

**PERBANDINGAN ALGORITMA *K-MEANS* DAN ALGORITMA
K-MEDOIDS DALAM PENGELOMPOKAN KOMODITAS TANAMAN
BIOFARMAKA DI PROVINSI JAWA TENGAH TAHUN 2018**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Jurusan
Statistika



Disusun Oleh:

Arfa Nawal Farachi

16 611 024

**JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2020**

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : Perbandingan Algoritma *K-Means* Dan Algoritma *K-Medoids* Dalam Pengelompokan Komoditas Tanaman Biofarmaka di Provinsi Jawa Tengah Tahun 2018

Nama Mahasiswa : Arfa Nawal Farachi

Nomor Mahasiswa : 16611024

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 31 Agustus 2020

Pembimbing



(Prof. Akhmad Fauzy, S.Si., M.Si., Ph.D.)

HALAMAN PENGESAHAN

TUGAS AKHIR

**Perbandingan Algoritma *K-Means* Dan Algoritma *K-Medoids* Dalam
Pengelompokan Komoditas Tanaman Biofarmaka di Provinsi Jawa Tengah
Tahun 2018**

Nama Mahasiswa : Arfa Nawal Farachi

Nomor Mahasiswa : 16611024

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 4 SEPTEMBER 2020**

Nama Penguji

Tanda tangan

1. Prof. Akhmad Fauzy, S.Si., M.Si.,
Ph.D.
2. Achmad Fauzan, S.Pd., M.Si.
3. Muhammad Hasan Sidiq
Kurniawan, S.Si., M.Sc.



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)

KATA PENGANTAR



Assalamu'alaikum Warahmatullaahi Wabarakaatuh

Alhamdulillah *rabbi'l'aalamiin*, Puji Syukur kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah, dan inayah-Nya, sehingga penulis dapat menyusun dan menyelesaikan tugas akhir yang berjudul **“Perbandingan Algoritma *K-means* dan Algoritma *K-Medoids* dalam Pengelompokan Komoditas Tanaman Biofarmaka di Jawa Tengah Tahun 2018”** dengan lancar dan sebaik-baiknya sebagai salah satu persyaratan yang harus dipenuhi dalam menyelesaikan jenjang strata satu di Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Shalawat serta salam semoga selalu tercurah kepada junjungan Nabi Agung Muhammad SAW semoga mendapatkan syafaatnya diyaumul qiyamah nanti.

Penyelesaian tugas akhir ini tidak terlepas dari dukungan, bantuan, arahan, dan bimbingan dari berbagai pihak. Untuk itu pada kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua Orang Tua beserta keluarga besar penulis yang telah banyak berkorban dan berjuang mulai dari tenaga, materi, pikiran untuk dapat menyekolahkan saya diperguruan tinggi, dan selalu memotivasi, menghibur, dan mendoakan yang terbaik sehingga saya selalu mendapatkan hidayah dari Allah SWT selama masa perkuliahan hingga saat ini.
2. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
3. Bapak Dr. Edy Widodo, S.Si., M.Si. Kepala Program Studi Statistika Universitas Islam Indonesia.
4. Bapak Prof. Akhmad Fauzy, S.Si., M.Si., Ph.D. selaku dosen Pembimbing Tugas Akhir yang sudah membimbing penulis.

5. Seluruh staff, pengajar Program Studi Statistika Universitas Islam Indonesia yang telah memberikan bekal ilmu dan bantuan kepada penulis.
6. Bapak Ibu Dosen Program Studi Statistika, Universitas Islam Indonesia yang telah mengajari penulis dikelas dari semester awal hingga akhir, sehingga saya dapat memperoleh banyak pengalaman dan ilmu terbaik yang dapat saya tuangkan dalam penyusunan tugas akhir ini.
7. Keluarga besar IKS (Ikatan Keluarga Statistika), sebagai Organisasi yang membawahi mahasiswa statistika FMIPA UII
8. Teman seperjuangan bimbingan TA pak Prof. Fauzy, yang telah bekerjasama dengan baik, menemani penyusunan tugas akhir dan bertukar pikiran dalam penyempurnaan penyusunan skripsi hingga terlaksananya pendadaran.
9. Seluruh teman-teman jurusan statistika angkatan 2016 yang telah banyak memberi informasi terkait akademik dan banyak bertukar ilmu selama 4 tahun berjuang bersama untuk menyanggah gelar sarjana.
10. Serta seluruh pihak-pihak lainnya yang tidak dapat disebutkan satu per satu, yang telah banyak membantu terselesaikannya tugas akhir ini dengan perjuangan.

Demikianlah yang dapat disampaikan, semoga Allah SWT senantiasa melimpahkan rahmat dan ridho-Nya kepada semua pihak yang telah membantu penulis. Semoga Tugas Akhir ini dapat memberikan manfaat bagi penulis khususnya dan umumnya bagi semua pihak yang membutuhkan. Akhir kata, semoga Allah SWT senantiasa melimpahkan rahmat serta hidayah-Nya kepada kita semua, Amin amin ya robbal 'alamiin.

Wassalamu'alaikum Warahmatullaahi Wabarakaatuh

Yogyakarta, 31 Agustus 2020

Arfa Nawal Farachi

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN PERSETUJUAN PEMBIMBING.....	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
PERNYATAAN	xiii
INTISARI.....	xiv
ABSTRACT	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	6
1.6. Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
2.1. Kajian Pustaka.....	8
BAB III LANDASAN TEORI	12
3.1. Pertanian.....	12
3.2. Cangkupan Pertanian	13
3.3. Subsektor Holtikultural.....	14
3.4. Statistika Deskriptif	15
3.5. Asumsi Analisis Kelompok.....	15
3.5.1 Sampel Representatif	15
3.5.2 Tidak Ada Multikolinieritas	15

3.5.3	Analisis Komponen Utama	17
3.6.	<i>Clustering</i>	20
3.6.1	<i>Hierarchical Methods</i> (Metode Hierarki)	22
3.6.2	Pengklasteran Sekatan (<i>Partitioning</i>) atau <i>Non</i> Hierarki.....	23
3.7.	Penentuan Jumlah Cluster menggunakan metode <i>Elbow</i>	34
3.8.	Ukuran Jarak dalam Pengklasteran.....	36
3.8.1	Jarak Euclidean.....	36
3.8.2	Jarak Kuadrat Euclidean	36
3.8.3	Jarak <i>Mahalanobis</i>	36
3.8.4	Jarak Manhattan atau City-Block	37
3.9.	Pemilihan Metode Terbaik dengan Simpangan Baku.....	37
BAB IV METODOLOGI PENELITIAN		39
4.1.	Populasi dan Sampel Penelitian.....	39
4.2.	Jenis dan Sumber Data Penelitian	39
4.3.	Variabel Penelitian dan Definisi Operasional Variabel	39
4.4.	Metode Pengumpulan Data	42
4.5.	Metode Analisis Data.....	42
4.6.	Tahapan Penelitian.....	42
BAB V HASIL DAN PEMBAHASAN.....		44
5.1.	Analisis Deskriptif	44
5.2.	Asumsi Cluster	49
5.2.1	Sampel Representatif	49
5.2.2	Uji Multikolinieritas	49
5.3.	Analisis Komponen Utama	50
5.4.	Penerapan Cluster	52
5.4.1.	Analisis <i>Cluster K-means</i>	52
5.4.2.	Rasio Simpangan Baku dalam kelompok (S_w) dan Simpangan Baku antar kelompok (S_b) dalam <i>K-means</i>	55
5.4.3.	Analisis <i>Cluster K-medoids</i>	55

5.4.4. Rasio Simpangan Baku dalam kelompok (S_w) dan Simpangan Baku antar kelompok (S_b) dalam <i>K-medoids</i>	58
5.5. Perbandingan Metode Terbaik	58
BAB VI KESIMPULAN DAN SARAN	61
6.1. Kesimpulan.....	61
6.2. Saran	62
DAFTAR PUSTAKA	63
LAMPIRAN	68



DAFTAR TABEL

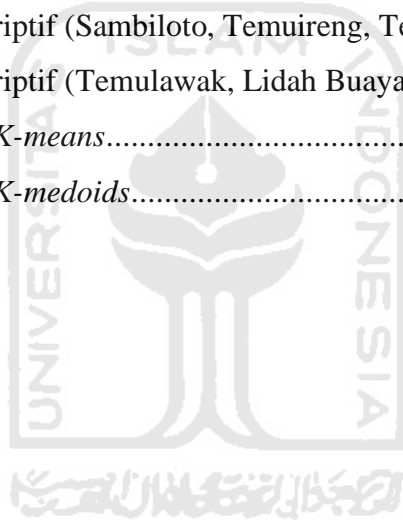
Tabel 2.1 Penelitian Sebelumnya	8
Tabel 3.1 Kondisi awal data	25
Tabel 3.2 <i>Centroid</i> awal	25
Tabel 3.3 Hasil perhitungan jarak ke setiap <i>centroid</i>	26
Tabel 3.4 Anggota klaster.....	27
Tabel 3.5 Anggota klaster 2.....	27
Tabel 3.6 Anggota klaster 3.....	27
Tabel 3.7 <i>Centroid</i> baru hasil iterasi 1	28
Tabel 3.8 Hasil perhitungan jarak ke setiap <i>centroid</i> baru hasil iterasi 1	28
Tabel 3.9 Kondisi awal data	30
Tabel 3.10 <i>Medoids</i>	30
Tabel 3.11 Hasil perhitungan jarak ke setiap <i>medoids</i>	31
Tabel 3.12 <i>Non Medoids</i>	31
Tabel 3.13 Hasil perhitungan jarak ke setiap <i>non medoids</i>	32
Tabel 3.14 <i>Non Medoids</i> baru.....	33
Tabel 3.15 Hasil perhitungan jarak ke setiap <i>non medoids</i>	33
Tabel 3.16 Hasil pengklasteran data simulasi dengan <i>K-medoids Clustering</i>	34
Tabel 4.1 Definisi Operasional Variabel.....	39
Tabel 5.1 Hasil Perhitungan Korelasi	49
Tabel 5.2 Nilai <i>Eigenvalues</i>	51
Tabel 5.3 Anggota <i>Cluster K-means</i>	53
Tabel 5.4 Nilai Rata-rata variabel <i>cluster K-means</i>	53
Tabel 5.5 Rasio S_w dan S_b <i>K-means</i>	55
Tabel 5.6 Anggota <i>Cluster K-medoids</i>	56

Tabel 5.7 Nilai Rata-rata variabel <i>cluster K-medoids</i>	57
Tabel 5.8 Rasio S_w dan S_b <i>K-medoids</i>	58
Tabel 5.9 Perbandingan <i>cluster K-means</i> dan <i>K-medoids</i>	59
Tabel 5.10 Perbandingan Nilai Rasio Simpangan Baku	60



DAFTAR GAMBAR

Gambar 3.1 Pengelompokan ideal	21
Gambar 3.2 Grafik metode <i>Elbow</i>	35
Gambar 4.1 Flowchart Penelitian	43
Gambar 5.1 Analisis Deskriptif (Jahe, Dringo, Kapulaga).....	44
Gambar 5.2 Analisis Deskriptif (Kejibeling, Kencur, Kunyit).....	45
Gambar 5.3 Analisis Deskriptif (Laos, Lempuyang, Mengkudu).....	46
Gambar 5.4 Analisis Deskriptif (Sambiloto, Temuireng, Temukunci).....	47
Gambar 5.5 Analisis Deskriptif (Temulawak, Lidah Buaya, Mahkota Dewa).....	48
Gambar 5.6 Hasil <i>Cluster K-means</i>	52
Gambar 5.7 Hasil <i>Cluster K-medoids</i>	56



DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian	68
Lampiran 2. <i>Syntax PCA</i>	71
Lampiran 3. <i>Syntax Cluster Kmeans</i>	71
Lampiran 4. <i>Syntax Cluster K-Medoids</i>	71
Lampiran 5. Korelasi.....	71



PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tugas akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk tugas akhir. Tugas akhir ini diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 31 Agustus 2020



Arfa Nawal Farachi

**PERBANDINGAN ALGORITMA K-MEANS DAN ALGORITMA
K-MEDOIDS DALAM PENGELOMPOKAN KOMODITAS TANAMAN
BIOFARMAKA DI PROVINSI JAWA TENGAH TAHUN 2018**

Arfa Nawal Farachi

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

INTISARI

Tanaman biofarmaka (tanaman obat) adalah tanaman yang bermanfaat untuk obat-obatan, kosmetik dan kesehatan yang dikonsumsi atau digunakan dari bagian-bagian tanaman seperti daun, batang, buah, umbi (rimpang) ataupun akar. Salah satu upaya yang dapat meningkatkan dan memelihara produktivitas di sektor tanaman biofarmaka yaitu dengan mengelompokkan ke dalam kelompok-kelompok berdasarkan kesamaan karakteristik yang dimiliki dari setiap wilayah yang ada di Provinsi Jawa Tengah. Salah satu cara dalam melakukan pembagian di setiap kelompok wilayah yaitu menggunakan analisis cluster. Analisis cluster adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan. Tujuan dari penelitian ini adalah untuk mengetahui hasil cluster yang terbentuk dan perbandingan hasil kinerja dengan menggunakan Algoritma *K-Means* dan Algoritma *K-Medoids* untuk pengelompokan kabupaten/kota berdasarkan variabel tanaman obat di Provinsi Jawa Tengah. Hasil cluster yang terbentuk dengan metode *Cluster K-Means* dan metode *Cluster K-Medoids* adalah sama. Diperoleh hasil clusternya yaitu *cluster* 1 ada 2 kabupaten, *cluster* 2 ada 32 kabupaten/kota dan *cluster* 3 ada 1 kabupaten. Dari hasil pengelompokan kedua metode tersebut merupakan metode yang terbaik karena nilai rasio simpangan baku hasilnya sama yaitu 0,333333.

