1):(a+b+c),]
data.c4.average=data.average.sort[(a+b+c+1):51,]
xlc1bar.average=matrix(c(rep(mean(data.c1.average$data.X1), a),
rep(mean(data.c1.average$data.X2),a),
rep(mean(data.c1.average$data.X3),a),rep(mean(data.c1.average$data.X4),a),re
p(mean(data.c1.average$data.X5),a),rep(mean(data.c1.average$data.X6),a)),
ncol=6)
xlc2bar.average=matrix(c(rep(mean(data.c2.average$data.X1), b),
rep(mean(data.c2.average$data.X2),b),
rep(mean(data.c2.average$data.X3),b),rep(mean(data.c2.average$data.X4),b),re
p(mean(data.c2.average$data.X5),b),rep(mean(data.c2.average$data.X6),b)),
ncol=6)
xlc3bar.average=matrix(c(rep(mean(data.c3.average$data.X1), c),
rep(mean(data.c3.average$data.X2),c),
rep(mean(data.c3.average$data.X3),c),rep(mean(data.c3.average$data.X4),c),re
p(mean(data.c3.average$data.X5),c),rep(mean(data.c3.average$data.X6),c)),
ncol=6)

```

```

xlc4bar.average=matrix(c(rep(mean(data.c4.average$data.X1), d),
rep(mean(data.c4.average$data.X2),d),
rep(mean(data.c4.average$data.X3),d),rep(mean(data.c4.average$data.X4),d),re
p(mean(data.c4.average$data.X5),d),rep(mean(data.c4.average$data.X6),d)),
ncol=6)
xil.c1.average=matrix(c(data.c1.average$data.X1,data.c1.average$data.X2,data.
c1.average$data.X3,data.c1.average$data.X4,data.c1.average$data.X5,data.c1.a
verage$data.X6),ncol=6)
xil.c2.average=matrix(c(data.c2.average$data.X1,data.c2.average$data.X2,data.
c2.average$data.X3,data.c2.average$data.X4,data.c2.average$data.X5,data.c2.a
verage$data.X6),ncol=6)
xil.c3.average=matrix(c(data.c3.average$data.X1,data.c3.average$data.X2,data.
c3.average$data.X3,data.c3.average$data.X4,data.c3.average$data.X5,data.c3.a
verage$data.X6),ncol=6)
xil.c4.average=matrix(c(data.c4.average$data.X1,data.c4.average$data.X2,data.
c4.average$data.X3,data.c4.average$data.X4,data.c4.average$data.X5,data.c4.a
verage$data.X6),ncol=6)
x1.average=sum((xil.c1.average-xlc1bar.average)^2)
x2.average=sum((xil.c2.average-xlc2bar.average)^2)
x3.average=sum((xil.c3.average-xlc3bar.average)^2)
x4.average=sum((xil.c4.average-xlc4bar.average)^2)
SSW.average.c4=sum(x1.average,x2.average,x3.average,x4.average)
SSB.average.c4=SST-SSW.average.c4
Rsquare.average.c4=SSB.average.c4/SST
hasil.average4=c(SST, SSW.average.c4, SSB.average.c4, Rsquare.average.c4)
#ringkasan
hasil4=rbind(hasil.single4,hasil.complete4,hasil.average4)
colnames(hasil4)=c("SST", "SSW", "SSB", "R-Square")
print ("Hasil Pengelompokkan")
print (hasil)
print ("Hasil R-Square")
print (hasil4)

```

Lampiran 6. *Syntax* Analisis Kelompok Data Numerik dengan Jumlah Kelompok 5

```

datanumerik = data.frame(data$X1, data$X2, data$X3, data$X4,
data$X5, data$X6)
d=dist(datanumerik, method="euclidean")
fit.sing=hclust(d,method="single")
fit.comp=hclust(d, method="complete")
fit.aver=hclust(d,method="average")
single=cutree(fit.sing,k=5)
complete=cutree(fit.comp,k=5)
average=cutree(fit.aver,k=5)
hasil=data.frame(single, complete, average)
y=c(((data$X1-mean(data$X1))^2,(data$X2-mean(data$X2))^2, (data$X3-
mean(data$X3))^2, (data$X4-mean(data$X4))^2, (data$X5-mean(data$X5))^2,
(data$X6-mean(data$X6))^2)
SST=sum(y)
#single
kelompok.single=hasil$single
data.single=data.frame(single, data$X1,data$X2, data$X3, data$X4, data$X5,
data$X6)
data.single.sort=data.single[order(data.single$single),]
a=length(data.single.sort$single[data.single.sort$single==1])
b=length(data.single.sort$single[data.single.sort$single==2])
c=length(data.single.sort$single[data.single.sort$single==3])
d=length(data.single.sort$single[data.single.sort$single==4])
e=length(data.single.sort$single[data.single.sort$single==5])
data.c1.single=data.single.sort[1:a,]
data.c2.single=data.single.sort[(a+1):(a+b),]
data.c3.single=data.single.sort[(a+b+1):(a+b+c),]
data.c4.single=data.single.sort[(a+b+c+1):(a+b+c+d),]
data.c5.single=data.single.sort[(a+b+c+d+1):51,]
x1c1bar.single=matrix(c(rep(mean(data.c1.single$data.X1), a),
rep(mean(data.c1.single$data.X2),a),
rep(mean(data.c1.single$data.X3),a),rep(mean(data.c1.single$data.X4),a),
rep(mean(data.c1.single$data.X5),a), rep(mean(data.c1.single$data.X6),a)),
ncol=6)

```



```

xlc2bar.single=matrix(c(rep(mean(data.c2.single$data.X1), b),
rep(mean(data.c2.single$data.X2),b),
rep(mean(data.c2.single$data.X3),b),rep(mean(data.c2.single$data.X4),b),
rep(mean(data.c2.single$data.X5),b), rep(mean(data.c2.single$data.X6),b)),
ncol=6)
xlc3bar.single=matrix(c(rep(mean(data.c3.single$data.X1), c),
rep(mean(data.c3.single$data.X2),c),
rep(mean(data.c3.single$data.X3),c),rep(mean(data.c3.single$data.X4),c),
rep(mean(data.c3.single$data.X5),c), rep(mean(data.c3.single$data.X6),c)),
ncol=6)
xlc4bar.single=matrix(c(rep(mean(data.c4.single$data.X1), d),
rep(mean(data.c4.single$data.X2),d),
rep(mean(data.c4.single$data.X3),d),rep(mean(data.c4.single$data.X4),d),
rep(mean(data.c4.single$data.X5),d), rep(mean(data.c4.single$data.X6),d)),
ncol=6)
xlc5bar.single=matrix(c(rep(mean(data.c5.single$data.X1), e),
rep(mean(data.c5.single$data.X2),e),
rep(mean(data.c5.single$data.X3),e),rep(mean(data.c5.single$data.X4),e),
rep(mean(data.c5.single$data.X5),e), rep(mean(data.c5.single$data.X6),e)),
ncol=6)
xil.c1.single=matrix(c(data.c1.single$data.X1,data.c1.single$data.X2,data.c1.single
gle$data.X3,data.c1.single$data.X4, data.c1.single$data.X5,
data.c1.single$data.X6),ncol=6)
xil.c2.single=matrix(c(data.c2.single$data.X1,data.c2.single$data.X2,data.c2.single
gle$data.X3,data.c2.single$data.X4, data.c2.single$data.X5,
data.c2.single$data.X6),ncol=6)
xil.c3.single=matrix(c(data.c3.single$data.X1,data.c3.single$data.X2,data.c3.single
gle$data.X3,data.c3.single$data.X4, data.c3.single$data.X5,
data.c3.single$data.X6),ncol=6)
xil.c4.single=matrix(c(data.c4.single$data.X1,data.c4.single$data.X2,data.c4.single
gle$data.X3,data.c4.single$data.X4, data.c4.single$data.X5,
data.c4.single$data.X6),ncol=6)
xil.c5.single=matrix(c(data.c5.single$data.X1,data.c5.single$data.X2,data.c5.single
gle$data.X3,data.c5.single$data.X4, data.c5.single$data.X5,
data.c5.single$data.X6),ncol=6)
x1.single=sum((xil.c1.single-xlc1bar.single)^2)
x2.single=sum((xil.c2.single-xlc2bar.single)^2)
x3.single=sum((xil.c3.single-xlc3bar.single)^2)

```