Kata kunci: *Tanaman Biofarmaka, Analisis Cluster, Cluster K-Means, Cluster K-Medoids*

**COMPARISON OF K-MEANS ALGORITHM AND K-MEDOIDS ALGORITHM
IN THE GROUPING OF BIOFARMAKA PLANT COMMODITIES IN
CENTRAL JAVA PROVINCE YEAR 2018**

Arfa Nawal Farachi

Department of Statistics, Faculty of Mathematics and Natural Science

Islamic University of Indonesia

ABSTRACT

Biopharmaca (medicinal plants) are plants that are useful for medicine, cosmetics and health which are consumed or used from plant parts such as leaves, stems, fruit, tubers (rhizomes) or roots. One of the efforts that can increase and maintain productivity in the biopharmaceutical sector is by grouping them into groups based on the similarities in characteristics of each region in Central Java Province. One way to divide each regional group is to use cluster analysis. Cluster analysis is a method used to divide data sets into groups based on predetermined similarities. The purpose of this study was to determine the results of clusters formed and the comparison of performance results using the K-Means Algorithm and the K-Medoids Algorithm for district / city grouping based on medicinal plant variables in Central Java Province. The results of the clusters formed by the Cluster K-Means method and the K-Medoids Cluster method are the same. The results obtained from the clusters are cluster 1 there are 2 districts, cluster 2 has 32 districts / cities and cluster 3 has 1 district. From the results of grouping the two methods is the best method because the value of the standard deviation ratio is the same, namely 0.3333333.

Keywords: *Biopharmaca Plants, Cluster Analysis, K-Means Cluster, K-Medoids Cluster*

BAB I

PENDAHULUAN

1.1. Latar Belakang

Tanaman obat sangat bermanfaat dalam memenuhi kebutuhan hidup manusia. Dalam dunia farmasi, tanaman obat merupakan sumber bahan baku obat tradisional maupun modern. Sekarang ini ada kecenderungan masyarakat untuk mengkonsumsi obat tradisional, karena adanya perubahan gaya hidup *back to nature* dan mahalanya obat-obatan modern yang membuat permintaan tanaman obat semakin tinggi, tidak hanya di Indonesia tetapi juga dunia. Tanaman obat mungkin tidak sepopuler jenis tanaman lain, khususnya tanaman penghasil bahan makanan seperti buah-buahan, umbi-umbian dan sebagainya. Namun bagi sebagian orang pencinta alam, tanaman obat merupakan tanaman yang sangat populer, apalagi dengan perubahan pola hidup yang saat ini sudah mengglobal yang dikenal dengan istilah *back to nature*. *Back to nature* bukan hanya menjangkit pada pola konsumsi masyarakat, namun sudah merambah juga ke sektor-sektor lain termasuk pengobatan. Secara global juga sudah terjadi perubahan pola pengobatan masyarakat ke obat-obat tradisional yang terbuat dari bahan alami (Badan Pengkajian dan Pengembangan Perdagangan, 2017).

Tanaman obat sendiri memiliki ribuan jenis spesies. Dari total sekitar 40.000 jenis tumbuh-tumbuhan obat yang telah dikenal di dunia, 30.000-nya disinyalir berada di Indonesia. Jumlah tersebut mewakili 90% dari tanaman obat yang terdapat di wilayah Asia. Dari jumlah tersebut, 25% diantaranya atau sekitar 7.500 jenis sudah diketahui memiliki khasiat herbal atau tanaman obat. Namun hanya 1.200 jenis tanaman yang sudah dimanfaatkan untuk bahan baku obat-obatan herbal atau jamu (PT. Sido Muncul, 2015). Tidak mengherankan jika kemudian Indonesia dikenal dengan julukan *live laboratory*. Di Indonesia sendiri, meskipun disinyalir 90% total jenis tumbuhan-tumbuhan berkhasiat jamu ada di Indonesia, ternyata hanya terdapat sekitar 9.000-spesies tanaman yang diduga memiliki khasiat obat. Dari jumlah tersebut, baru

sekitar 5% yang dimanfaatkan sebagai bahan fitofarmaka sedangkan sekitar 1000-an jenis tanaman sudah dimanfaatkan untuk bahan baku jamu. Fitofarmaka adalah sediaan obat yang telah dibuktikan keamanan dan khasiatnya, bahan bakunya terdiri dari simplisia atau sediaan galenik yang telah memenuhi persyaratan yang berlaku. Fitofarmaka harus didukung oleh hasil pengujian, dengan protokol pengujian yang jelas dan dapat dipertanggung jawabkan.

Di Indonesia tanaman obat juga sering dikategorikan sebagai tanaman Biofarmaka. Tanaman biofarmaka mencakup 15 (lima belas) jenis tanaman, meliputi jahe, laos/lengkuas, kencur, kunyit, lempuyang, temulawak, temuireng, temukunci, dlingo/dringo, kapulaga, mengkudu/pace, mahkota dewa, kejibeling, sambiloto, dan lidah buaya (Statistik Hortikultura, 2014).

Di Indonesia daerah penghasil tanaman biofarmaka terbesar ada di Provinsi Jawa Tengah. Hal ini dapat dilihat dari sekitar 676 ribu ton tanaman yang dihasilkan di Indonesia pada tahun 2018, 57 ribu ton diantaranya berasal dari Provinsi Jawa Tengah. Dari seluruh kabupaten atau kota yang ada di Jawa Tengah hanya Kota Magelang, Surakarta, Pekalongan dan Tegal yang memiliki produksi tanaman biofarmaka paling kecil. Daerah penghasil di Jawa Tengah yang paling besar ada di Kabupaten Wonogiri. Pada tahun 2017 Kabupaten Wonogiri menghasilkan kurang lebih 34 ribu ton dan pada tahun 2018 turun menjadi sekitar 31 ribu ton. Penurunan ini dikarenakan lahan yang digunakan untuk menanam juga mengalami pengurangan. Dari seluruh kecamatan yang ada di Wonogiri Kecamatan Kismantoro yang hasil produksinya paling besar. Setidaknya 8 ribu ton dihasilkan pada tahun 2017 dan 10 ribu ton tanaman biofarmaka dihasilkan di Kecamatan Kismantoro pada tahun 2018 (Badan Pusat Statistik, 2017).

Jahe merupakan komoditas yang produksinya paling banyak dari semua jenis tanaman yang masuk dalam jenis tanaman biofarmaka. Produksi jahe di Jawa Tengah sendiri dari tahun 2014 hingga tahun 2016 selalu mengalami kenaikan namun pada tahun 2017 dan 2018 mengalami penurunan. Total produksi jahe dari tahun 2014 hingga tahun 2018 adalah 178 ribu ton. Komoditas terbanyak kedua yaitu kunyit dengan total produksi tidak kurang dari 25 ribu ton untuk setiap tahunnya pada 2014

hingga 2018. Total produksi kunyit di Jawa Tengah dari tahun 2014 hingga tahun 2018 adalah 148 ribu ton. Lengkuas sebagai komoditas terbesar ketiga dengan total produksi tidak kurang dari 15 ribu ton untuk tiap tahunnya dari 2014 hingga 2018 dengan total produksi yang mencapai setidaknya 80 ribu ton. Komoditas terbesar keempat adalah kencur dengan total produksi dari tahun 2014 hingga 2018 adalah sebesar 42.435.926 kg (Badan Pusat Statistik, 2017).

Untuk produksi keseluruhan yang ada di Provinsi Jawa Tengah pada tahun 2014 sebesar 112 ribu ton turun menjadi 93 ribu ton pada tahun 2015. Kemudian pada tahun 2016 mengalami kenaikan menjadi 101 ribu ton dan mulai mengalami penurunan pada tahun 2017 menjadi 99 ribu ton. Pada tahun 2018 tetap mengalami penurunan lagi yaitu menjadi 57 ribu ton tanaman. Penurunan pada tahun 2016 hingga tahun 2018 disebabkan karena berkurangnya lahan yang tersedia untuk menanam tanaman biofarmaka (Badan Pusat Statistik, 2017). Hal ini pemerintah sebagai pelaku dan penentu kebijakan harus berupaya bersama dalam rangka meningkatkan dan memelihara produktivitas di sektor tanaman biofarmaka. Salah satu upaya yang dapat rangka meningkatkan dan memelihara produktivitas di sektor tanaman biofarmaka yaitu dengan mengelompokkan ke dalam kelompok-kelompok berdasarkan kesamaan karakteristik yang dimiliki dari setiap wilayah yang ada di Provinsi Jawa Tengah, pengelompokan ini bertujuan agar informasi terkait jenis tanaman biofarmaka yang tersebar di setiap wilayah yang ada menjadi lebih efisien dan spesifik. Salah satu cara dalam melakukan pembagian di setiap kelompok wilayah yaitu menggunakan analisis klaster.

Analisis klaster merupakan salah satu teknik dalam *data mining* yang bertujuan untuk mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok obyek lainnya. *Data mining* sendiri menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT dalam Larose (2006) adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. Sedangkan analisis klaster dalam *data mining* (dikenal juga dengan istilah *clustering*)

adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan (Gorunescu, 2011).

Di antara banyaknya analisis kluster yang ada, terdapat dua jenis analisis kluster yang memiliki algoritma yang masih saling berkaitan, yaitu *K-means* dan *K-medoids Clustering*. Kedua metode tersebut merupakan metode pengklasteran sekatan (*partitioning*) yang tentu saja lebih cepat dibanding metode hierarki dan lebih menguntungkan apabila jumlah objek sangat besar.

K-means adalah metode pengklasteran berbasis jarak yang membagi data ke dalam sejumlah kluster dan algoritma ini bekerja hanya pada atribut numerik. Sedangkan *K-medoids* adalah algoritma yang masih berkaitan dengan algoritma *K-means*, di mana *K-medoids* merupakan versi umum dari algoritma *K-means* yang bekerja dengan mengukur jarak dan mempunyai komputasi yang lebih intensif. Keduanya sama-sama partisional (memecah *dataset* menjadi kelompok-kelompok) dan berusaha meminimalkan *squared error*, jarak antara titik berlabel yang berada dalam kluster dan titik yang ditunjuk sebagai pusat kluster itu. Yang membedakan *K-medoids* dengan *K-means* adalah *K-medoids* memilih data *point* sebagai pusatnya (Yusuf dan Novian, 2014).

Berdasarkan latar belakang yang telah diuraikan tersebut, maka penulis mengambil judul **“Perbandingan Algoritma *K-Means* dan Algoritma *K-Medoids* dalam Pengelompokan Komoditas Tanaman Biofarmaka di Jawa Tengah Tahun 2018”**, di mana alasan penggunaan kedua metode tersebut adalah selain karena lebih cepat dibandingkan dengan metode hierarki dan jumlah observasi sangat besar, penggunaan metode *K-means* dan *K-Medods* pada kasus ini bertujuan untuk melihat perbandingan apakah terdapat perbedaan hasil dari kedua metode pengklasteran tersebut walaupun keduanya masih memiliki algoritma yang saling berkaitan. Dari hasil perbandingan tersebut maka akan diperoleh metode terbaik untuk pengklasteran tanaman biofarmaka di Jawa Tengah tahun 2018.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah yang akan dikaji dalam penelitian ini yaitu:

1. Bagaimana analisis deskriptif tanaman biofarmaka di Provinsi Jawa Tengah pada tahun 2018?
2. Bagaimana hasil pengklasteran komoditas tanaman biofarmaka menggunakan metode *K-means* dan *K-medoids* di Provinsi Jawa Tengah pada tahun 2018?
3. Bagaimana hasil perbandingan pengklasteran komoditas tanaman biofarmaka dengan menggunakan analisis kluster *K-means* dan *K-medoids* pada tahun 2018?

1.3. Batasan Masalah

Agar penelitian ini tidak menyimpang dan mengambang dari tujuan yang semula direncanakan sehingga mempermudah mendapatkan informasi, maka penulis menetapkan batasan-batasan masalah sebagai berikut:

1. Penelitian ini menggunakan data jumlah produksi (dalam kilogram) tanaman Biofarmaka tahun 2018 di Jawa Tengah dengan jumlah sebanyak 15 variabel, yang didapatkan pada *website* Badan Pusat Statistik Jawa Tengah.
2. Penulis menggunakan bantuan *software Microsoft Excel* dan *R Studio* untuk melakukan proses analisis data.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, maka tujuan yang akan dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengetahui hasil analisis deskriptif dari data tanaman biofarmaka di Provinsi Jawa Tengah tahun 2018.
2. Mengetahui hasil pengklasteran komoditas tanaman biofarmaka menggunakan metode *K-means* dan *K-medoids* di Provinsi Jawa Tengah tahun 2018.
3. Mengetahui hasil perbandingan pengklasteran komoditas tanaman biofarmaka menggunakan metode *K-means* dan *K-medoids* di Provinsi Jawa Tengah tahun 2018.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian tugas akhir ini adalah:

1. Memberikan pengetahuan mengenai hasil analisis deskriptif dari data tanaman biofarmaka di Provinsi Jawa Tengah tahun 2018 dengan bantuan *software Microsoft Excel*.
2. Mengetahui implementasi hasil pengklasteran tanaman biofarmaka dengan menggunakan metode *K-means* dan *K-medoids*.
3. Memberikan pengetahuan mengenai hasil perbandingan dari analisis menggunakan metode *K-means* dan *K-medoids* untuk diambil mana metode yang terbaik.
4. Sebagai acuan bagi pemerintah Semarang tentang daerah-daerah yang menghasilkan produksi tanaman biofarmaka.

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan tugas akhir ini dapat diuraikan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini akan dibahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini memaparkan penelitian-penelitian terdahulu yang berhubungan dengan permasalahan yang diteliti dan menjadi acuan konseptual.

BAB III LANDASAN TEORI

Pada bab ini akan dibahas tentang teori-teori dan konsep yang berhubungan dengan penelitian yang dilakukan dan mendukung dalam pemecahan masalahnya. Selain itu, bab ini juga memuat teori-teori dalam pelaksanaan pengumpulan dan pengolahan data serta saat melakukan penganalisaan.

BAB IV METODOLOGI PENELITIAN

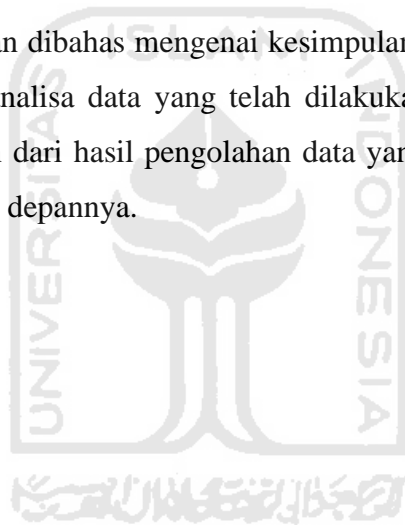
Bab ini memaparkan populasi dan sampel, jenis dan sumber data, metode analisis data, Instrumen penelitian dan tahapan penelitian.

BAB V ANALISIS DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai analisa yang dilakukan terhadap hasil pengumpulan, pengolahan dan analisa data yang diperoleh dari hasil penelitian.

BAB VI PENUTUP

Pada bab ini akan dibahas mengenai kesimpulan yang diperoleh dari hasil penelitian dan analisa data yang telah dilakukan, serta saran-saran yang dapat diterapkan dari hasil pengolahan data yang dapat menjadi masukan yang berguna ke depannya.