```

x4.single=sum((xil.c4.single-xlc4bar.single)^2)
x5.single=sum((xil.c5.single-xlc5bar.single)^2)
SSW.single.c5=sum(x1.single,x2.single,x3.single,x4.single,x5.single)
SSB.single.c5=SST-SSW.single.c5
Rsquare.single.c5=SSB.single.c5/SST
hasil.single5=c(SST, SSW.single.c5, SSB.single.c5, Rsquare.single.c5)
kelompok.complete=hasil$complete
data.complete=data.frame(complete, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.complete.sort=data.complete[order(data.complete$complete),]
a=length(data.complete.sort$complete[data.complete.sort$complete==1])
b=length(data.complete.sort$complete[data.complete.sort$complete==2])
c=length(data.complete.sort$complete[data.complete.sort$complete==3])
d=length(data.complete.sort$complete[data.complete.sort$complete==4])
e=length(data.complete.sort$complete[data.complete.sort$complete==5])
data.c1.complete=data.complete.sort[1:a,]
data.c2.complete=data.complete.sort[(a+1):(a+b),]
data.c3.complete=data.complete.sort[(a+b+1):(a+b+c),]
data.c4.complete=data.complete.sort[(a+b+c+1):(a+b+c+d),]
data.c5.complete=data.complete.sort[(a+b+c+d+1):51,]
xlc1bar.complete=matrix(c(rep(mean(data.c1.complete$data.X1), a),
rep(mean(data.c1.complete$data.X2),a),
rep(mean(data.c1.complete$data.X3),a),rep(mean(data.c1.complete$data.X4),a)
, rep(mean(data.c1.complete$data.X5),a),
rep(mean(data.c1.complete$data.X6),a)), ncol=6)
xlc2bar.complete=matrix(c(rep(mean(data.c2.complete$data.X1), b),
rep(mean(data.c2.complete$data.X2),b),
rep(mean(data.c2.complete$data.X3),b),rep(mean(data.c2.complete$data.X4),b)
, rep(mean(data.c2.complete$data.X5),b),
rep(mean(data.c2.complete$data.X6),b)), ncol=6)
xlc3bar.complete=matrix(c(rep(mean(data.c3.complete$data.X1), c),
rep(mean(data.c3.complete$data.X2),c),
rep(mean(data.c3.complete$data.X3),c),rep(mean(data.c3.complete$data.X4),c)
, rep(mean(data.c3.complete$data.X5),c),
rep(mean(data.c3.complete$data.X6),c)), ncol=6)

```

```

xlc4bar.complete=matrix(c(rep(mean(data.c4.complete$data.X1), d),
rep(mean(data.c4.complete$data.X2),d),
rep(mean(data.c4.complete$data.X3),d),rep(mean(data.c4.complete$data.X4),d)
, rep(mean(data.c4.complete$data.X5),d),
rep(mean(data.c4.complete$data.X6),d)), ncol=6)
xlc5bar.complete=matrix(c(rep(mean(data.c5.complete$data.X1), e),
rep(mean(data.c5.complete$data.X2),e),
rep(mean(data.c5.complete$data.X3),e),rep(mean(data.c5.complete$data.X4),e)
, rep(mean(data.c5.complete$data.X5),e),
rep(mean(data.c5.complete$data.X6),e)), ncol=6)
xil.c1.complete=matrix(c(data.c1.complete$data.X1,data.c1.complete$data.X2,
data.c1.complete$data.X3,data.c1.complete$data.X4,data.c1.complete$data.X5,
data.c1.complete$data.X6 ),ncol=6)
xil.c2.complete=matrix(c(data.c2.complete$data.X1,data.c2.complete$data.X2,
data.c2.complete$data.X3,data.c2.complete$data.X4,data.c2.complete$data.X5,
data.c2.complete$data.X6),ncol=6)
xil.c3.complete=matrix(c(data.c3.complete$data.X1,data.c3.complete$data.X2,
data.c3.complete$data.X3,data.c3.complete$data.X4,data.c3.complete$data.X5,
data.c3.complete$data.X6),ncol=6)
xil.c4.complete=matrix(c(data.c4.complete$data.X1,data.c4.complete$data.X2,
data.c4.complete$data.X3,data.c4.complete$data.X4,data.c4.complete$data.X5,
data.c4.complete$data.X6),ncol=6)
xil.c5.complete=matrix(c(data.c5.complete$data.X1,data.c5.complete$data.X2,
data.c5.complete$data.X3,data.c5.complete$data.X4,data.c5.complete$data.X5,
data.c5.complete$data.X6),ncol=6)
x1.complete=sum((xil.c1.complete-xlc1bar.complete)^2)
x2.complete=sum((xil.c2.complete-xlc2bar.complete)^2)
x3.complete=sum((xil.c3.complete-xlc3bar.complete)^2)
x4.complete=sum((xil.c4.complete-xlc4bar.complete)^2)
x5.complete=sum((xil.c5.complete-xlc5bar.complete)^2)
SSW.complete.c5=sum(x1.complete,x2.complete,x3.complete,x4.complete,x5.c
omplete)
SSB.complete.c5=SST-SSW.complete.c5
Rsquare.complete.c5=SSB.complete.c5/SST
hasil.complete5=c(SST, SSW.complete.c5, SSB.complete.c5,
Rsquare.complete.c5)

```

```

kelompok.average=hasil$average
data.average=data.frame(average, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.average.sort=data.average[order(data.average$average),]
a=length(data.average.sort$average[data.average.sort$average==1])
b=length(data.average.sort$average[data.average.sort$average==2])
c=length(data.average.sort$average[data.average.sort$average==3])
d=length(data.average.sort$average[data.average.sort$average==4])
e=length(data.average.sort$average[data.average.sort$average==5])
data.c1.average=data.average.sort[1:a,]
data.c2.average=data.average.sort[(a+1):(a+b),]
data.c3.average=data.average.sort[(a+b+1):(a+b+c),]
data.c4.average=data.average.sort[(a+b+c+1):(a+b+c+d),]
data.c5.average=data.average.sort[(a+b+c+d+1):51,]
xlc1bar.average=matrix(c(rep(mean(data.c1.average$data.X1), a),
rep(mean(data.c1.average$data.X2),a),
rep(mean(data.c1.average$data.X3),a),rep(mean(data.c1.average$data.X4),a),re
p(mean(data.c1.average$data.X5),a),rep(mean(data.c1.average$data.X6),a)),
ncol=6)
xlc2bar.average=matrix(c(rep(mean(data.c2.average$data.X1), b),
rep(mean(data.c2.average$data.X2),b),
rep(mean(data.c2.average$data.X3),b),rep(mean(data.c2.average$data.X4),b),re
p(mean(data.c2.average$data.X5),b),rep(mean(data.c2.average$data.X6),b)),
ncol=6)
xlc3bar.average=matrix(c(rep(mean(data.c3.average$data.X1), c),
rep(mean(data.c3.average$data.X2),c),
rep(mean(data.c3.average$data.X3),c),rep(mean(data.c3.average$data.X4),c),re
p(mean(data.c3.average$data.X5),c),rep(mean(data.c3.average$data.X6),c)),
ncol=6)
xlc4bar.average=matrix(c(rep(mean(data.c4.average$data.X1), d),
rep(mean(data.c4.average$data.X2),d),
rep(mean(data.c4.average$data.X3),d),rep(mean(data.c4.average$data.X4),d),re
p(mean(data.c4.average$data.X5),d),rep(mean(data.c4.average$data.X6),d)),
ncol=6)
xlc5bar.average=matrix(c(rep(mean(data.c5.average$data.X1), e),
rep(mean(data.c5.average$data.X2),e),
rep(mean(data.c5.average$data.X3),e),rep(mean(data.c5.average$data.X4),e),re
p(mean(data.c5.average$data.X5),e),rep(mean(data.c5.average$data.X6),e)),
ncol=6)

```

```

xil.c1.average=matrix(c(data.c1.average$data.X1,data.c1.average$data.X2,data.
c1.average$data.X3,data.c1.average$data.X4,data.c1.average$data.X5,data.c1.a
verage$data.X6),ncol=6)

xil.c2.average=matrix(c(data.c2.average$data.X1,data.c2.average$data.X2,data.
c2.average$data.X3,data.c2.average$data.X4,data.c2.average$data.X5,data.c2.a
verage$data.X6),ncol=6)

xil.c3.average=matrix(c(data.c3.average$data.X1,data.c3.average$data.X2,data.
c3.average$data.X3,data.c3.average$data.X4,data.c3.average$data.X5,data.c3.a
verage$data.X6),ncol=6)

xil.c4.average=matrix(c(data.c4.average$data.X1,data.c4.average$data.X2,data.
c4.average$data.X3,data.c4.average$data.X4,data.c4.average$data.X5,data.c4.a
verage$data.X6),ncol=6)

xil.c5.average=matrix(c(data.c5.average$data.X1,data.c5.average$data.X2,data.
c5.average$data.X3,data.c5.average$data.X4,data.c5.average$data.X5,data.c5.a
verage$data.X6),ncol=6)
x1.average=sum((xil.c1.average-xlc1bar.average)^2)
x2.average=sum((xil.c2.average-xlc2bar.average)^2)
x3.average=sum((xil.c3.average-xlc3bar.average)^2)
x4.average=sum((xil.c4.average-xlc4bar.average)^2)
x5.average=sum((xil.c5.average-xlc5bar.average)^2)

SSW.average.c5=sum(x1.average,x2.average,x3.average,x4.average,x5.average
)
SSB.average.c5=SST-SSW.average.c5
Rsquare.average.c5=SSB.average.c5/SST
hasil.average5=c(SST, SSW.average.c5, SSB.average.c5, Rsquare.average.c5)
#ringkasan
hasil5=rbind(hasil.single5,hasil.complete5,hasil.average5)
colnames(hasil5)=c("SST", "SSW", "SSB", "R-Square")
print ("Hasil Pengelompokkan")
print (hasil)
print ("Hasil R-Square")
print (hasil5)

```

Lampiran 7. *Syntax* Analisis Kelompok Data Numerik dengan Jumlah Kelompok 6

```

datanumerik = data.frame(data$X1, data$X2, data$X3, data$X4,
data$X5, data$X6)
d=dist(datanumerik, method="euclidean")
fit.sing=hclust(d,method="single")
fit.comp=hclust(d, method="complete")
fit.aver=hclust(d,method="average")
single=cutree(fit.sing,k=6)
complete=cutree(fit.comp,k=6)
average=cutree(fit.aver,k=6)
hasil=data.frame(single, complete, average)
y=c((data$X1-mean(data$X1))^2,(data$X2-mean(data$X2))^2, (data$X3-
mean(data$X3))^2, (data$X4-mean(data$X4))^2, (data$X5-mean(data$X5))^2,
(data$X6-mean(data$X6))^2)
SST=sum(y)
kelompok.single=hasil$single
data.single=data.frame(single, data$X1,data$X2, data$X3, data$X4, data$X5,
data$X6)
data.single.sort=data.single[order(data.single$single),]
a=length(data.single.sort$single[data.single.sort$single==1])
b=length(data.single.sort$single[data.single.sort$single==2])
c=length(data.single.sort$single[data.single.sort$single==3])
d=length(data.single.sort$single[data.single.sort$single==4])
e=length(data.single.sort$single[data.single.sort$single==5])
f=length(data.single.sort$single[data.single.sort$single==6])
data.c1.single=data.single.sort[1:a,]
data.c2.single=data.single.sort[(a+1):(a+b),]
data.c3.single=data.single.sort[(a+b+1):(a+b+c),]
data.c4.single=data.single.sort[(a+b+c+1):(a+b+c+d),]
data.c5.single=data.single.sort[(a+b+c+d+1):(a+b+c+d+e),]
data.c6.single=data.single.sort[(a+b+c+d+e+1):51,]
xlc1bar.single=matrix(c(rep(mean(data.c1.single$data.X1), a),
rep(mean(data.c1.single$data.X2),a),
rep(mean(data.c1.single$data.X3),a),rep(mean(data.c1.single$data.X4),a),
rep(mean(data.c1.single$data.X5),a), rep(mean(data.c1.single$data.X6),a)),
ncol=6)

```

```

xlc2bar.single=matrix(c(rep(mean(data.c2.single$data.X1), b),
rep(mean(data.c2.single$data.X2),b),
rep(mean(data.c2.single$data.X3),b),rep(mean(data.c2.single$data.X4),b),
rep(mean(data.c2.single$data.X5),b), rep(mean(data.c2.single$data.X6),b)),
ncol=6)
xlc3bar.single=matrix(c(rep(mean(data.c3.single$data.X1), c),
rep(mean(data.c3.single$data.X2),c),
rep(mean(data.c3.single$data.X3),c),rep(mean(data.c3.single$data.X4),c),
rep(mean(data.c3.single$data.X5),c), rep(mean(data.c3.single$data.X6),c)),
ncol=6)
xlc4bar.single=matrix(c(rep(mean(data.c4.single$data.X1), d),
rep(mean(data.c4.single$data.X2),d),
rep(mean(data.c4.single$data.X3),d),rep(mean(data.c4.single$data.X4),d),
rep(mean(data.c4.single$data.X5),d), rep(mean(data.c4.single$data.X6),d)),
ncol=6)
xlc5bar.single=matrix(c(rep(mean(data.c5.single$data.X1), e),
rep(mean(data.c5.single$data.X2),e),
rep(mean(data.c5.single$data.X3),e),rep(mean(data.c5.single$data.X4),e),
rep(mean(data.c5.single$data.X5),e), rep(mean(data.c5.single$data.X6),e)),
ncol=6)
xlc6bar.single=matrix(c(rep(mean(data.c6.single$data.X1), f),
rep(mean(data.c6.single$data.X2),f),
rep(mean(data.c6.single$data.X3),f),rep(mean(data.c6.single$data.X4),f),
rep(mean(data.c6.single$data.X5),f), rep(mean(data.c6.single$data.X6),f)),
ncol=6)
xil.c1.single=matrix(c(data.c1.single$data.X1,data.c1.single$data.X2,data.c1.single
$data.X3,data.c1.single$data.X4, data.c1.single$data.X5,
data.c1.single$data.X6),ncol=6)
xil.c2.single=matrix(c(data.c2.single$data.X1,data.c2.single$data.X2,data.c2.single
$data.X3,data.c2.single$data.X4, data.c2.single$data.X5,
data.c2.single$data.X6),ncol=6)
xil.c3.single=matrix(c(data.c3.single$data.X1,data.c3.single$data.X2,data.c3.single
$data.X3,data.c3.single$data.X4, data.c3.single$data.X5,
data.c3.single$data.X6),ncol=6)
xil.c4.single=matrix(c(data.c4.single$data.X1,data.c4.single$data.X2,data.c4.single
$data.X3,data.c4.single$data.X4, data.c4.single$data.X5,
data.c4.single$data.X6),ncol=6)

```