BAB II TINJAUAN PUSTAKA

2.1. Kajian Pustaka

Penelitian sebelumnya menjadi acuan untuk penelitian ini guna mengetahui hubungan dengan penelitian terdahulu dan menghindari unsur plagiasi atau duplikasi dengan penelitian sebelumnya. Selain itu, penelitian sebelumnya menunjukkan bahwa penelitian yang dilakukan mempunyai kebermanfaatan, sehingga dapat memberikan kontribusi pada perkembangan ilmu pengetahuan dan mampu memberikan solusi terhadap permasalahan aspek kehidupan dengan teknologi.

TABEL 2.1 PENELITIAN SEBELUMNYA

Peneliti	Judul Penelitian	Metode Penelitian	Hasil
(Flowrensia, 2018)	Perbandingan Penggerombolan <i>K-means</i> dan <i>K-medoids</i> pada Data Yang Mengandung Pencilan	<i>K-means</i> dan <i>K-medoids</i>	metode <i>K-medoids</i> mempunyai nilai rata-rata tingkat salah klasifikasi yang lebih rendah dan signifikan pada kondisi proporsi pencilan 5 persen, sedangkan pada kondisi proporsi pencilan 10 persen dan 15 persen hasil rata-rata tingkat salah klasifikasinya tidak berbeda signifikan dengan metode <i>K-means</i>
(Triyanto, 2015)	Algoritma <i>K-medoids</i> untuk Penentuan Strategi Pemasaran Produk	<i>K-medoids</i>	Hasil dari penelitian ini menghasilkan 5 <i>cluster</i> dengan <i>cluster</i> pertama terdiri dari 909 <i>record</i> transaksi, <i>cluster</i> kedua

Peneliti	Judul Penelitian	Metode Penelitian	Hasil
			<p>terdiri dari 166 <i>record</i> transaksi, <i>cluster</i> ketiga terdiri dari 66 <i>record</i> transaksi, <i>cluster</i> keempat terdiri dari 132 <i>record</i> transaksi, <i>cluster</i> kelima terdiri dari 87 <i>record</i> transaksi.</p> <p>Strategi pemasaran produk dapat dilakukan dengan melakukan promosi pada <i>cluster</i> kelima yang memiliki kombinasi jumlah barang dibeli yang paling tinggi.</p>
(Hilda, 2015)	<p>Perbandingan <i>K-means</i> dan <i>K-medoids Clustering</i> Terhadap Kelayakan Puskesmas di DIY Tahun 2015</p>	<p><i>K-means</i> dan <i>K-medoids</i></p>	<p>metode <i>K-means Clustering</i> dapat diketahui bahwa pada tahun 2015 di DIY terdapat 31 Puskesmas dengan kondisi layak, 52 Puskesmas dengan kondisi cukup layak, dan 38 Puskesmas dengan kondisi kurang layak. Sedangkan dengan metode <i>K-medoids Clustering</i> dapat diketahui bahwa pada tahun 2015 di DIY terdapat 35 Puskesmas dengan kondisi layak, 49 Puskesmas dengan</p>

Peneliti	Judul Penelitian	Metode Penelitian	Hasil
			kondisi cukup layak, dan 37 Puskesmas dengan kondisi kurang layak,
(Ramadhan, Efendi & Mustakim, 2017)	Perbandingan <i>K-means</i> dan <i>Fuzzy C-Means</i> untuk Pengelompokan Data User Knowledge Modeling	<i>K-means</i> dan <i>C-Means</i>	Hasil dari penelitian ini menunjukkan bahwa metode FCM adalah metode yang lebih baik daripada <i>K-means</i> untuk melakukan <i>clustering</i> pada data user knowledge modeling dikarenakan nilai validasinya bernilai mendekati 1.
(Aryuni, Madyatmadja & Miranda, 2018)	Penerapan <i>K-means</i> dan <i>K-medoids Clustering</i> Pada Data Internet Banking Di Bank XYZ	<i>K-means</i> dan <i>K-medoids</i>	Hasil penelitian menunjukkan bahwa nilai k yang paling optimal pada <i>k-means</i> adalah 3 dengan nilai average within <i>centroid</i> distance (W) sebesar 35,241. Sedangkan untuk <i>k-medoids</i> , nilai k yang paling optimal adalah 3 dengan nilai average within <i>centroid</i> distance (W) sebesar 88,849. Algoritme <i>k-means</i> memiliki performa yang lebih baik daripada <i>k-medoids</i> .

Peneliti	Judul Penelitian	Metode Penelitian	Hasil
(Wicaksono, 2019)	Analisis <i>K-Means Cluster</i> Perbandingan 3 Tahun Produksi Tanaman Biofarmaka di Provinsi Jawa Tengah Tahun 2015-2017	<i>K-Means</i>	Dengan <i>cluster</i> pada tahun 2015 sebagai pembanding untuk tahun yang lainnya, maka terdapat total 13 perpindahan kluster pada tahun 2016 dan 6 pada tahun 2017. Wilayah yang mengalami perpindahan kluster antara lain Kab. Wonosobo sebanyak dua kali, Kab. Boyolali dua kali, Kab. Kudus, Kab. Pekalongan dua kali, Kab. Rembang, Kab. Wonogiri, Kab. Semarang, Kab. Magelang, Kab. Brebes dan Kab. Purworejo.

Kelebihan dari penelitian ini dari penelitian-penelitian sebelumnya adalah pengelompokan komoditas tanaman biofarmaka di Provinsi Jawa Tengah tahun 2018 menggunakan perbandingan terbaik dari algoritma *K-means* dan algoritma *K-medoids* dalam membantu pemerintah untuk mengetahui pengelompokan tanaman biofarmaka.

BAB III

LANDASAN TEORI

3.1. Pertanian

Pertanian adalah kegiatan pemanfaatan sumber daya hayati yang dilakukan manusia untuk menghasilkan bahan pangan, bahan baku industri, atau sumber energi, serta untuk mengelola lingkungan hidupnya. Kegiatan pemanfaatan sumberdaya hayati yang termasuk dalam pertanian biasa dipahami orang sebagai budidaya tanaman atau bercocok tanam (bahasa Inggris: *crop cultivation*) serta pembesaran hewan ternak (*raising*), meskipun cakupannya dapat pula berupa pemanfaatan mikroorganisme dan bioenzim dalam pengolahan produk lanjutan, seperti pembuatan keju dan tempe, atau sekedar ekstraksi semata, seperti penangkapan ikan atau eksploitasi hutan. Bagian terbesar penduduk dunia bermata pencaharian dalam bidang-bidang di lingkup pertanian, namun pertanian hanya menyumbang 4% dari PDB dunia. Sejarah Indonesia sejak masa kolonial sampai sekarang tidak dapat dipisahkan dari sektor pertanian dan perkebunan, karena sektor-sektor ini memiliki arti yang sangat penting dalam menentukan pembentukan berbagai realitas ekonomi dan sosial masyarakat di berbagai wilayah Indonesia. Usaha tani (*farming*) adalah bagian inti dari pertanian karena menyangkut sekumpulan kegiatan yang dilakukan dalam budidaya. Petani adalah sebutan bagi mereka yang menyelenggarakan usaha tani, sebagai contoh "petani tembakau" atau "petani ikan". Pelaku budidaya hewan ternak (*livestock*) secara khusus disebut sebagai peternak (Inti Mediatama, 2018). Berdasarkan data BPS tahun 2017, bidang pertanian di Indonesia menyediakan lapangan kerja bagi sekitar 31,86% penduduk meskipun hanya menyumbang sekitar 13,53% dari total pendapatan domestik bruto. Kelompok ilmu-ilmu pertanian mengkaji pertanian dengan dukungan ilmu-ilmu pendukungnya. Karena pertanian selalu terikat dengan ruang dan waktu, ilmu-ilmu pendukung, seperti ilmu tanah, meteorologi, teknik pertanian, biokimia, dan statistika juga dipelajari dalam pertanian (BPS, 2017).

3.2. Cangkupan Pertanian

Pertanian dalam pengertian yang luas yaitu kegiatan manusia untuk memperoleh hasil yang berasal dari tumbuh-tumbuhan dan atau hewan yang pada mulanya dicapai dengan jalan sengaja menyempurnakan segala kemungkinan yang telah diberikan oleh alam guna mengembangbiakkan tumbuhan dan atau hewan tersebut (*Van Aarsten, 1953*). Dalam arti sempit, pertanian diartikan sebagai segala aspek biofisik yang berkaitan dengan usaha penyempurnaan budidaya tanaman untuk memperoleh produksi fisik yang maksimum (*Sumantri, 1980*).

Indonesia merupakan salah satu negara agraris dimana, sebagian besar penduduknya tinggal di perdesaan dengan mata pencaharian sebagai petani. Penduduk Indonesia pada umumnya mengkonsumsi hasil pertanian untuk makanan pokok mereka. Pertanian di Indonesia perlu ditingkatkan produksinya semaksimal mungkin menuju swasembada pangan akan tetapi, tantangan untuk mencapai hal tersebut sangat besar karena luas wilayah pertanian yang semakin lama semakin sempit, penyimpangan iklim, pengembangan komoditas lain, teknologi yang belum modern, dan masalah yang satu ini adalah masalah yang sering meresahkan hati para petani yaitu hama dan penyakit yang menyerang tanaman yang dibudidayakan. Hasil produksi tanaman padi di Indonesia belum bisa memenuhi target kebutuhan masyarakat karena ada di beberapa daerah di Indonesia yang masih mengalami kelaparan (*Agriculture Sector Review Indonesia, 2003*).

Luas pertanian di Indonesia yang semakin menyempit hal inilah yang menjadi tantangan terbesar saat ini yang harus dihadapi akan tetapi, ada cara yang dapat dilakukan untuk mengantisipasinya yaitu dengan cara melakukan pembangunan sektor pertanian. Pembangunan adalah suatu proses perubahan sosial dengan partisipasi yang luas dalam suatu masyarakat yang dimaksudkan untuk kemajuan sosial dan material (termasuk bertambah besarnya kebebasan, keadilan dan kualitas lainnya yang dihargai) untuk mayoritas rakyat melalui kontrol yang lebih besar yang mereka peroleh terhadap lingkungan mereka (*Rogers, 1994*). Pembangunan pertanian adalah upaya-upaya

pengelolaan sumberdaya alam yang dilakukan untuk memastikan kapasitas produksi pertanian jangka panjang dan meningkatkan kesejahteraan petani melalui pilihan-pilihan pendekatan yang ramah terhadap lingkungan (Schultink, 1990).

Pertanian merupakan sektor ekonomi yang utama di Negara-Negara Berkembang. Peran atau kontribusi sektor pertanian dalam pembangunan ekonomi suatu negara menduduki posisi yang penting sekali. Hal ini antara lain disebabkan beberapa faktor (Totok Mardikanto, 2007:3). Pertama, sektor pertanian merupakan sumber persediaan bahan makanan dan bahan mentah yang dibutuhkan oleh suatu Negara. Kedua tekanan-tekanan demografis yang besar di negara-negara berkembang yang disertai dengan meningkatnya pendapatan dari sebagian penduduk menyebabkan kebutuhan tersebut terus meningkat. Ketiga, sektor pertanian harus dapat menyediakan faktor-faktor yang dibutuhkan untuk ekspansi sektor-sektor lain terutama sektor industri. Faktor-faktor ini biasanya berwujud modal, tenaga kerja, dan bahan mentah. Keempat, sektor pertanian merupakan sektor basis dari hubungan-hubungan pasar yang penting berdampak pada proses pembangunan. Sektor ini dapat pula menciptakan keterkaitan kedepan dan keterkaitan kebelakang yang bila disertai dengan kondisi-kondisi yang tepat dapat memberi sumbangan yang besar untuk pembangunan. Kelima, sektor ini merupakan sumber pemasukan yang diperlukan untuk pembangunan dan sumber pekerjaan dan pendapatan dari sebagian besar penduduk negara-negara berkembang yang hidup di pedesaan (Pratomo, 2010).

3.3. Subsektor Hortikultural

Subsektor hortikultura merupakan komponen penting dalam pembangunan pertanian yang terus tumbuh dan berkembang dari waktu ke waktu. Pasar produk komoditas hortikultura bukan hanya untuk memenuhi kebutuhan pasar di dalam negeri saja, melainkan juga sebagai komoditas ekspor yang dapat menghasilkan devisa negara. Di lain pihak, konsumen semakin menyadari arti penting produk hortikultura yang bukan hanya untuk memenuhi kebutuhan pangan semata, tetapi juga mempunyai

manfaat untuk kesehatan, estetika dan menjaga lingkungan hidup (Badan Penelitian dan Pengembangan Pertanian, 2015).

3.4. Statistika Deskriptif

Statistika deskriptif merupakan bagian statistika yang membahas tentang metode-metode untuk menyajikan data sehingga menarik dan informatif. Secara umum statistika deskriptif dapat diartikan sebagai metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Statistika deskriptif juga memberikan informasi hanya mengenai data yang dipunyai dan sama sekali tidak menarik inferensia atau kesimpulan apapun tentang gugus data induknya yang lebih besar (Walpole, 1998).

3.5. Asumsi Analisis Kelompok

Menurut Hair, dkk., (1998) terdapat dua asumsi dalam analisis kelompok yaitu sampel yang representatif dan tidak ada multikolinieritas.

3.5.1 Sampel Representatif

Penggunaan sampel dalam penelitian harus dapat mewakili populasi atau representatif. Penggunaan sampel yang representatif akan memberikan hasil yang maksimal dan sesuai dengan kondisi populasi yang ada. Namun apabila penelitian menggunakan populasi maka dapat disimpulkan bahwa asumsi representatif terpenuhi. (Hair, dkk., 1998).

3.5.2 Tidak Ada Multikolinieritas

Multikolinieritas adalah suatu peristiwa dimana terjadi korelasi yang kuat antara dua atau lebih variabel kelompok. Multikolinieritas merupakan masalah yang perlu diperhatikan dalam analisis multivariat pada umumnya, karena pengaruhnya yang sangat besar dalam menghasilkan solusi, sehingga mengganggu proses analisis. Namun dalam analisis kelompok efeknya berbeda, yaitu variabel-variabel yang terjadi multikolinieritas secara implisit dibobot lebih besar. Karena alasan ini peneliti dianjurkan untuk menguji variabel kelompok mana yang secara substansial menimbulkan multikolinieritas. (Hair, dkk., 1998).

Pengujian terhadap multikolinieritas dalam data salah satunya dengan menggunakan koefisien korelasi. Koefisien korelasi merupakan indeks atau bilangan yang digunakan untuk mengukur keeratan hubungan antar variabel. Perhitungan koefisien korelasi yang dapat digunakan untuk data dengan skala pengukuran interval dan rasio adalah koefisien korelasi *pearson*. (Hasan, 2002). Berikut adalah persamaan koefisien korelasi *pearson*:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.1)$$

Dengan:

r_{xy} = koefisien korelasi antara variabel x dan y

x_i, y_i = variabel bebas x dan y pada data ke- i

\bar{x}, \bar{y} = rata-rata data variabel x dan y

Berikut uji hipotesis untuk melihat multikolinieritas dalam data berdasarkan koefisien korelasi dan *P-value*:

a. Hipotesis Penelitian

$$H_0 : \rho_{xy} = 0$$

$$H_1 : \rho_{xy} \neq 0$$

b. Tingkat Signifikansi

$$\alpha = 0.05$$

c. Daerah Kritis

Tolak H_0 jika: $r_{xy} > r_{tabel}$ atau $P\text{-value} < \alpha$

d. Statistik Uji

Nilai r_{xy}

e. Keputusan

Tolak H_0 atau Gagal Tolak H_0

f. Kesimpulan

Kesimpulan yang dapat ditarik dari keputusan yang diambil

Beberapa hal yang dapat dilakukan apabila data mengandung multikolinieritas yaitu dapat mengurangi variabel dengan jumlah yang sama pada setiap set atau dapat menggunakan salah satu *distance measures*, seperti *Mahalanobis distance* (Hair, dkk., 1998). Selain itu, menurut Supranto (2004) multikolinieritas juga dapat ditangani dengan menggunakan analisis komponen utama yang mereduksi variabel menjadi beberapa faktor sehingga tidak mengandung multikolinieritas.

3.5.3 Analisis Komponen Utama

Analisis faktor merupakan suatu analisis yang digunakan untuk mereduksi dan meringkas data. Faktor-faktor diekstraksi sehingga faktor pertama memberikan andil terbesar terhadap seluruh varian dari seluruh variabel asli, faktor kedua menyumbang terbesar kedua, faktor ketiga menyumbang terbesar ketiga, dan begitu seterusnya sehingga proses pencarian faktor dihentikan setelah sumbangan terhadap seluruh varian variabel mencapai 60% atau lebih. Dimungkinkan pula untuk mengekstraksi faktor sehingga faktor tidak berkorelasi sesamanya seperti di dalam Analisis Komponen Utama atau *Principal Component Analysis* (PCA) (Supranto, 2004).

Dalam analisis komponen utama jumlah varian data yang terkandung dalam semua variabel asli dipertimbangkan, analisis ini direkomendasikan apabila tujuan peneliti adalah menentukan banyaknya faktor yang diekstraksi minimum (sedikit mungkin) tetapi menyerap sebagian besar informasi yang terkandung pada semua variabel asli untuk analisis multivariat selanjutnya (Supranto, 2004).

Sebelum melakukan analisis faktor, dilakukan pengujian terhadap data dengan menggunakan *uji bartlett* dan KMO. Berikut penjelasan untuk kedua uji tersebut:

a. Uji *Bartlett*

Pengujian ini digunakan untuk melihat apakah variabel yang digunakan berkorelasi dengan variabel lainnya. Pengujian dilakukan dengan menggunakan statistik *chi-square*, apabila nilai *chi-square* lebih dari nilai *chi-square* tabel maka variabel mengandung korelasi. (Usman dan Sobari, 2013). Berikut uji hipotesis untuk uji *bartlett*:

i. Hipotesis

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

ii. Tingkat Signifikansi

$$\alpha = 0.05$$

iii. Daerah Kritis

$$\text{Tolak } H_0 \text{ jika: } \chi_{hitung}^2 > \chi_{(\alpha; db)}^2 \text{ atau } P\text{-value} < \alpha$$

iv. Statistik Uji

$$\text{Nilai } \chi_{hitung}^2 \text{ dan } P\text{-value}$$

v. Keputusan

$$\text{Tolak } H_0 \text{ atau Gagal Tolak } H_0$$

vi. Kesimpulan

Kesimpulan yang dapat ditarik dari keputusan yang diambil

b. Uji *Kaiser Mayer Olkin* (KMO)

KMO merupakan indeks yang digunakan untuk meneliti ketepatan analisis faktor. Analisis faktor tepat untuk digunakan ketika nilai KMO berada diantara 0,5 hingga 1, apabila kurang dari 0,5 analisis faktor dikatakan tidak tepat. (Supranto, 2004). Berikut persamaan untuk menghitung nilai KMO:

$$KMO = \frac{\sum_i^n \sum_{j \neq i}^n r_{ij}^2}{\sum_i^n \sum_{j \neq i}^n r_{ij}^2 + \sum_i^n \sum_{j \neq i}^n a_{ij}^2} \quad (3.2)$$

Dengan:

$$a_{ij} = - \frac{v_{ij}}{\sqrt{v_{ii}v_{jj}}}$$

- a_{ij} = koefisien korelasi parsial dari variabel i dan j
 r_{ij} = koefisien korelasi sederhana dari variabel i dan j
 v_{ij} = invers matriks korelasi dari variabel i dan j
 i, j = 1, 2, 3... n

Penentuan banyak faktor yang dibentuk dapat dilakukan dengan dua metode, yaitu dengan melihat nilai *eigenvalues* dan persentase varian. Berikut penjelasan untuk masing-masing metode:

1. Penentuan berdasarkan *Eigenvalues*

Suatu *eigenvalues* menunjukkan besarnya sumbangan dari faktor terhadap varian seluruh variabel asli. Dalam pendekatan ini, hanya faktor dengan nilai *eigenvalues* yang lebih dari satu yang dipertahankan atau dimasukkan dalam model. Sedangkan yang kurang dari satu tidak dimasukkan karena dianggap tidak lebih baik dari variabel asli, sebab variabel asli telah dibakukan yang berarti rata-ratanya nol dan variannya satu (Supranto, 2004).

2. Penentuan berdasarkan Persentase Varian

Banyaknya faktor yang diekstraksi ditentukan sedemikian rupa sehingga kumulatif persentase varian yang diekstraksi oleh faktor mencapai suatu level tertentu yang memuaskan. Menurut Supranto (2004), ekstraksi faktor dihentikan ketika kumulatif persentase varian sudah mencapai minimal 60% atau 75% dari seluruh varian variabel asli.

Ekstraksi faktor yang belum menghasilkan komponen faktor utama yang jelas mengharuskan dilakukannya rotasi faktor. Tujuan dari rotasi faktor ini yaitu agar dapat memperoleh struktur faktor yang lebih sederhana sehingga mudah diinterpretasikan. Ada beberapa metode rotasi faktor yang dapat digunakan:

- a. *Varimax Method*, adalah metode rotasi orthogonal untuk meminimalisasi jumlah indikator yang mempunyai *factor loading* tinggi pada tiap faktor.
- b. *Quartimax Method*, adalah metode rotasi yang meminimalisasi jumlah faktor yang digunakan untuk menjelaskan indikator.

- c. *Equamax Method*, adalah metode gabungan antara *varimax method* dan *quartimax method* (Widarjono, 2010).

Tujuan dilakukan analisis faktor pada penelitian ini yaitu untuk mencari variabel yang independen atau tidak mengandung multikolinieritas, maka penghitungan skor atau nilai faktor harus dilakukan untuk membentuk variabel baru guna melakukan analisis selanjutnya, dalam hal ini analisis kelompok. Berikut persamaan untuk menghitung nilai faktor :

$$F_i = w_{i1}X_1 + w_{i2}X_2 + \dots + w_{ij}X_j \quad (3.3)$$

Dengan :

F_i = skor faktor untuk faktor ke- i

w_{ij} = *weight* atau *factor score coefficient* faktor ke- i dan variabel ke- j

X_j = variabel ke- j

i = banyak faktor; $i = 1, 2, 3, \dots, n$

j = banyak variabel; $j = 1, 2, 3, \dots, n$

3.6. *Clustering*

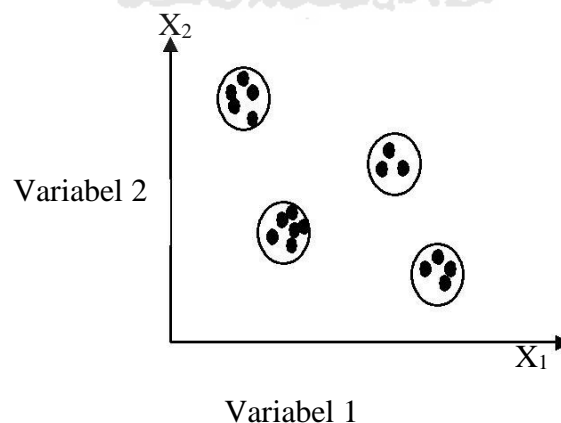
Analisis *cluster* merupakan suatu teknik analisis statistik yang ditujukan untuk menempatkan sekumpulan obyek ke dalam dua atau lebih grup berdasarkan kesamaan-kesamaan obyek atas dasar berbagai karakteristik (Simamora, 2005).

Dalam analisis cluster terdapat beberapa istilah penting yang perlu diketahui yaitu:

- Agromeration Schedule*, merupakan daftar yang memberikan informasi tentang obyek atau kasus yang akan dikelompokkan di setiap tahap pada proses analisis cluster dengan metode hirarki.
- Rata-rata cluster (*Cluster Centroid*), adalah nilai rata-rata variabel dari semua obyek atau observasi dalam cluster tertentu.
- Pemusatan *cluster* (*Cluster Centrens*), adalah titik awal dimulai pengelompokan di dalam *cluster* non hirarki.
- Keanggotaan *cluster* adalah keanggotaan yang menunjukkan *cluster* untuk tiap obyek yang menjadi anggotanya.

- e. Dendogram yaitu suatu alat grafis untuk menyajikan hasil dari analisis cluster yang dilakukan oleh peneliti. Dendogram berguna untuk menunjukkan anggota cluster yang ada jika akan ditentukan berapa cluster yang seharusnya dibentuk.
- f. Jarak antara pusat cluster (*Distance Between Cluster Center*) merupakan jarak yang menunjukkan bagaimana terpisahnya pasangan individu.

Analisis cluster meneliti seluruh hubungan interdependensi, tidak ada perbedaan antara variabel bebas dan tak bebas (*Independent* dan *Dependent Variables*) dalam analisis kelompok. Tidak adanya perbedaan ini menunjukkan bahwa analisis kelompok tidak dipergunakan untuk mengetahui pengaruh dari variabel bebas terhadap variabel tidak bebas, namun hanya mengklasifikasikan objek ke dalam kelompok yang relatif homogen. Menurut Supranto (2004) analisis kelompok adalah metode yang digunakan untuk mengklasifikasi objek atau kasus (responden) ke dalam kelompok yang relatif homogen, yang disebut cluster atau kelompok. Objek/ kasus dalam setiap kelompok cenderung mirip satu sama lain dan berbeda jauh (tidak sama) dengan objek dari kelompok lainnya. Selain itu, setiap objek hanya masuk ke dalam satu kelompok saja, tidak terjadi tumpang tindih (*overlapping* atau *interaction*), seperti pada gambar 3.1 di bawah:



Gambar 3.1 Pengelompokan ideal

(Sumber: *Analisis Multivariat, Arti dan Interpretasi*, Supranto, 2004)

Gambar 3.4 menunjukkan hasil pengelompokan yang ideal, dimana setiap objek/kasus hanya masuk atau menjadi anggota dari salah satu kelompok (tidak mungkin menjadi anggota dari dua kelompok atau lebih). Analisis kelompok terbagi menjadi dua, yaitu metode hierarki dan non hierarki. Berikut penjelasan untuk masing-masing jenis:

3.6.1 *Hierarchical Methods* (Metode Hierarki)

Metode ini biasa digunakan untuk individu yang tidak terlalu banyak, dan jumlah kelompok yang hendak dibentuk belum diketahui. Pengelompokan ini disajikan dalam bentuk dendogram, yang mirip dengan struktur diagram pohon atau *tree diagram* (Usman dan Sobari, 2013). Metode hierarki terbagi menjadi dua, yaitu *Agglomerative* atau Metode Penggabungan dan *Divisive* atau Metode Pembagian. Metode *agglomerative* dimulai dengan setiap objek dalam satu kelompok yang terpisah. Kelompok dibentuk dengan mengelompokkan objek ke dalam kelompok yang semakin membesar. Proses ini dilanjutkan sampai semua objek menjadi anggota dari suatu kelompok tunggal (*a single cluster*). Sedangkan metode *divisive* dimulai dari semua objek dikelompokkan menjadi kelompok tunggal. Kemudian kelompok tersebut dibagi atau dipisah, sampai setiap objek berada dalam kelompok yang terpisah (Supranto, 2004).

Beberapa metode pengklasteran hierarki antara lain.

- a. *Single linkage* (jarak terdekat atau tautan tunggal), pengklasteran ini memberikan hasil bila kelompok-kelompok digabungkan menurut jarak antara anggota-anggota yang terdekat di antara dua kelompok (Prasetyo, 2012).
- b. *Complete linkage* (jarak terjauh atau tautan lengkap), pengklasteran ini terjadi bila kelompok-kelompok digabungkan menurut jarak antara anggota-anggota yang terjauh di antara dua kelompok (Prasetyo, 2012).
- c. *Average linkage* (jarak rata-rata atau tautan rata-rata), pengklasteran ini menggabungkan objek menurut jarak rata-rata pasangan-pasangan anggota masing-masing pada himpunan antara dua kelompok (Prasetyo, 2012).

- d. *Ward's method*, metode ini menggunakan perhitungan yang lengkap dan memaksimalkan homogenitas di dalam satu kelompok (Dillon dan Goldstein, 1984).

3.6.2 Pengklasteran Sekatan (*Partitioning*) atau *Non Hierarki*

Pengklasteran berbasis *partitioning* menghasilkan partisi dari data sehingga objek dalam klaster lebih mirip satu sama lain daripada objek yang ada dalam klaster lain (Triyanto, 2015). Berbeda dengan pengklasteran hierarki, prosedur pengklasteran sekatan tidak dilakukan secara bertahap dan jumlah klasternya juga ditentukan terlebih dahulu (Machfudhoh dan Wahyuningsih, 2013). Beberapa metode pengklasteran sekatan yang digunakan dalam penelitian ini antara lain.

a. *K-means Clustering*

Algoritma *K-means* merupakan salah satu algoritma dengan partitional, karena *K-means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya (Madhulatha, 2012).