```

xil.c5.single=matrix(c(data.c5.single$data.X1,data.c5.single$data.X2,data.c5.single$data.X3,data.c5.single$data.X4, data.c5.single$data.X5,
data.c5.single$data.X6),ncol=6)
xil.c6.single=matrix(c(data.c6.single$data.X1,data.c6.single$data.X2,data.c6.single$data.X3,data.c6.single$data.X4, data.c6.single$data.X5,
data.c6.single$data.X6),ncol=6)
x1.single=sum((xil.c1.single-xlc1bar.single)^2)
x2.single=sum((xil.c2.single-xlc2bar.single)^2)
x3.single=sum((xil.c3.single-xlc3bar.single)^2)
x4.single=sum((xil.c4.single-xlc4bar.single)^2)
x5.single=sum((xil.c5.single-xlc5bar.single)^2)
x6.single=sum((xil.c6.single-xlc6bar.single)^2)
SSW.single.c6=sum(x1.single,x2.single,x3.single,x4.single,x5.single,x6.single)
SSB.single.c6=SST-SSW.single.c6
Rsquare.single.c6=SSB.single.c6/SST
hasil.single6=c(SST, SSW.single.c6, SSB.single.c6, Rsquare.single.c6)
#complete
kelompok.complete=hasil$complete
data.complete=data.frame(complete, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.complete.sort=data.complete[order(data.complete$complete),]
a=length(data.complete.sort$complete[data.complete.sort$complete==1])
b=length(data.complete.sort$complete[data.complete.sort$complete==2])
c=length(data.complete.sort$complete[data.complete.sort$complete==3])
d=length(data.complete.sort$complete[data.complete.sort$complete==4])
e=length(data.complete.sort$complete[data.complete.sort$complete==5])
f=length(data.complete.sort$complete[data.complete.sort$complete==6])
data.c1.complete=data.complete.sort[1:a,]
data.c2.complete=data.complete.sort[(a+1):(a+b),]
data.c3.complete=data.complete.sort[(a+b+1):(a+b+c),]
data.c4.complete=data.complete.sort[(a+b+c+1):(a+b+c+d),]
data.c5.complete=data.complete.sort[(a+b+c+d+1):(a+b+c+d+e),]
data.c6.complete=data.complete.sort[(a+b+c+d+e+1):51,]
xlc1bar.complete=matrix(c(rep(mean(data.c1.complete$data.X1), a),
rep(mean(data.c1.complete$data.X2),a),
rep(mean(data.c1.complete$data.X3),a),rep(mean(data.c1.complete$data.X4),a)
, rep(mean(data.c1.complete$data.X5),a),
rep(mean(data.c1.complete$data.X6),a)), ncol=6)

```



```

xlc2bar.complete=matrix(c(rep(mean(data.c2.complete$data.X1), b),
rep(mean(data.c2.complete$data.X2),b),
rep(mean(data.c2.complete$data.X3),b),rep(mean(data.c2.complete$data.X4),b)
, rep(mean(data.c2.complete$data.X5),b),
rep(mean(data.c2.complete$data.X6),b)), ncol=6)
xlc3bar.complete=matrix(c(rep(mean(data.c3.complete$data.X1), c),
rep(mean(data.c3.complete$data.X2),c),
rep(mean(data.c3.complete$data.X3),c),rep(mean(data.c3.complete$data.X4),c)
, rep(mean(data.c3.complete$data.X5),c),
rep(mean(data.c3.complete$data.X6),c)), ncol=6)
xlc4bar.complete=matrix(c(rep(mean(data.c4.complete$data.X1), d),
rep(mean(data.c4.complete$data.X2),d),
rep(mean(data.c4.complete$data.X3),d),rep(mean(data.c4.complete$data.X4),d)
, rep(mean(data.c4.complete$data.X5),d),
rep(mean(data.c4.complete$data.X6),d)), ncol=6)
xlc5bar.complete=matrix(c(rep(mean(data.c5.complete$data.X1), e),
rep(mean(data.c5.complete$data.X2),e),
rep(mean(data.c5.complete$data.X3),e),rep(mean(data.c5.complete$data.X4),e)
, rep(mean(data.c5.complete$data.X5),e),
rep(mean(data.c5.complete$data.X6),e)), ncol=6)
xlc6bar.complete=matrix(c(rep(mean(data.c6.complete$data.X1), f),
rep(mean(data.c6.complete$data.X2),f),
rep(mean(data.c6.complete$data.X3),f),rep(mean(data.c6.complete$data.X4),f),
rep(mean(data.c6.complete$data.X5),f),
rep(mean(data.c6.complete$data.X6),f)), ncol=6)
xil.c1.complete=matrix(c(data.c1.complete$data.X1,data.c1.complete$data.X2,
data.c1.complete$data.X3,data.c1.complete$data.X4,data.c1.complete$data.X5,
data.c1.complete$data.X6 ),ncol=6)
xil.c2.complete=matrix(c(data.c2.complete$data.X1,data.c2.complete$data.X2,
data.c2.complete$data.X3,data.c2.complete$data.X4,data.c2.complete$data.X5,
data.c2.complete$data.X6),ncol=6)
xil.c3.complete=matrix(c(data.c3.complete$data.X1,data.c3.complete$data.X2,
data.c3.complete$data.X3,data.c3.complete$data.X4,data.c3.complete$data.X5,
data.c3.complete$data.X6),ncol=6)
xil.c4.complete=matrix(c(data.c4.complete$data.X1,data.c4.complete$data.X2,
data.c4.complete$data.X3,data.c4.complete$data.X4,data.c4.complete$data.X5,
data.c4.complete$data.X6),ncol=6)

```

```

xil.c5.complete=matrix(c(data.c5.complete$data.X1,data.c5.complete$data.X2,
data.c5.complete$data.X3,data.c5.complete$data.X4,data.c5.complete$data.X5,
data.c5.complete$data.X6),ncol=6)
xil.c6.complete=matrix(c(data.c6.complete$data.X1,data.c6.complete$data.X2,
data.c6.complete$data.X3,data.c6.complete$data.X4,data.c6.complete$data.X5,
data.c6.complete$data.X6),ncol=6)
x1.complete=sum((xil.c1.complete-xlc1bar.complete)^2)
x2.complete=sum((xil.c2.complete-xlc2bar.complete)^2)
x3.complete=sum((xil.c3.complete-xlc3bar.complete)^2)
x4.complete=sum((xil.c4.complete-xlc4bar.complete)^2)
x5.complete=sum((xil.c5.complete-xlc5bar.complete)^2)
x6.complete=sum((xil.c6.complete-xlc6bar.complete)^2)
SSW.complete.c6=sum(x1.complete,x2.complete,x3.complete,x4.complete,x5.c
omplete,x6.complete)
SSB.complete.c6=SST-SSW.complete.c6
Rsquare.complete.c6=SSB.complete.c6/SST
hasil.complete6=c(SST, SSW.complete.c6, SSB.complete.c6,
Rsquare.complete.c6)
#average
kelompok.average=hasil$average
data.average=data.frame(average, data$X1,data$X2, data$X3, data$X4,
data$X5, data$X6)
data.average.sort=data.average[order(data.average$average),]
a=length(data.average.sort$average[data.average.sort$average==1])
b=length(data.average.sort$average[data.average.sort$average==2])
c=length(data.average.sort$average[data.average.sort$average==3])
d=length(data.average.sort$average[data.average.sort$average==4])
e=length(data.average.sort$average[data.average.sort$average==5])
f=length(data.average.sort$average[data.average.sort$average==6])
data.c1.average=data.average.sort[1:a,]
data.c2.average=data.average.sort[(a+1):(a+b),]
data.c3.average=data.average.sort[(a+b+1):(a+b+c),]
data.c4.average=data.average.sort[(a+b+c+1):(a+b+c+d),]
data.c5.average=data.average.sort[(a+b+c+d+1):(a+b+c+d+e),]
data.c6.average=data.average.sort[(a+b+c+d+e+1):51,]

```