Algoritma *K-means* menggunakan proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan menghasilkan jumlah *cluster* akhir sebagai *output*. Jika algoritma diperlukan untuk menghasilkan *cluster* K maka akan ada K awal dan K akhir. Metode *K-means* akan memilih pola k sebagai titik awal *centroid* secara acak. Jumlah iterasi untuk mencapai *cluster centroid* akan dipengaruhi oleh calon *cluster centroid* awal secara random dimana jika posisi *centroid* baru tidak berubah. Nilai K yang dipilih menjadi pusat awal, akan dihitung dengan menggunakan rumus *Euclidean Distance* yaitu mencari jarak terdekat antara titik *centroid* dengan data/objek. Data yang memiliki jarak pendek atau terdekat dengan *centroid* akan membentuk sebuah *cluster* (C.M. HUNG, 2005).

K-means Clustering merupakan algoritma pengklasteran yang paling sederhana dibanding algoritma pengklasteran yang lain. *K-means Clustering* mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan

paling banyak dipraktikkan dalam tugas data *mining*. Pengklasteran merupakan suatu metode untuk mengelompokkan dokumen di mana dokumen dikelompokkan dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu *query* (Grossman dan Frieder, 2004).

K-means Clustering merupakan algoritma pengklasteran iteratif yang melakukan partisi set data ke dalam sejumlah k klaster yang sudah ditetapkan di awal. *K-means Clustering* sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-means Clustering* menjadi salah satu algoritma yang paling penting dalam bidang data *mining* (Wu dan Kumar, 2009 dalam Prasetyo, 2012).

K-means merupakan salah satu metode pengklasteran data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih klaster. Metode ini mempartisi data ke dalam klaster sehingga data berkarakteristik berbeda diklasterkan ke dalam klaster yang lain. Adapun tujuan pengklasteran data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengklasteran, yang pada umumnya berusaha meminimalkan fungsi objektif yang diset dalam proses pengklasteran, yang pada umumnya berusaha meminimalkan variasi di dalam suatu klaster dan memaksimalkan variasi antar klaster. Adapun langkah-langkah untuk *K-means Clustering* adalah sebagai berikut (Prasetyo, 2012):

1. Inisialisasi: tentukan nilai k sebagai jumlah klaster yang diinginkan dan matriks jarak yang diinginkan.
2. Pilih k data dari set data X sebagai *centroid*.
3. Alokasikan semua data ke *centroid* terdekat dengan matriks jarak yang sudah ditetapkan (memperbarui klaster ID pada setiap data).
4. Hitung kembali *centroid* berdasarkan data yang mengikuti klaster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu tidak ada data yang berpindah klaster.

Contoh penerapan *K-means Clustering* (Prasetyo,2012):

Tabel 3.1 Kondisi awal data

Data ke- <i>i</i>	X	Y
1	1	1
2	4	1
3	6	1
4	1	2
5	2	3
6	5	3
7	2	5
8	3	5
9	2	6
10	3	8

1. Inisialisasi, ditentukan k (jumlah kluster) dari n objek adalah 3, kemudian dipilih k data sebagai *centroid* awal, misalnya dipilih data ke 2, 4, dan 6.

Tabel 3.2 *Centroid* awal

<i>Centroid</i>	x	y
1	4	1
2	1	2
3	5	3

2. Iterasi 1, hitung jarak setiap data ke *centroid* terdekat. *Centroid* terdekat akan menjadi kluster yang diikuti oleh data tersebut. Dalam penelitian ini, jarak yang digunakan adalah jarak *Euclidean*. Berikut contoh perhitungan jarak ke setiap *centroid* pada data ke 1:

$$d_{x_1,c_1} = \sqrt{(-4)^2 - (-1)^2} = 3$$

$$d_{x_1,c_2} = \sqrt{(-1)^2 - (-2)^2} = 1$$

$$d_{x_1,c_3} = \sqrt{(-5)^2 - (-3)^2} = 4,47$$

Cara yang sama dapat dilakukan untuk perhitungan jarak ke setiap *centroid* pada data ke 2 sampai data ke 10, sehingga diperoleh tabel 3.3 berikut.

Tabel 3.3 Hasil perhitungan jarak ke setiap *centroid*

Data ke- <i>i</i>	Jarak ke <i>centroid</i>			Terdekat	Klaster yang diikuti
	1	2	3		
1	3	1	4,47	1	2
2	0	3,16	2,24	0	1
3	2	5,10	2,24	2	1
4	3,16	0	4,12	0	2
5	2,83	1,41	3	1,41	2
6	2,24	4,12	0	0	3
7	4,47	3,15	3,61	3,15	2
8	4,12	3,61	2,83	2,83	3
9	5,39	4,12	4,24	4,12	2
10	7,07	6,32	5,39	5,39	3

Selanjutnya hitung nilai *centroid* yang baru untuk setiap klaster berdasarkan data yang bergabung pada setiap klasternya. Adapun persamaan untuk perhitungan nilai *centroid* adalah sebagai berikut:

$$c_{kj} = \frac{1}{N_k} \sum_{i=1}^{N_k} x_{ij} \quad (3.4)$$

Dengan :

c_{kj} = *centroid* klaster ke k pada variabel ke j

N_k = jumlah data yang tergabung dalam klaster ke k

x_{ij} = nilai objek ke i pada variabel ke j

Sehingga perhitungan *centroid* baru untuk klaster 1 adalah

Tabel 3.4 Anggota kluster

Data anggota	X	Y
2	4	1
3	6	1

$$c_{11} = \frac{4+6}{2} = 5$$

$$c_{12} = \frac{1+1}{2} = 1$$

Perhitungan *centroid* baru untuk kluster 2 adalah

Tabel 3.5 Anggota kluster 2

Data anggota	X	Y
1	1	1
4	1	2
5	2	3
7	2	5
9	2	6

$$c_{21} = \frac{1+1+2+2+2}{5} = 1,60$$

$$c_{22} = \frac{1+2+3+5+6}{5} = 3,40$$

Perhitungan *centroid* baru untuk kluster 3 adalah

Tabel 3.6 Anggota kluster 3

Data anggota	X	Y
6	5	3
8	3	5
10	3	8

$$c_{31} = \frac{5+3+3}{3} = 3,67$$

$$c_{32} = \frac{3+5+8}{3} = 5,34$$

Sehingga diperoleh *centroid* baru pada table berikut.

Tabel 3.7 *Centroid* baru hasil iterasi 1

<i>Centroid</i>	<i>x</i>	<i>y</i>
1	5	1
2	1,60	3,40
3	3,67	5,34

3. Iterasi 2, hitung kembali jarak setiap data ke *centroid* baru hasil iterasi 1. Berikut contoh perhitungan jarak ke setiap *centroid* baru hasil iterasi 1 pada data ke 1:

$$d_{x_1, c_1} = \sqrt{(-5)^2 + (-1)^2} = 4$$

$$d_{x_1, c_2} = \sqrt{(-1,60)^2 + (-3,40)^2} = 2,47$$

$$d_{x_1, c_3} = \sqrt{(-3,67)^2 + (-5,34)^2} = 5,02$$

Cara yang sama dapat dilakukan untuk perhitungan jarak ke setiap *centroid* baru hasil iterasi 1 pada data ke 2 sampai data ke 10, sehingga diperoleh tabel 3.8 berikut.

Tabel 3.8 Hasil perhitungan jarak ke setiap *centroid* baru hasil iterasi 1

Data ke- <i>i</i>	Jarak ke <i>centroid</i>			Terdekat	Klaster yang diikuti
	1	2	3		
1	4	2,47	5,02	2,47	2
2	1	3,39	4,32	1	1
3	1	5,01	4,92	1	1
4	4,12	1,52	4,20	1,52	2
5	3,61	0,57	2,80	0,57	2
6	2	3,42	2,69	2	1
7	5	1,65	1,63	1,63	3
8	4,47	2,13	0,67	0,67	3
9	5,83	2,63	1,75	1,75	3

10	7,28	4,81	2,77	2,77	3
----	------	------	------	------	---

Karena masih terdapat data yang berpindah kluster, maka langkah selanjutnya adalah menghitung *centroid* hasil iterasi 2 yang dilanjutkan dengan iterasi 3 hingga kondisi konvergen tercapai.

b. *K-medoids Clustering*

K-medoids Clustering, juga dikenal sebagai *Partitioning Around Medoids (PAM)*, adalah varian dari metode *K-means*. Hal ini didasarkan pada penggunaan *medoids* bukan dari pengamatan *mean* yang dimiliki oleh setiap kluster, dengan tujuan mengurangi sensitivitas dari partisi sehubungan dengan nilai ekstrim yang ada dalam dataset (Vercellis, 2009).

K-medoids Clustering hadir untuk mengatasi kelemahan *K-means Clustering* yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data (Han dan Kamber, 2006).

K-medoids Clustering menggunakan metode pengklasteran partisi untuk mengklasterkan sekumpulan n objek menjadi sejumlah k kluster. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah kluster. Objek yang mewakili sebuah kluster disebut dengan *medoids*. Kluster dibangun dengan menghitung kedekatan yang dimiliki antara *medoids* dengan objek *non medoids* (Setyawati, 2017).

Menurut Han dan Kamber (2006), tahapan *K-medoids Clustering* adalah sebagai berikut:

1. Secara acak pilih k objek pada sekumpulan n objek sebagai *medoids*.
2. Ulangi.
3. Tempatkan objek *non medoids* ke dalam kluster yang paling dekat dengan *medoids*.
4. Secara acak pilih Orandom (sebuah objek *non medoids*).
5. Hitung total cost, S , dari pertukaran *medoids* O_j dengan Orandom .
6. Jika $S < 0$ maka tukar O_j dengan Orandom, untuk membentuk sekumpulan k objek baru sebagai *medoids*.

7. Hingga tidak ada perubahan.

Contoh penerapan *K-medoids Clustering*:

Tabel 3.9 Kondisi awal data

Data ke i	X	Y
1	66	402
2	31	182
3	49	258
4	50	289
5	51	281
6	65	464
7	75	387
8	162	964
9	113	706
10	61	329
11	48	290
12	59	311

Sumber : Data Simulasi

1. Ditentukan k (jumlah kluster) dari n objek adalah 3.
2. Tentukan *centroid* awal sebagai *medoids* dengan asumsi seperti pada tabel 3.15 berikut.

Tabel 3.10 *Medoids*

Nama	Keterangan	x	y
$C1$	Diambil data ke 8 sebagai pusat kluster ke 1	162	964
$C2$	Diambil data ke 7 sebagai pusat kluster ke 2	75	387
$C3$	Diambil data ke 2 sebagai pusat kluster ke 3	31	182

3. Tempatkan objek-objek *non medoids* ke dalam kluster yang paling dekat dengan *medois* berdasarkan jarak *Euclidean*. Berikut contoh perhitungan jarak pada data ke 1:

$$d_{x_1, c_1} = \sqrt{(66 - 162)^2 + (402 - 964)^2} = 570,14$$

$$d_{x_1, c_2} = \sqrt{(66 - 75)^2 + (402 - 387)^2} = 16,16$$

$$d_{x_1, c_3} = \sqrt{(66 - 31)^2 + (402 - 182)^2} = 329,28$$

Cara yang sama dapat dilakukan untuk perhitungan jarak data ke 2 sampai data ke 10, sehingga diperoleh sebagai berikut.

Tabel 3.11 Hasil perhitungan jarak ke setiap *medoids*

Data ke- <i>i</i>	Jarak ke <i>medoids</i>			Terdekat	Klaster yang diikuti
	<i>C1</i>	<i>C2</i>	<i>C3</i>		
1	5,70	16,16	329,28	16,16	2
2	796,07	314,87	0	0	3
3	714,00	130,05	273,08	130,05	2
4	684,23	100,44	282,08	100,44	2
5	691,96	108,06	278,21	108,06	2
6	509,32	77,32	374,22	77,32	2
7	583,98	0	314,87	0	2
8	0	583,98	796,07	0	1
9	262,61	321,62	560,16	262,61	1
10	642,98	59,03	290,02	59,03	2
11	683,57	99,92	284,31	99,92	2
12	661,07	77,10	283,10	77,10	2
Jumlah	6.799,94	1.888,56	4.065,40		
Total Cost	12.753,90				

4. Tentukan objek *non medoids* dengan asumsi sebagai berikut.

Tabel 3.12 *Non Medoids*

Nama	Keterangan	<i>x</i>	<i>y</i>
<i>D1</i>	Diambil data ke 3 sebagai pusat klaster ke 1	49	259
<i>D2</i>	Diambil data ke 5 sebagai pusat klaster ke 2	51	281

<i>D3</i>	Diambil data ke 9 sebagai pusat kluster ke 3	113	706
-----------	--	-----	-----

5. Ulangi langkah 3 untuk objek *non medoids* sehingga diperoleh sebagai berikut berikut.

Tabel 3.13 Hasil perhitungan jarak ke setiap *non medoids*

Data ke- <i>i</i>	Jarak ke <i>medoids</i>			Terdekat	Kluster yang diikuti
	<i>D1</i>	<i>D2</i>	<i>D3</i>		
1	144,01	121,93	307,61	121,93	2
2	273,08	278,21	560,16	273,08	1
3	0	22,09	451,56	0	1
4	30,02	8,06	421,73	8,06	2
5	22,09	0	429,50	0	2
6	205,62	183,53	246,71	183,53	2
7	130,05	108,06	321,62	108,06	2
8	714	691,96	262,51	262,61	3
9	451,56	429,50	0	0	3
10	71,02	49,03	380,57	49,03	2
11	31,02	9,49	421,05	9,49	2
12	52,95	31,05	398,67	31,05	1
Jumlah	2.125,42	1.932,91	4.201,79		
Total Cost	8.260,13				

6. Hitung nilai *S* dengan persamaan sebagai berikut (Setyawati, 2017):

$$S = \text{total cost baru} - \text{total cost lama}$$

Dengan

$$S = \text{selisih}$$

$$\text{Total cost baru} = \text{total cost untuk non medoids}$$

$$\text{Total cost lama} = \text{total cost untuk medoids}$$

$$\text{Cost} = \text{nilai jarak terdekat}$$

Sehingga diperoleh:

$$S = 8.260,13 - 12.752,90 = -4.493,77$$

7. Karena nilai $S < 0$ maka tukar *non medoids* dengan *medoids*, sehingga D1, D2, dan D3 menjadi objek *medoids*.
8. Tentukan objek *non medoids* baru dengan asumsi sebagai berikut.