```

xlc1bar.average=matrix(c(rep(mean(data.c1.average$data.X1), a),
rep(mean(data.c1.average$data.X2),a),
rep(mean(data.c1.average$data.X3),a),rep(mean(data.c1.average$data.X4),a),rep
p(mean(data.c1.average$data.X5),a),rep(mean(data.c1.average$data.X6),a)),
ncol=6)
xlc2bar.average=matrix(c(rep(mean(data.c2.average$data.X1), b),
rep(mean(data.c2.average$data.X2),b),
rep(mean(data.c2.average$data.X3),b),rep(mean(data.c2.average$data.X4),b),re
p(mean(data.c2.average$data.X5),b),rep(mean(data.c2.average$data.X6),b)),
ncol=6)
xlc3bar.average=matrix(c(rep(mean(data.c3.average$data.X1), c),
rep(mean(data.c3.average$data.X2),c),
rep(mean(data.c3.average$data.X3),c),rep(mean(data.c3.average$data.X4),c),re
p(mean(data.c3.average$data.X5),c),rep(mean(data.c3.average$data.X6),c)),
ncol=6)
xlc4bar.average=matrix(c(rep(mean(data.c4.average$data.X1), d),
rep(mean(data.c4.average$data.X2),d),
rep(mean(data.c4.average$data.X3),d),rep(mean(data.c4.average$data.X4),d),re
p(mean(data.c4.average$data.X5),d),rep(mean(data.c4.average$data.X6),d)),
ncol=6)
xlc5bar.average=matrix(c(rep(mean(data.c5.average$data.X1), e),
rep(mean(data.c5.average$data.X2),e),
rep(mean(data.c5.average$data.X3),e),rep(mean(data.c5.average$data.X4),e),re
p(mean(data.c5.average$data.X5),e),rep(mean(data.c5.average$data.X6),e)),
ncol=6)
xlc6bar.average=matrix(c(rep(mean(data.c6.average$data.X1), f),
rep(mean(data.c6.average$data.X2),f),
rep(mean(data.c6.average$data.X3),f),rep(mean(data.c6.average$data.X4),f),rep
(mean(data.c6.average$data.X5),f),rep(mean(data.c6.average$data.X6),f)),
ncol=6)
xil.c1.average=matrix(c(data.c1.average$data.X1,data.c1.average$data.X2,data
.c1.average$data.X3,data.c1.average$data.X4,data.c1.average$data.X5,data.c1.a
verage$data.X6),ncol=6)
xil.c2.average=matrix(c(data.c2.average$data.X1,data.c2.average$data.X2,data
.c2.average$data.X3,data.c2.average$data.X4,data.c2.average$data.X5,data.c2.a
verage$data.X6),ncol=6)
xil.c3.average=matrix(c(data.c3.average$data.X1,data.c3.average$data.X2,data
.c3.average$data.X3,data.c3.average$data.X4,data.c3.average$data.X5,data.c3.a
verage$data.X6),ncol=6)

```

```

xil.c4.average=matrix(c(data.c4.average$data.X1,data.c4.average$data.X2,data.
c4.average$data.X3,data.c4.average$data.X4,data.c4.average$data.X5,data.c4.a
verage$data.X6),ncol=6)

xil.c5.average=matrix(c(data.c5.average$data.X1,data.c5.average$data.X2,data.
c5.average$data.X3,data.c5.average$data.X4,data.c5.average$data.X5,data.c5.a
verage$data.X6),ncol=6)

xil.c6.average=matrix(c(data.c6.average$data.X1,data.c6.average$data.X2,data.
c6.average$data.X3,data.c6.average$data.X4,data.c6.average$data.X5,data.c6.a
verage$data.X6),ncol=6)
x1.average=sum((xil.c1.average-xlc1bar.average)^2)
x2.average=sum((xil.c2.average-xlc2bar.average)^2)
x3.average=sum((xil.c3.average-xlc3bar.average)^2)
x4.average=sum((xil.c4.average-xlc4bar.average)^2)
x5.average=sum((xil.c5.average-xlc5bar.average)^2)
x6.average=sum((xil.c6.average-xlc6bar.average)^2)

SSW.average.c6=sum(x1.average,x2.average,x3.average,x4.average,x5.average,
x6.average)
SSB.average.c6=SST-SSW.average.c6
Rsquare.average.c6=SSB.average.c6/SST
hasil.average6=c(SST, SSW.average.c6, SSB.average.c6, Rsquare.average.c6)
#ringkasan
hasil6=rbind(hasil.single6,hasil.complete6,hasil.average6)
colnames(hasil6)=c("SST", "SSW", "SSB", "R-Square")
print ("Hasil Pengelompokkan")
print (hasil)
print ("Hasil R-Square")
print (hasil6)

```

Lampiran 8. *Syntax* Nilai Rasio pada Analisis Kelompok Data Numerik dengan Jumlah Kelompok 4

```

datanumerik=data.frame(data$X1, data$X2, data$X3, data$X4, data$X5,
data$X6)
d=dist(datanumerik,method = "euclidean")
fit.single=hclust(d, method="single")
fit.complete=hclust(d,method="complete")
fit.average=hclust(d, method="average")
single=cutree(fit.single, k=4)
complete=cutree(fit.complete, k=4)
average=cutree(fit.average, k=4)
hasilkelompoknumerik=data.frame (data$SMA, single, complete, average)
mean.X=matrix(ncol=1, nrow=51)
for (i in 1:51){
mean.X[i,]=mean(datanumerik2[i,])
}
kelompok.single=hasilkelompoknumerik$single
data.single=data.frame(single, mean.X)
data.single.sort=data.single[order(data.single$single),]
data.single.sort
a=length(data.single.sort$single[data.single.sort$single==1])
b=length(data.single.sort$single[data.single.sort$single==2])
c=length(data.single.sort$single[data.single.sort$single==3])
d=length(data.single.sort$single[data.single.sort$single==4])
data.c1.single=data.single.sort[1:a,]
data.c2.single=data.single.sort[(a+1):(a+b),]
data.c3.single=data.single.sort[(a+b+1):(a+b+c),]
data.c4.single=data.single.sort[(a+b+c+1):51,]
mean.c1.single=mean(data.c1.single$mean.X)
mean.c2.single=mean(data.c2.single$mean.X)
mean.c3.single=mean(data.c3.single$mean.X)
mean.c4.single=mean(data.c4.single$mean.X)
s1.single=sqrt((sum((data.c1.single$mean.X-mean.c1.single)^2))/(a))
s2.single=sqrt((sum((data.c2.single$mean.X-mean.c2.single)^2))/(b))
s3.single=sqrt((sum((data.c3.single$mean.X-mean.c3.single)^2))/(c))
s4.single=sqrt((sum((data.c4.single$mean.X-mean.c4.single)^2))/(d))
jumlah.s.single=sum(s1.single, s2.single, s3.single, s4.single)

```