Tabel 3.14 Non *Medoids* baru

Nama	Keterangan	x	y
<i>E1</i>	Diambil data ke 8 sebagai pusat klaster ke 1	162	964
<i>E2</i>	Diambil data ke 10 sebagai pusat klaster ke 2	61	329
<i>E3</i>	Diambil data ke 12 sebagai pusat klaster ke 3	59	311

9. Ulangi langkah 3 untuk objek *non medoids* baru sehingga diperoleh sebagai berikut.

Tabel 3.15 Hasil perhitungan jarak ke setiap *non medoids*

Data ke- i	Jarak ke <i>medoids</i>			Terdekat	Klaster yang diikuti
	<i>E1</i>	<i>E2</i>	<i>E3</i>		
1	570,14	73,71	91,27	73,71	2
2	796,07	290,02	283,10	283,10	3
3	714	71,02	52,95	52,95	3
4	684,23	41,48	23,77	23,77	3
5	691,96	49,03	31,05	31,05	3
6	509,32	135,06	153,12	135,06	2
7	583,98	59,03	77,10	59,03	2
8	0	642,98	661,07	0	1
9	262,61	380,57	398,67	262,61	1
10	642,98	0	18,11	0	2
11	683,57	41,11	23,71	23,71	3
12	661,07	18,11	0	0	3
Jumlah	6.799,94	1.801,59	1.813,92		

Total Cost	10.415,45
------------	-----------

10. Hitung kembali nilai S sebagai berikut:

$$S = 10.415,45 - 8.260,13 = 2.155,32$$

11. Karena nilai $S > 0$ maka proses pengklasteran dihentikan. Sehingga diperoleh anggota tiap klaster sebagai berikut.

Tabel 3.16 Hasil pengklasteran data simulasi dengan *K-medoids Clustering*

Data ke i	X	Y	Klaster yang diikuti
1	66	402	2
2	31	182	1
3	49	258	1
4	50	289	2
5	51	281	2
6	65	464	2
7	75	387	2
8	162	964	3
9	113	706	3
10	61	329	2
11	48	290	2
12	59	311	1

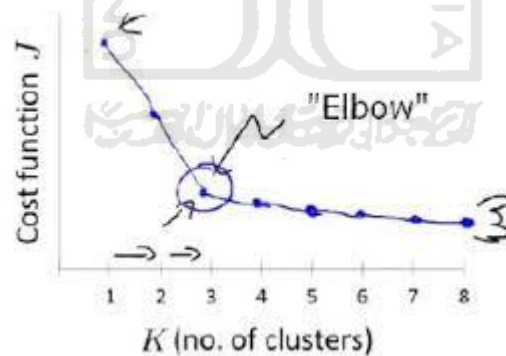
3.7. Penentuan Jumlah Cluster menggunakan metode *Elbow*

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik (Madhulatha, 2012). Metode ini memberikan ide/gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. Dan selain itu persentase perhitungan yang dihasilkan

menjadi pembanding antara jumlah *cluster* yang ditambah. Hasil persentase yang berbeda dari setiap nilai *cluster* dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai *cluster* tersebut yang terbaik. Untuk mendapatkan perbandingannya adalah dengan menghitung *SSE* (*Sum of Square Error*) dari masing-masing nilai *cluster*. Karena semakin besar jumlah *cluster* K maka nilai *SSE* akan semakin kecil (Irwanto, 2012).

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_K} \|X_i - C_K\|_2^2 \quad (3.5)$$

Setelah dilihat akan ada beberapa nilai K yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai K akan turun secara perlahan-lahan sampai hasil dari nilai K tersebut stabil. Misalnya nilai *cluster* $K=2$ ke $K=3$, kemudian dari $K=3$ ke $K=4$, terlihat penurunan drastis membentuk siku pada titik $K=3$ maka nilai *cluster* k yang ideal adalah $K=3$ (Kodinariya dan Makwana, 2013).



Gambar 3.2 Grafik metode *Elbow*

Algoritma Metode *Elbow* dalam menentukan nilai K yaitu:

1. Mulai
2. Inisialisasi awal nilai K
3. Naikkan nilai K
4. Hitung hasil *sum of square error* dari tiap nilai K

5. Melihat hasil *sum of square error* dari nilai K yang turun secara drastis
6. Tetapkan nilai K yang berbentuk siku
7. Selesai.

3.8. Ukuran Jarak dalam Pengklasteran

3.8.1 Jarak *Euclidean*

Jarak *Euclidean* adalah ukuran ketidakmiripan yang sering digunakan, merupakan jarak geometris diruang multidimensional (Anonim, 2008c). Jarak ini digunakan jika variabel yang digunakan tidak berkorelasi satu sama lain atau saling *ortogonal*, yang memiliki satuan dan skala pengukuran yang sama (Anonim, 2008e). Jarak ini cukup fleksibel untuk dilakukan modifikasi dalam mengatasi kelemahan data. Jarak *Euclidean* merupakan jarak terpendek yang didapat antara dua titik dalam perhitungan.

Ukuran jarak *Euclidean* antar dua objek $\mathbf{x}' = [x_1, x_2, \dots, x_p]$ dan $\mathbf{y}' = [y_1, y_2, \dots, y_p]$ yang berdimensi p adalah (Johnson & Winchern, 2002):

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (3.6)$$

3.8.2 Jarak Kuadrat *Euclidean*

Jarak kuadrat *Euclid* adalah jumlah kuadrat yang berbeda dari nilai antara dua objek pada seluruh variabel (Fiedling, 2007). Jarak kuadrat *Euclid* antara dua objek x dan pada ruang berdimensi p adalah

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^p (x_i - y_i)^2 \quad (3.7)$$

3.8.3 Jarak *Mahalanobis*

Jika diantara variabel-variabel yang digunakan saling berkorelasi maka perlu dilakukan transformasi data. Transformasi ini berfungsi untuk menghilangkan pengaruh keragaman data atau dengan kata lain, semua variabel memberikan kontribusi yang sama untuk jarak (Anonim, 2008a). Jika tidak dilakukan transformasi data dapat digunakan jarak *Mahalanobis*.

Jarak ini menggunakan variabel dengan sampel matriks varian-kovarian, karena matriks kovarian juga menggunakan rata-rata korelasi diantara variabel. Jarak *Mahalanobis* antara objek x dan y dapat dinyatakan dalam bentuk berikut ini:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})} \quad (3.8)$$

dimana \mathbf{S}^{-1} adalah invers matriks varian-kovarian.

3.8.4 Jarak *Manhattan* atau *City-Block*

Jarak *City-Block* atau *Manhattan* adalah jumlah nilai perbedaan mutlak untuk setiap variabel (Supranto, 2004). Ukuran jarak ini menghasilkan jarak yang serupa dengan jarak Euclid untuk beberapa kasus tertentu. Jarak ini mempunyai kelebihan yaitu dapat mendeteksi keberadaan *outlier* dengan baik (Agusta, 2007). Jarak *City-Block* atau *Manhattan* diturunkan dari persamaan jarak *Mahalanobis* untuk $m=1$, sehingga jarak *City-Block* atau *Manhattan* dinyatakan sebagai berikut:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^p |x_i - y_i| \quad (3.9)$$

3.9. Pemilihan Metode Terbaik dengan Simpangan Baku

Sebuah metode pengelompokan yang baik jika mempunyai nilai simpangan baku dalam kelompok (s_w) yang minimum dan nilai simpangan baku antar kelompok (s_b) yang maksimum (Barakbah dan Arai, 2004). Dengan rumus (s_w) (Bunkers et.al, 1996:136) sebagai berikut:

$$S_w = \frac{1}{K} \sum_{k=1}^K S_k \quad (3.10)$$

Dengan :

K = Banyaknya kelompok yang terbentuk

S_k = Simpangan baku kelompok ke- k

Jika diberikan *cluster* c_k , dimana $k=1, \dots, p$ dan setiap *cluster* memiliki anggota x_i , dimana $i=1, \dots, n$ dan n adalah jumlah anggota dari setiap *cluster* dan \bar{x}_k adalah rata-rata dari *cluster* k maka untuk mencari nilai simpangan baku ke- k (S_k) digunakan rumus berikut:

$$S_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_k)^2} \quad (3.11)$$

Jika terdapat rata-rata variabel dalam setiap *cluster* k (\bar{X}_k) maka komponen dari setiap *cluster* berbeda, dan simpangan baku antar kelompok (s_b) dapat dirumuskan sebagai berikut :

$$S_b = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2} \quad (3.12)$$

Dengan :

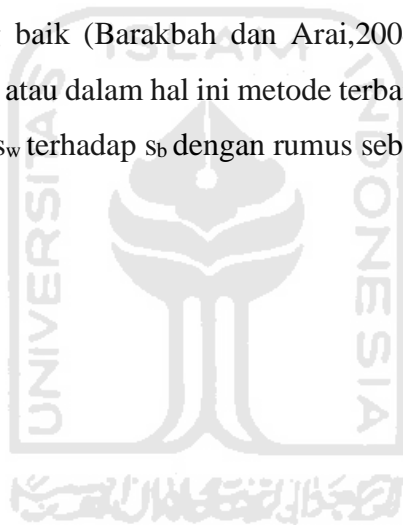
K = Banyaknya kelompok yang terbentuk

\bar{X}_k = Rataan kelompok ke- k

\bar{X} = Rataan keseluruhan kelompok

Pengelompokan yang baik (Barakbah dan Arai,2004) akan memiliki nilai s_w minimum dan s_b maksimum atau dalam hal ini metode terbaik menghasilkan nilai rasio simpangan baku minimum s_w terhadap s_b dengan rumus sebagai berikut:

$$S = \frac{s_w}{s_b} \quad (3.13)$$



BAB IV METODOLOGI PENELITIAN

4.1. Populasi dan Sampel Penelitian

Populasi yang digunakan dalam penelitian ini adalah tanaman biofarmaka di Jawa Tengah Tahun 2018. Sedangkan sampel yang digunakan dalam penelitian ini adalah data jumlah produksi tanaman biofarmaka Tahun 2018.

4.2. Jenis dan Sumber Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari website Badan Pusat Statistik Provinsi Jawa Tengah dengan data Produksi Tanaman Biofarmaka Provinsi Jawa Tengah Tahun 2018 dan diperoleh 16 Variabel.

4.3. Variabel Penelitian dan Definisi Operasional Variabel

Variabel yang di gunakan pada penelitian ini di tampilkan dalam tabel 4.1 berikut tentang penjelasan dan definisi operasional penelitian:

Tabel 4.1 Definisi Operasional Variabel

Nama Variabel	Definisi Operasional Variabel
Kabupaten/Kota	Kabupaten adalah pembagian wilayah administratif di Indonesia setelah provinsi, yang dipimpin oleh seorang bupati. Sedangkan Kota adalah pusat permukiman dan kegiatan penduduk yang mempunyai batas wilayah administrasi yang diatur dalam peraturan perundang-undangan serta pemukiman yang telah memperlihatkan watak dan ciri kehidupan perkotaan.
Jahe	Jahe adalah tanaman rimpang yang sangat populer sebagai rempah-rempah dan bahan obat.

Nama Variabel	Definisi Operasional Variabel
Dringo	Dringo adalah salah satu tumbuhan penyusun ramuan anti malaria dan kanker serta termasuk peringkat ke-17 sebagai tumbuhan yang paling banyak digunakan pada semua ramuan di seluruh etnis yang diteliti.
Kapulaga	Kapulaga adalah sejenis rempah yang dihasilkan dari biji beberapa tanaman dari genera <i>Elettaria</i> dan <i>Amomum</i> dalam keluarga <i>Zingiberaceae</i> (keluarga jahe-jahean).
Kejibeling	Kejibeling adalah obat herbal untuk membantu meluruhkan batu ginjal dan batu kandung kemih. Obat ini tersedia dalam bentuk pil dan kapsul.
Kencur	Kencur adalah salah satu jenis <u>empon-empon</u> /tanaman obat yang tergolong dalam suku temu-temuan (<i>Zingiberaceae</i>).
Kunyit	Kunyit adalah termasuk salah satu tanaman <u>rempah-rempah</u> dan <u>obat</u> asli dari wilayah Asia Tenggara dan tergolong dalam kelompok jahe-jahean.
Laos/Lengkuas	Laos/Lengkuas adalah jenis tumbuhan umbi-umbian yang bisa hidup di daerah dataran tinggi maupun dataran rendah. Umumnya masyarakat memanfaatkannya sebagai campuran bumbu masak dan pengobatan tradisional.
Lempuyang	Lempuyang adalah sejenis <u>rempah-rempah</u> yang berkhasiat obat.
Mengkudu	Mengkudu berasal dari daerah Asia Tenggara, tergolong dalam family <i>Rubiaceae</i> . Secara tradisional mengkudu juga sering digunakan sebagai bahan obat-obatan.

Nama Variabel	Definisi Operasional Variabel
Sambiloto	Sambiloto merupakan tumbuhan berkhasiat obat berupa terna tegak yang tingginya bisa mencapai 90 sentimeter. Asalnya diduga dari Asia tropika.
Temuireng	Temuireng merupakan tanaman temu-temuan yang memiliki rimpang berwarna gelap atau kehitaman.
Temukunci	Temukunci adalah sejenis rempah-rempah yang rimpangnya dipakai sebagai <u>bumbu</u> dalam masakan <u>Asia Tenggara</u> . Bentuk temu kunci agak berbeda dengan temu-temuan yang lain karena tumbuhnya yang vertikal ke bawah.
Temulawak	Temulawak adalah tumbuhan obat yang tergolong dalam suku temu-temuan (<i>Zingiberaceae</i>). Ia berasal dari Indonesia, khususnya Pulau Jawa, kemudian menyebar ke beberapa tempat di kawasan wilayah biogeografi <i>Malesia</i> .
Lidah Buaya	Lidah Buaya adalah spesies tumbuhan dengan daun berdaging tebal dari <i>genus Aloe</i> .
Mahkota Dewa	Pohon Mahkota Dewa (<i>Phaleria macrocarpa</i>) atau dalam bahasa melayu pohon simalakama, yang buahnya juga di sebut buah simalakama. disinilah muncul pribahasa bagaikan buah simalakama. karena buah ini dikenal sebagai buah beracun yang bisa mematikan bagi para pemakannya jika kita tidak pandai mengolahnya juga sebagai salah satu tanaman obat di Indonesia. Asalnya dari Papua/Irian Jaya.