```

Sw.single=jumlah.s.single/4
y1.single=(mean.c1.single-mean(mean.X))^2
y2.single=(mean.c2.single-mean(mean.X))^2
y3.single=(mean.c3.single-mean(mean.X))^2
y4.single=(mean.c4.single-mean(mean.X))^2
jumlah.single=sum(y1.single, y2.single, y3.single, y4.single)
sb.single=sqrt(jumlah.single/(4-1))
rasio.single=Sw.single/sb.single
hasil.single=c(Sw.single, sb.single,rasio.single)
kelompok.complete=hasilkelompoknumerik$complete
data.complete=data.frame(complete, mean.X)
data.complete.sort=data.complete[order(data.complete$complete),]
data.complete.sort
a=length(data.complete.sort$complete[data.complete.sort$complete==1])
b=length(data.complete.sort$complete[data.complete.sort$complete==2])
c=length(data.complete.sort$complete[data.complete.sort$complete==3])
d=length(data.complete.sort$complete[data.complete.sort$complete==4])
data.c1.complete=data.complete.sort[1:a,]
data.c2.complete=data.complete.sort[(a+1):(a+b),]
data.c3.complete=data.complete.sort[(a+b+1):(a+b+c),]
data.c4.complete=data.complete.sort[(a+b+c+1):51,]
mean.c1.complete=mean(data.c1.complete$mean.X)
mean.c2.complete=mean(data.c2.complete$mean.X)
mean.c3.complete=mean(data.c3.complete$mean.X)
mean.c4.complete=mean(data.c4.complete$mean.X)
s1.complete=sqrt((sum((data.c1.complete$mean.X-mean.c1.complete)^2))/(a))
s2.complete=sqrt((sum((data.c2.complete$mean.X-mean.c2.complete)^2))/(b))
s3.complete=sqrt((sum((data.c3.complete$mean.X-mean.c3.complete)^2))/(c))
s4.complete=sqrt((sum((data.c4.complete$mean.X-mean.c4.complete)^2))/(d))
jumlah.s.complete=sum(s1.complete, s2.complete, s3.complete, s4.complete)
Sw.complete=jumlah.s.complete/4
y1.complete=(mean.c1.complete-mean(mean.X))^2
y2.complete=(mean.c2.complete-mean(mean.X))^2
y3.complete=(mean.c3.complete-mean(mean.X))^2
y4.complete=(mean.c4.complete-mean(mean.X))^2
jumlah.complete=sum(y1.complete, y2.complete, y3.complete, y4.complete)
sb.complete=sqrt(jumlah.complete/(4-1))
rasio.complete=Sw.complete/sb.complete

```

```

hasil.complete=c(Sw.complete, sb.complete, rasio.complete)
kelompok.average=hasilkelompoknumerik$average
data.average=data.frame(average, mean.X)
data.average.sort=data.average[order(data.average$average),]
a=length(data.average.sort$average[data.average.sort$average==1])
b=length(data.average.sort$average[data.average.sort$average==2])
c=length(data.average.sort$average[data.average.sort$average==3])
d=length(data.average.sort$average[data.average.sort$average==4])
data.c1.average=data.average.sort[1:a,]
data.c2.average=data.average.sort[(a+1):(a+b),]
data.c3.average=data.average.sort[(a+b+1):(a+b+c),]
data.c4.average=data.average.sort[(a+b+c+1):51,]
mean.c1.average=mean(data.c1.average$mean.X)
mean.c2.average=mean(data.c2.average$mean.X)
mean.c3.average=mean(data.c3.average$mean.X)
mean.c4.average=mean(data.c4.average$mean.X)
s1.average=sqrt((sum((data.c1.average$mean.X-mean.c1.average)^2))/(a))
s2.average=sqrt((sum((data.c2.average$mean.X-mean.c2.average)^2))/(b))
s3.average=sqrt((sum((data.c3.average$mean.X-mean.c3.average)^2))/(c))
s4.average=sqrt((sum((data.c4.average$mean.X-mean.c4.average)^2))/(d))
jumlah.s.average=sum(s1.average, s2.average, s3.average, s4.average)
Sw.average=jumlah.s.average/4
y1.average=(mean.c1.average-mean(mean.X))^2
y2.average=(mean.c2.average-mean(mean.X))^2
y3.average=(mean.c3.average-mean(mean.X))^2
y4.average=(mean.c4.average-mean(mean.X))^2
jumlah.average=sum(y1.average, y2.average, y3.average, y4.average)
sb.average=sqrt(jumlah.average/(4-1))
rasio.average=Sw.average/sb.average
hasil.average=c(Sw.average, sb.average, rasio.average)
hasil=rbind(hasil.single, hasil.complete, hasil.average)
print ("Nilai Rasio antara Sw dan Sb")
print ( "Metode      Sw      Sb      Rasio")
print (hasil)

```

Lampiran 9. *Syntax* Analisis Pengelompokkan Data Kategorik dengan K-modes

```
inputkat  
inputkat[,2:4]  
cl=kmodes(inputkat[,2:4], 2, iter.max = 10, weighted = FALSE)  
plot(jitter(as.matrix(inputkat[,2:4])), col = cl$cluster)  
points(cl$modes, col = 1:5, pch = 8)
```


Lampiran 10. *Syntax* Analisis Pengelompokkan Data Campuran dengan Ensembl SWFM

```

datanumerik = data.frame(data$X1, data$X2, data$X3,
data$X4, data$X5, data$X6)
datakategorik = data.frame(data$x7, data$x8, data$x9)
# Pengelompokan Numerik
d = dist(datanumerik, method = "euclidean")
fit.sin = hclust(d, method = "single")
k = 4
single = cutree(fit.sin, k=4)
# Pengelompokan Kategorik
datakategorik1 = c(data$x7, data$x8, data$x9)
datakategorik2 = matrix (datakategorik1,51,3)
n = 2
cl = kmodes(inputkat[,2:4], 2, iter.max = 10, weighted = FALSE)
names(cl)
cl$cluster
as.matrix(cl$cluster)
t(as.matrix(cl$cluster))
u=t(as.matrix(cl$cluster))
# 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(single[1],u[1])
data2 = c(single[2],u[2])
data3 = c(single[3],u[3])
data4 = c(single[4],u[4])
data5 = c(single[5],u[5])
data6 = c(single[6],u[6])
data7 = c(single[7],u[7])

```

```
data8 = c(single[8],u[8])
data9 = c(single[9],u[9])
data10 = c(single[10],u[10])
data11 = c(single[11],u[11])
data12 = c(single[12],u[12])
data13 = c(single[13],u[13])
data14 = c(single[14],u[14])
data15 = c(single[15],u[15])
data16 = c(single[16],u[16])
data17 = c(single[17],u[17])
data18 = c(single[18],u[18])
data19 = c(single[19],u[19])
data20 = c(single[20],u[20])
data21 = c(single[21],u[21])
data22 = c(single[22],u[22])
data23 = c(single[23],u[23])
data24 = c(single[24],u[24])
data25 = c(single[25],u[25])
data26 = c(single[26],u[26])
data27 = c(single[27],u[27])
data28 = c(single[28],u[28])
data29 = c(single[29],u[29])
data30 = c(single[30],u[30])
data31 = c(single[31],u[31])
data32 = c(single[32],u[32])
data33 = c(single[33],u[33])
data34 = c(single[34],u[34])
data35 = c(single[35],u[35])
data36 = c(single[36],u[36])
data37 = c(single[37],u[37])
data38 = c(single[38],u[38])
data39 = c(single[39],u[39])
data40 = c(single[40],u[40])
data41 = c(single[41],u[41])
data42 = c(single[42],u[42])
data43 = c(single[43],u[43])
data44 = c(single[44],u[44])
data45 = c(single[45],u[45])
```

```

data46 = c(single[46],u[46])
data47 = c(single[47],u[47])
data48 = c(single[48],u[48])
data49 = c(single[49],u[49])
data50 = c(single[50],u[50])
data51 = c(single[51],u[51])
data=rbind(data1,data2,data3,data4,data5,data6,data7,data8,data9,data10,data11
,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,data39,data40,data41,data42,data43,data44
,data45,data46,data47,data48,data49,data50,data51)
sim = matrix (1, nrow = 51, ncol = 51)
rownames(sim) = c(1:51)
colnames(sim) = c(1:51)
for (i in 1:(nrow(sim)-1)){
  for (j in (i+1):nrow(sim)){
    x = data[i,]
    y = data[j,]
    sim [i,j] = sij(x,y)
    sim [j,i] = sim[i,j]
  }
}
sim
wi = 0.5
F = wi*(1-sim)
jarak = as.dist(F)
fit = hclust(jarak,method = "single")
hasil.k2 = cutree(fit, k=2)
hasil.k3 = cutree(fit, k=3)
hasil.k4 = cutree(fit, k=4)
hasil.k5 = cutree(fit, k=5)
hasil.k6 = cutree(fit, k=6)
hasil.kelompok =
data.frame(hasil.k2,hasil.k3,hasil.k4,hasil.k5,hasil.k6)
SWFM = data.frame(hasil.kelompok)
write.csv(SWFM,file = "Data Hasil Ensemble SWFM.csv")
SWFM = read.csv("Data Hasil Ensemble SWFM.csv", header=TRUE,
sep=",")

```

```

X1=outputkel$X1
X2=outputkel$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))
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/(51-2))
SW.3 = sqrt(SSW.3/(51-3))
SW.4 = sqrt(SSW.4/(51-4))
SW.5 = sqrt(SSW.5/(51-5))
SW.6 = sqrt(SSW.6/(51-6))

```

```
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 Ensembl SWFM",
xlab = "Jumlah Kluster", ylab = "Nilai Ratio")
lines(u,z,col = "Green")
print("Hasil Pengelompokan SWFM")
print(hasil.kelompok)
print("-----")
print("Nilai Ratio untuk Setiap Nilai K")
print(Ratio.SWFM)
```

Lampiran 11. Output untuk Analisis Kelompok pada Pada Data Numerik dengan Jumlah Kelompok 2

```
> print (hasil)
single complete average
1  1  1  1
2  1  1  1
3  1  1  1
4  1  1  1
5  1  1  1
6  1  1  1
7  1  1  1
8  1  1  1
9  1  1  1
10 1  1  1
11 1  1  1
12 1  1  1
13 1  1  1
14 1  1  1
15 1  1  1
16 1  1  1
17 1  1  1
18 1  1  1
19 1  1  1
20 1  1  1
21 1  1  1
22 1  1  1
23 1  1  1
24 1  1  1
25 1  1  1
26 1  1  1
27 1  1  1
28 1  1  1
29 2  2  2
30 1  1  1
31 1  1  1
32 1  1  1
```

```
33 1 1 1
34 1 1 1
35 1 1 1
36 1 1 1
37 1 1 1
38 1 1 1
39 1 1 1
40 1 1 1
41 1 1 1
42 1 1 1
43 1 1 1
44 1 1 1
45 1 1 1
46 1 1 1
47 1 1 1
48 1 1 1
49 1 1 1
50 1 1 1
51 1 1 1
> print ("Hasil R-Square")
[1] "Hasil R-Square"
> print (hasil2)
      SST      SSW      SSB R-Square
hasil.single2 397258167467 75979362595 321278804872 0.8087406
hasil.complete2 397258167467 75979362595 321278804872 0.8087406
hasil.average2 397258167467 75979362595 321278804872 0.8087406
```

Lampiran 12. Output untuk Analisis Kelompok pada Pada Data Numerik dengan Jumlah Kelompok 3

```
> print (hasil)
single complete average
1 1 1 1
2 1 1 1
3 1 1 1
4 1 1 1
5 1 1 1
6 1 1 1
7 1 1 1
8 1 1 1
9 1 1 1
10 1 1 1
11 1 1 1
12 1 1 1
13 1 1 1
14 1 1 1
15 1 1 1
16 1 1 1
17 1 1 1
18 1 1 1
19 1 1 1
20 1 1 1
21 1 1 1
22 2 2 2
23 1 1 1
24 1 1 1
25 1 1 1
26 1 1 1
27 1 1 1
28 1 1 1
29 3 3 3
30 1 1 1
31 1 1 1
32 1 1 1
```



```
33 1 1 1
34 1 1 1
35 1 1 1
36 1 1 1
37 1 1 1
38 1 1 1
39 1 1 1
40 1 1 1
41 2 2 2
42 1 1 1
43 1 1 1
44 1 1 1
45 1 1 1
46 1 1 1
47 1 1 1
48 1 1 1
49 1 1 1
50 1 1 1
51 1 1 1
> print ("Hasil R-Square")
[1] "Hasil R-Square"
> print (hasil3)
      SST      SSW      SSB R-Square
hasil.single3 397258167467 31181521532 366076645935 0.9215082
hasil.complete3 397258167467 31181521532 366076645935 0.9215082
hasil.average3 397258167467 31181521532 366076645935 0.9215082
```

Lampiran 13. Output untuk Analisis Kelompok pada Pada Data Numerik dengan Jumlah Kelompok 4

```
> print (hasil)
single complete average
1 1 1 1
2 1 2 1
3 1 2 1
4 1 2 1
5 1 1 1
6 1 2 1
7 1 1 1
8 1 2 1
9 1 2 1
10 1 2 1
11 1 2 1
12 1 2 1
13 1 2 1
14 1 2 1
15 1 1 1
16 1 1 1
17 1 2 1
18 1 1 1
19 1 2 1
20 1 2 1
21 1 1 2
22 2 3 3
23 1 2 1
24 1 2 1
25 1 1 1
26 1 2 1
27 1 2 1
28 1 1 1
29 3 4 4
30 1 1 2
31 1 2 1
32 1 2 1
```

```
33  1  2  1
34  1  1  1
35  1  2  1
36  1  1  1
37  1  1  2
38  1  1  2
39  1  2  1
40  1  1  2
41  4  3  3
42  1  2  1
43  1  2  1
44  1  1  1
45  1  2  1
46  1  1  2
47  1  2  1
48  1  2  1
49  1  2  1
50  1  2  1
51  1  2  1
> print ("Hasil R-Square")
[1] "Hasil R-Square"
> print (hasil4)
      SST      SSW      SSB R-Square
hasil.single4 397258167467 29873425141 367384742327 0.9248010
hasil.complete4 397258167467 10039781501 387218385966 0.9747273
hasil.average4 397258167467 12512430977 384745736490 0.9685030
```

Lampiran 14. Output untuk Analisis Kelompok pada Pada Data Numerik dengan Jumlah Kelompok 5