4.4. Metode Pengumpulan Data

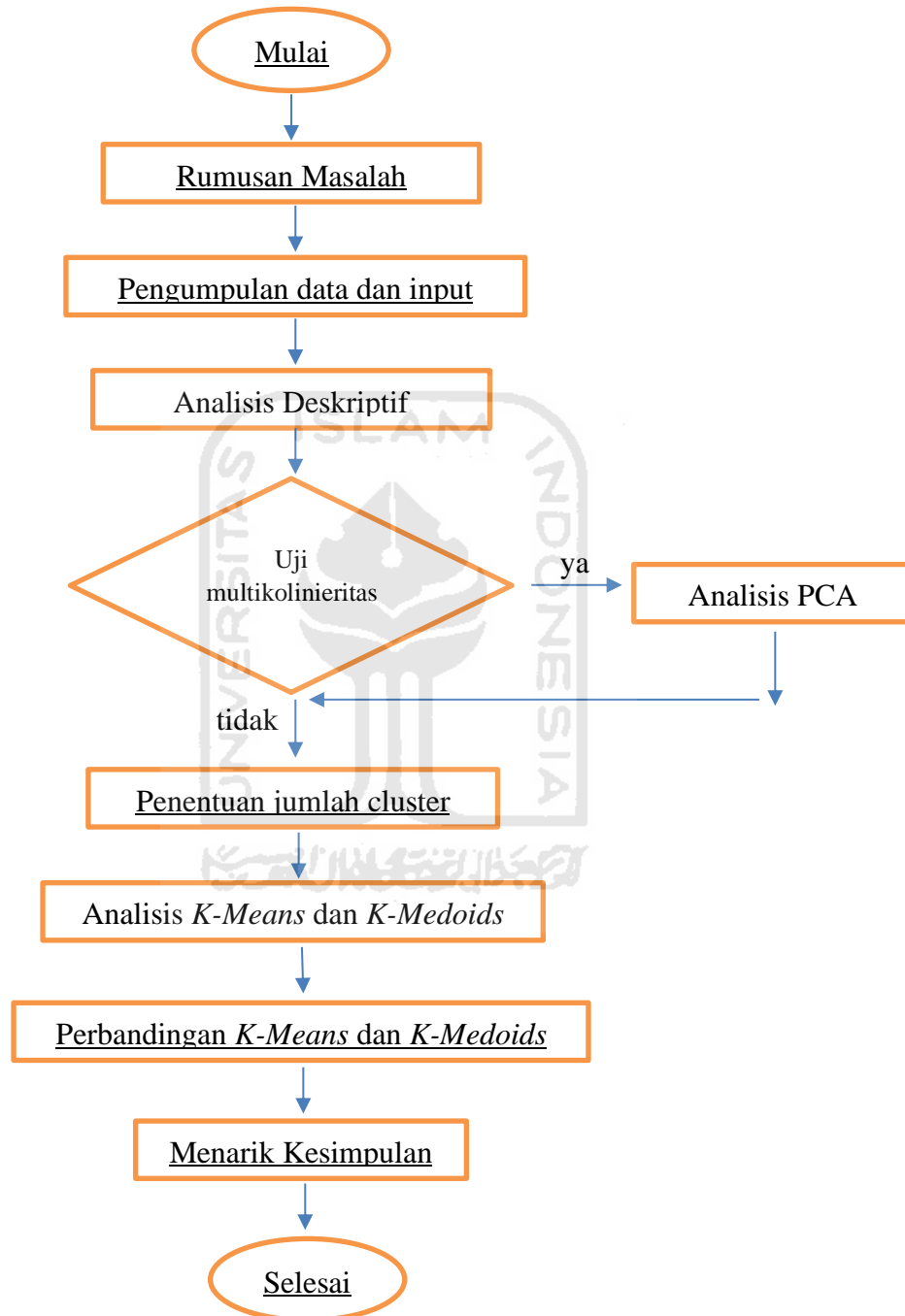
Data dalam penelitian ini menggunakan data sekunder dimana data merupakan publikasi dari Badan Pusat Statistika Provinsi Jawa Tengah. Data tersebut di kumpulkan melalui situs Website Badan Pusat Statistika Provinsi Jawa Tengah dengan cara mengumpulkan data dari setiap variabel.

4.5. Metode Analisis Data

Pada penelitian ini digunakan dua metode analisis yakni analisis deskriptif untuk mendeksripsikan data yang akan diproses secara garis besarnya, dan mengklasifikasikannya menggunakan algoritma pada metode *K-means* dan *K-medoids Clustering* dimana *K-means* merupakan merupakan salah satu algoritma dengan partitional, karena *K-means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya dan *K-medoids Clustering* merupakan varian dari metode *K-means*. Hal ini didasarkan pada penggunaan *medoids* bukan dari pengamatan *mean* yang dimiliki oleh setiap klaster, dengan tujuan mengurangi sensitivitas dari partisi sehubungan dengan nilai ekstrim yang ada dalam dataset. Proses untuk menjalankan algoritma tersebut dibantu dengan *software RStudio*.

4.6. Tahapan Penelitian

Tahapan atau langkah dalam penelitian ini digambarkan dalam *flowchart* melalui **Gambar 4.1** berikut ini:



Gambar 4.1 Flowchart Penelitian

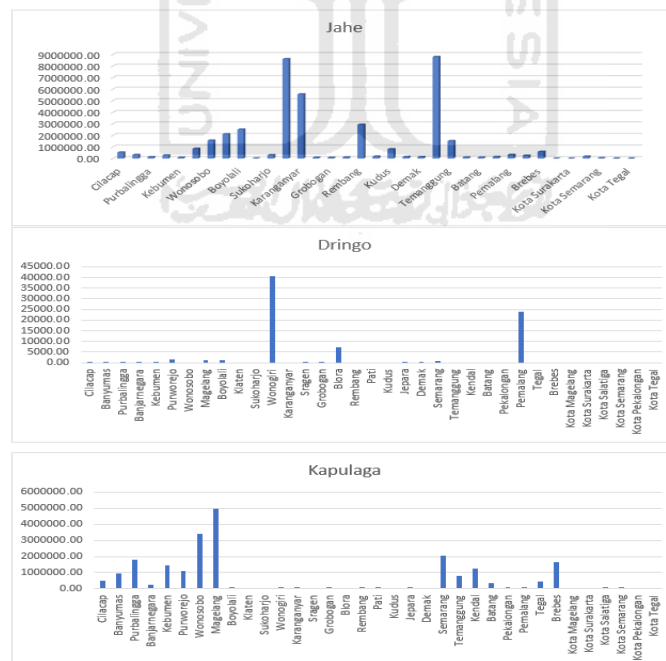
BAB V

HASIL DAN PEMBAHASAN

Pada bagian ini akan membahas tahapan-tahapan analisis serta hasil yang diperoleh untuk pengkalsteran komoditas Tanaman Biofarmaka di Provinsi Jawa Tengah tahun 2018. Analisis yang dilakukan menggunakan analisis deskriptif, metode *Cluster K-means* dan *Cluster K-medoids*.

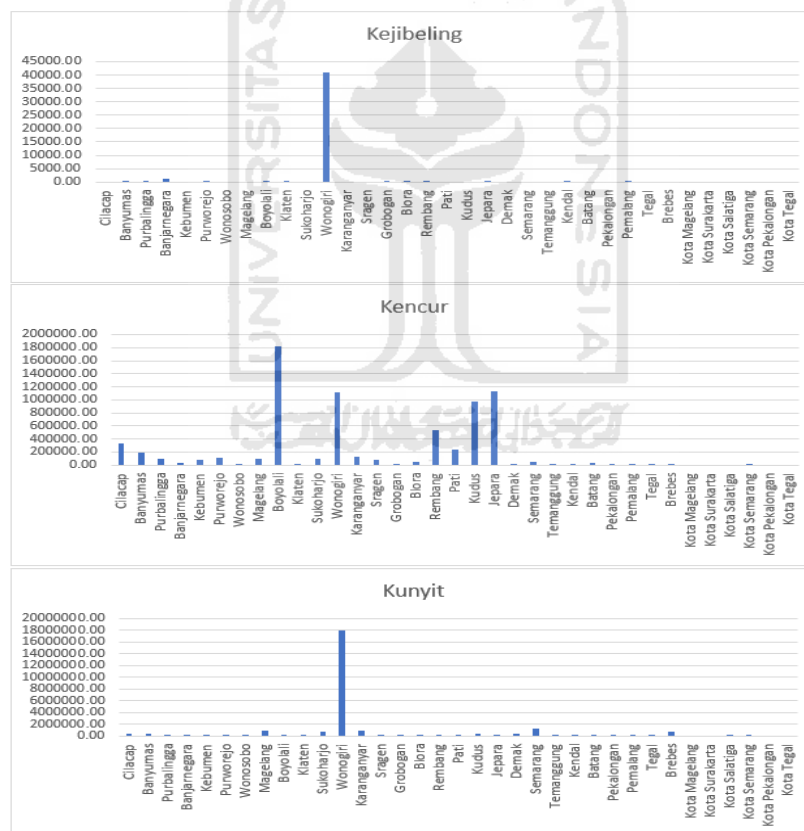
5.1. Analisis Deskriptif

Analisis deskriptif mampu menggambarkan kondisi dari petani tanaman pada setiap Kabupaten/Kota di Provinsi Jawa Tengah. Berikut merupakan grafik yang menggambarkan kondisi jumlah produksi tanaman Biofarmaka yang dihasilkan, meliputi Jahe, Dringo, Kapulaga, Kejibeling, Kencur, Kunyit, Laos/Lengkuas, Lempuyang, Mengkudu, Sambiloto, Temuireng, Temukunci, Temulawak, Lidah Buaya dan Mahkota Dewa.



Gambar 5.1 Analisis Deskriptif (Jahe, Dringo, Kapulaga)

Hasil grafik pada **Gambar 5.1** dapat dilihat bahwa produksi terbanyak tanaman Jahe di Provinsi Jawa Tengah ada di Kabupaten Semarang sebesar 8.793.952 kg dengan nilai rata-ratanya sebesar 1.119.955,80 kg dan nilai terendah ada di 4 Kota yaitu Kota Magelang, Kota Surakarta, Kota Pekalongan dan Kota Tegal sebesar 0 kg yang artinya kota tersebut tidak produksi tanaman Jahe. Produksi tanaman Dringo terbanyak di Provinsi Jawa Tengah ada di Kabupaten Wonogiri sebesar 40.624 kg dengan nilai rata-ratanya sebesar 2.244,66 kg. Selanjutnya produksi tanaman Kapulaga terbanyak terdapat di Kabupaten Magelang sebanyak 4.960.835 kg dengan nilai rata-ratanya sebesar 609.088,03 kg.



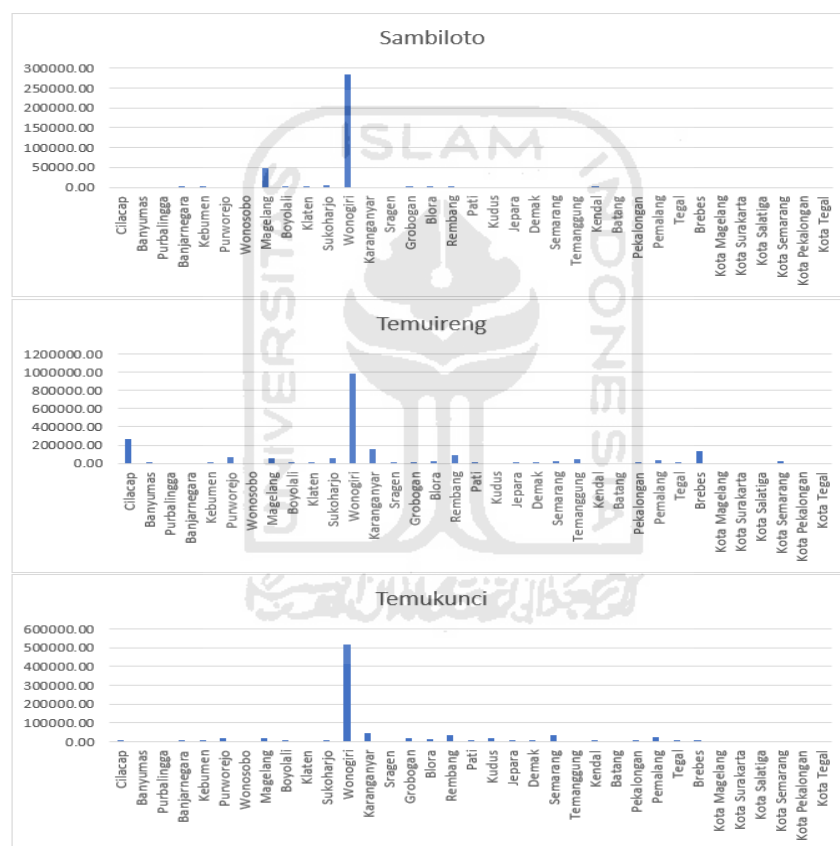
Gambar 5.2 Analisis Deskriptif (Kejibeling, Kencur, Kunyit)

Hasil grafik pada **Gambar 5.2** dapat dilihat bahwa produksi terbanyak tanaman Kejobeling di Provinsi Jawa Tengah ada di Kabupaten Wonogiri sebesar 40.994 kg dengan nilai rata-ratanya sebesar 1.269,17 kg dan mayoritas Kabupaten/Kota di Provinsi Jawa Tengah tidak produksi tanaman Kejobeling. Produksi tanaman Kencur terbanyak di Provinsi Jawa Tengah ada di Kabupaten Boyolali sebesar 1.811.480 kg dengan nilai rata-ratanya sebesar 208.522,86 kg dan yang tidak produksi tanaman Kencur ada di 5 Kota yaitu Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Pekalongan dan Kota Tegal. Selanjutnya produksi tanaman Kunyit terbanyak terdapat di Kabupaten Wonogiri sebanyak 17.935.420 kg dengan nilai rata-ratanya sebesar 735.653,31 kg dan yang tidak produksi tanaman Kunyit ada di 4 Kota yaitu Kota Magelang, Kota Surakarta, Kota Pekalongan dan Kota Tegal.