```
> print (hasil)
  single complete average
1     1     1     1
2     1     2     1
3     1     2     1
4     1     2     1
5     1     1     1
6     1     2     1
7     1     1     1
8     1     2     1
9     1     2     1
10    1     2     1
11    1     2     1
12    1     2     1
13    1     2     1
14    1     2     1
15    1     1     1
16    1     1     1
17    1     2     1
18    1     1     1
19    1     2     1
20    1     2     1
21    2     3     2
22    3     4     3
23    1     2     1
24    1     2     1
25    1     1     1
26    1     2     1
27    1     2     1
28    1     1     1
29    4     5     4
30    2     3     2
31    1     2     1
32    1     2     1
```

```
33  1  2  1
34  1  1  1
35  1  2  1
36  1  1  1
37  2  3  2
38  2  3  2
39  1  2  1
40  2  3  2
41  5  4  5
42  1  2  1
43  1  2  1
44  1  1  1
45  1  2  1
46  2  3  2
47  1  2  1
48  1  2  1
49  1  2  1
50  1  2  1
51  1  2  1
> print ("Hasil R-Square")
[1] "Hasil R-Square"
> print (hasil5)
      SST      SSW      SSB R-Square
hasil.single5 397258167467 11204334586 386053832881 0.9717958
hasil.complete5 397258167467 4627354867 392630812601 0.9883518
hasil.average5 397258167467 11204334586 386053832881 0.9717958
```

Lampiran 15. Output untuk Analisis Kelompok pada Pada Data Numerik dengan Jumlah Kelompok 6

```
> print (hasil)
  single complete average
1     1     1     1
2     1     2     2
3     1     2     2
4     1     2     2
5     1     1     1
6     1     2     2
7     2     1     1
8     1     2     2
9     1     2     2
10    1     2     1
11    1     2     2
12    1     2     2
13    1     2     2
14    1     2     2
15    1     1     1
16    1     1     1
17    1     2     2
18    1     1     1
19    1     2     2
20    1     2     1
21    3     3     3
22    4     4     4
23    1     2     2
24    1     2     2
25    1     1     1
26    1     2     2
27    1     2     2
28    1     1     1
29    5     5     5
30    3     3     3
31    1     2     2
32    1     2     2
```

```
33  1  2  2
34  1  1  1
35  1  2  2
36  1  1  1
37  3  3  3
38  3  3  3
39  1  2  2
40  3  3  3
41  6  6  6
42  1  2  2
43  1  2  2
44  1  1  1
45  1  2  2
46  3  3  3
47  1  2  2
48  1  2  2
49  1  2  2
50  1  2  2
51  1  2  2
> print ("Hasil R-Square")
[1] "Hasil R-Square"
> print (hasil6)
      SST      SSW      SSB R-Square
hasil.single6 397258167467 9321955754 387936211714 0.9765343
hasil.complete6 397258167467 3319258475 393938908992 0.9916446
hasil.average6 397258167467 3054794831 394203372636 0.9923103
```

Lampiran 16. *Output* untuk Analisis Kelompok pada Pada Data Kategorik dengan Jumlah Kelompok 2

K-modes clustering with 2 clusters of sizes 46, 5

Cluster modes:

Z1 Z2 Z3

1 0 1 0

2 1 1 1

Clustering vector:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
28

1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1

29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51

1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1

Within cluster simple-matching distance by cluster:

[1] 20 3

Available components:

[1] "cluster" "size" "modes" "withindiff" "iterations" "weighted"

Lampiran 17. Hasil Pengelompokan Terbaik Dari Data Numerik dan Kategorik.

X1	X2	X1	X2
1	1	1	1
1	1	1	1
1	1	1	1
1	2	1	2
1	1	1	1
1	1	1	1
1	1	1	1
1	2	4	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	2	1	1
1	1	1	1
1	1	1	1
1	2	1	1
2	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
3	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

Lampiran 18. Output untuk Analisis Kelompok pada Pada Data Campuran dengan Ensembel SWFM

```
> print("Hasil Pengelompokan SWFM")
[1] "Hasil Pengelompokan SWFM"
> print(hasil.kelompok)
  hasil.k2 hasil.k3 hasil.k4 hasil.k5 hasil.k6
1      1      1      1      1      1
2      1      1      1      1      1
3      1      1      1      1      1
4      1      1      1      1      1
5      1      1      1      1      1
6      1      1      1      1      1
7      1      1      2      2      2
8      1      1      1      1      1
9      1      1      1      1      1
10     1      1      1      1      1
11     1      1      1      1      1
12     1      1      1      1      1
13     1      1      1      1      1
14     1      1      1      1      1
15     1      1      1      1      1
16     1      1      1      1      1
17     1      1      1      1      1
18     1      1      1      1      1
19     1      1      1      1      1
20     1      1      1      1      1
21     1      1      1      1      1
22     1      1      2      3      3
23     1      1      1      1      1
24     1      1      1      1      1
25     1      1      1      1      1
26     1      1      1      1      1
27     1      1      1      1      1
28     1      1      1      1      1
29     1      2      3      4      4
```

```

30  1  1  1  1  1
31  1  1  1  1  1
32  1  1  1  1  1
33  1  1  1  1  1
34  1  1  1  1  1
35  1  1  1  1  1
36  1  1  1  1  1
37  1  1  1  1  1
38  1  1  1  1  1
39  1  1  1  1  1
40  1  1  1  1  1
41  2  3  4  5  5
42  1  1  1  1  1
43  1  1  1  1  1
44  1  1  1  1  1
45  1  1  1  1  1
46  1  1  1  1  1
47  1  1  1  1  1
48  1  1  1  1  1
49  1  1  1  1  1
50  1  1  1  1  1
51  1  1  1  1  6
> print("-----")
[1] "-----"
> print("Nilai Ratio untuk Setiap Nilai K")
[1] "Nilai Ratio untuk Setiap Nilai K"
> print(Ratio.SWFM)
      [,1]
Ratio.2 0.10769418
Ratio.3 0.06339664
Ratio.4 0.06909463
Ratio.5 0.06713177
Ratio.6 0.33575184

```

BIODATA PENULIS



Aryo Joko Prakoso atau biasa dipanggil Joko merupakan anak dari pasangan Bapak Ariyadi dan Ibu Wiji Asri. Penulis lahir di Trenggalek 28 April 1995. Penulis memulai jenjang pendidikan di SDN Jagalan III Kediri (2001-2007), kemudian melanjutkan pendidikan di SMP Al-Huda Kediri (2007-2010). Penulis melanjutkan pendidikannya di SMAN 2 Kediri (2010-2013). Pada tahun 2013, penulis melanjutkan pendidikan di jenjang perguruan tinggi di jurusan Statistika ITS melalui jalur undangan SNMPTN dan memperoleh NRP 1313100059. Selama kuliah penulis aktif di organisasi kemahasiswaan ITS tingkat jurusan dan UKM yakni HIMASTA-ITS dan UKM SepakBola ITS sebagai staff Departemen Kewirausahaan pada periode 2014/2015 dan staff PSDU di UKM pada periode yang sama 2014/2015. Selain itu penulis juga turut berpartisipasi dalam kepanitiaan seperti Pekan Raya Statistika (PRS) 2015 dan lainnya.

Penulis menerima segala kritikan, masukan, dan saran yang bersifat membangun demi meningkatkan manfaat Tugas Akhir ini. Segala kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini dapat menghubungi kontak berikut.

Email : prakoso05joko@gmail.com

Line : [aryo.joko](https://line.me/tv/aryo.joko)

Steam : steamcommunity.com/profiles/76561198300695908/

No HP : 085749698218

(halaman ini sengaja dikosongkan)