Gambar 5.3 Analisis Deskriptif (Laos, Lempuyang, Mengkudu)

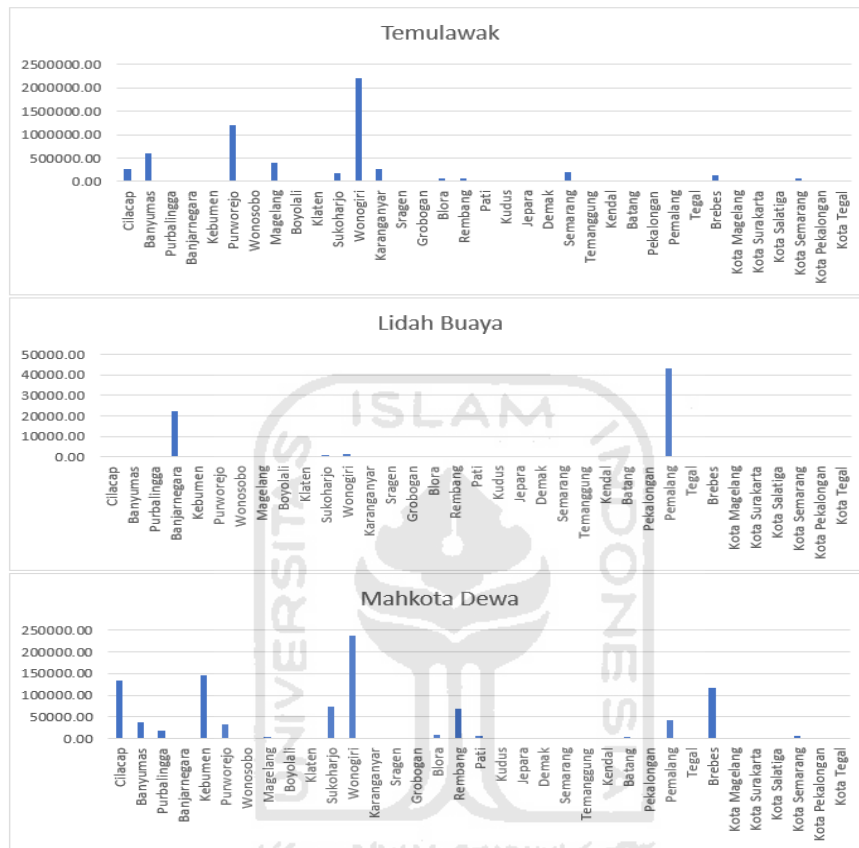
Hasil grafik pada **Gambar 5.3** dapat dilihat bahwa produksi terbanyak tanaman Laos/Lengkuas di Provinsi Jawa Tengah ada di Kabupaten Wonogiri sebesar 4.018.666 kg dan nilai rata-ratanya sebesar 536.160,86 kg. Produksi tanaman Lempuyang terbanyak di Provinsi Jawa Tengah ada di Kabupaten Wonogiri sebesar 972.336 kg dengan nilai rata-ratanya sebesar 60.125,89 kg. Selanjutnya produksi tanaman Mengkudu terbanyak juga terdapat di Kabupaten Wonogiri sebanyak 184.008 kg dengan nilai rata-ratanya sebesar 20.649,09 kg.



Gambar 5.4 Analisis Deskriptif (Sambiloto, Temuireng, Temukunci)

Hasil grafik pada **Gambar 5.4** dapat dilihat bahwa produksi terbanyak tanaman Sambiloto di Provinsi Jawa Tengah ada di Kabupaten Wonogiri sebesar 283.769 kg dengan nilai rata-ratanya sebesar 9.693,26 kg. Produksi tanaman Temuireng terbanyak di Provinsi Jawa Tengah ada di Kabupaten Wonogiri sebesar 983.513 kg dengan nilai rata-ratanya sebesar 57.404,54 kg. Selanjutnya produksi tanaman Temukunci

terbanyak juga terdapat di Kabupaten Wonogiri sebanyak 517.088 kg dengan nilai rata-ratanya sebesar 21.544,83 kg.



Gambar 5.5 Analisis Deskriptif (Temulawak, Lidah Buaya, Mahkota Dewa)

Hasil grafik pada **Gambar 5.5** dapat dilihat bahwa produksi terbanyak tanaman Temulawak di Provinsi Jawa Tengah ada di Kabupaten Wonogiri sebesar 2.196.177 kg dengan nilai rata-ratanya sebesar 167.456,66 kg. Produksi tanaman Lidah Buaya terbanyak di Provinsi Jawa Tengah ada di Kabupaten Pemalang sebesar 43.225 kg dengan nilai rata-ratanya sebesar 2.017,63 kg. Selanjutnya produksi tanaman Temukunci terbanyak terdapat di Kabupaten Wonogiri sebanyak 236.120 kg dengan nilai rata-ratanya sebesar 26.950 kg.

Dari hasil analisis deskriptif tersebut dapat disimpulkan bahwa Kabupaten/Kota di Provinsi Jawa Tengah yang menghasilkan produksi tanaman biofarmaka terbanyak terdapat di Kabupaten Wonogiri. Faktor utamanya yaitu mempunyai lahan pertanian

yang luas sehingga bisa produksi tanaman biofarmaka terbanyak. Dan untuk daerah yang menghasilkan jumlah produksi sedikit disebabkan karena luas wilayah pertanian yang semakin lama semakin sempit, penyimpangan iklim dan pengembangan komoditas lain.

5.2. Asumsi Cluster

5.2.1 Sampel Representatif

Populasi pada penelitian ini adalah seluruh Kabupaten/ Kota di Jawa Tengah. Penelitian menggunakan populasi sehingga asumsi sampel representatif sudah terpenuhi.

5.2.2 Uji Multikolinieritas

Uji multikolinieritas pada penelitian ini penulis menggunakan nilai korelasi dan dari hasil korelasi penulis hanya mengambil beberapa *output* saja. Berikut uji hipotesis dan hasil *output* nilai korelasi:

a. Hipotesis Penelitian

$$H_0 : \rho_{xy} = 0$$

$$H_1 : \rho_{xy} \neq 0$$

b. Tingkat Signifikansi

$$\alpha = 0,05$$

c. Daerah Kritis

Tolak H_0 jika : $P\text{-value} < \alpha$ atau nilai korelasi variabel $> r_{(0,05;35)}$

; dimana $r_{(0,05;35)} = 0.334$

d. Statistik Uji

Tabel 5.1 Hasil Perhitungan Korelasi

r_x	r_y	r_{xy}	Keputusan
Jahe	Dringo	0,4805392	Tolak H_0
Dringo	Kencur	0,29271704	Gagal Tolak H_0
Kencur	Kunyit	0,37645537	Tolak H_0
Lidah Buaya	Sambiloto	-0,01917565	Gagal Tolak H_0

Mengkudu	Temulawak	0,82099157	Tolak H ₀
----------	-----------	------------	----------------------

e. Kesimpulan

Dengan tingkat signifikansi sebesar 5% data yang ada menjelaskan bahwa dari hasil statistik uji terdapat keputusan Tolak H₀ (terdapat multikolinieritas) dan Gagal Tolak H₀ (tidak ada multikolinieritas).

Dengan menggunakan uji multikolinieritas seperti diatas dapat diartikan bahwa masih terdapat beberapa variabel yang terdapat multikolinieritas dan ada beberapa variabel lain yang tidak terdapat multikolinieritas. Sehingga dapat disimpulkan bahwa asumsi multikolinieritas pada penelitian ini tidak terpenuhi. Selanjutnya dilakukan penanganan untuk melakukan reduksi factor terhadap variabel-variabel yang diteliti dengan menggunakan analisis komponen utama atau biasa disebut PCA.

5.3. Analisis Komponen Utama

Untuk melakukan analisis komponen utama sebelumnya perlu dilakukan uji *Bartlett* dan KMO.

1. Uji *Bartlett*

Uji *Bartlett* digunakan untuk menguji hipotesis apakah antar variabel saling berkorelasi atau tidak berkorelasi. Berikut hasil uji *Bartlett*:

i. Hipotesis

H₀ : $\rho = 0$ (Tidak terdapat Korelasi)

H₁ : $\rho \neq 0$ (Terdapat Korelasi)

ii. Tingkat Signifikansi

$\alpha = 0,05$

iii. Daerah Kritis

Tolak H₀ jika *P-value* < α

iv. Statistik Uji

Nilai *P-value* yaitu 2,2e-16

v. Keputusan

Tolak H₀ karena *P-value* (2,2e-16) < α (0,05)

vi. Kesimpulan

Dengan tingkat signifikasni 5% data yang ada menjelaskan bahwa terdapat korelasi antar variabel.

2. KMO (*Kaiser Mayer Olkin*)

KMO digunakan untuk meneliti ketepatan analisis faktor, berdasarkan hasil *output* pada *Rstudio* yang akan dilampirkan pada lampiran diketahui bahwa nilai KMO yaitu sebesar 0,7485761. Nilai ini berada di antara 0,5 hingga 1 yang berarti bahwa analisis faktor tepat digunakan atau layak digunakan.

Untuk menentukan jumlah faktor yang terbentuk, peneliti menggunakan nilai *eigenvalues*. Faktor dengan nilai *eigenvalues* lebih dari satu akan dipertahankan, sedangkan apabila kurang dari satu maka faktor tidak diikutsertakan dalam model. Selain itu, menurut Supranto (2004) ekstraksi faktor dihentikan apabila persentase kumulatif varian paling sedikit 60% atau 75% dari seluruh variabel asli.

Tabel 5.2 Nilai *Eigenvalues*

Faktor	<i>Initial Eigenvalues</i>		
	<i>Standar Deviation</i>	<i>Proportion of Variance</i>	<i>Cumulative Proportion</i>
1	3,0909937	0,6369495	0,6369495
2	1,18801673	0,09409225	0,73104174
3	1,10067455	0,08076563	0,81180737
4	0,90586717	0,05470635	0,86651372
5	0,83461736	0,04643908	0,91295280
6	0,65360019	0,02847955	0,94143234
7	0,60258934	0,02420759	0,96563994
8	0,52995033	0,01872316	0,98436310
9	0,319511928	0,006805858	0,991168953
10	0,25904555	0,00447364	0,99564259

11	0,203675278	0,002765575	0,998408167
12	0,1132330900	0,0008547822	0,9992629496
13	0,0786977779	0,0004128893	0,9996758390
14	0,059240948	0,000233966	0,999909805
15	3,678213e-02	9,019501e-05	1,00000000

Berdasarkan **Tabel 5.2** dapat dilihat bahwa faktor 1 sampai faktor 3 yang memiliki nilai *Eigenvalues* lebih dari satu. Sehingga faktor yang terbentuk pada analisis komponen utama ini sebanyak 3 faktor. Faktor 1 menyumbang varian sebesar 63,7 %, faktor 2 menyumbang varian sebesar 9,4% dan faktor 3 menyumbang varian sebesar 8,07 sehingga kumulatif varian yang dapat digambarkan oleh ketiga faktor tersebut adalah sebesar 81,2 % yang lebih dari 60% yang artinya ketiga faktor tersebut dapat menggambarkan data.

5.4. Penerapan Cluster

5.4.1. Analisis Cluster K-means

Cluster K-means merupakan salah satu metode yang digunakan oleh penulis. Metode *K-means* mengelompokkan data didasarkan pada jarak yang terdekat dengan pusatnya. Pada penelitian ini menggunakan jarak Euclidian, metode *K-means* tergantung pada inisiasi nilai pusat awal *cluster* yang diberikan (Santosa, 2007). Dari hasil *cluster* menggunakan *Software RStudio* dengan 3 *cluster* sehingga terbentuk anggota *cluster*. berikut hasil *output RStudio* dan anggota *cluster*:

```
Clustering vector:
 [1] 2 2 2 2 2 2 2 2 2 2 3 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
 [35] 2
```

Gambar 5.6 Hasil *Cluster K-means*

Pada **Gambar 5.6** dari 3 *cluster* dihasilkan pada *cluster* 1 terdapat 2 anggota, *cluster* 2 terdapat 32 anggota dan *cluster* 3 terdapat 1 anggota. Berikut tabel hasil *cluster* beserta anggota-anggotanya:

Tabel 5.3 Anggota *Cluster K-means*

Cluster	Jumlah Anggota	Kabupaten/Kota
1	2	Karanganyar dan Semarang
2	32	Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan dan Kota Tegal
3	1	Wonogiri

Pada **Tabel 5.3** dapat diketahui bahwa hasil pengelompokan menggunakan metode *K-means* yaitu pada *Cluster 1* terdiri dari 2 Kabupaten, *Cluster 2* terdiri dari 32 Kabupaten/Kota dan *Cluster 3* terdiri dari 1 Kabupaten.

Kemudian untuk membedakan hasil pengelompokan yang terbentuk maka dihitung nilai rata-rata setiap variabelnya. Berikut hasil rata-rata setiap variabelnya:

Tabel 5.4 Nilai Rata-rata variabel *cluster K-means*

Variabel	Cluster 1	Cluster 2	Cluster 3
Jahe	7.174.372,0	507.657,9	8.604.656,0
Dringo	362,5	1.162,938	40.624,0
Kapulaga	1.029.997,5	601.258,4	17.816,0
Kejibeling	0,0000	107,0938	40.994,00
Kencur	89.052,5	187.668,3	1.115.858,0
Kunyit	1.097.248,5	175.560,9	17.935.420,0
Laos/Lengkuas	1.733.663,0	352.488,7	4.018.666,0

Lempuyang	50.147,50	32.242,97	972.336,00
Mengkudu	637,50	16.794,84	184.008,00
Sambiloto	0,000	1.734,219	283.769,00
Temuireng	90.337,00	26.405,38	983.513,0
Temukunci	41.006,50	4.842,75	517.088,0
Temulawak	233.186,00	99.951,06	2.196.177,00
Lidah Buaya	0,000	2.163,062	1.399,0
Mahkota Dewa	0,000	22.098,38	23.6120,0
Rata-Rata Total	769.334,04	135.475,8	2.476.562,9

Setelah *cluster* terbentuk maka tahap selanjutnya adalah memberi ciri spesifik untuk menggambarkan isi *cluster* tersebut dan berdasarkan **tabel 5.4** rata-rata jumlah tanaman biofarmaka di Kabupaten/Kota Provinsi Jawa Tengah, maka diperoleh sebagai berikut:

I. *Cluster* 1

Cluster 1 beranggotakan 2 kabupaten dimana *cluster* pertama memiliki rata-rata total variabel yang lebih tinggi dari *cluster* 2 dan lebih rendah dari *cluster* 3 yaitu 769334.04 dan dapat dikelompokkan menjadi kelompok kabupaten/kota yang jumlah produksi tanaman biofarmakanya kategori menengah. Dan juga terdapat produksi tanaman Kapulaga tertinggi dibandingkan dari anggota *cluster* lainnya, akan tetapi di *cluster* 1 terdapat anggota Kabupaten/Kota yang tidak produksi tanaman biofarmaka yaitu Kejibeling, Sambiloto, Lidah Buaya dan Mahkota Dewa. Maka *cluster* 1 ini dikategorikan *cluster* sedang.

II. *Cluster* 2

Cluster 2 beranggotakan 32 kabupaten/kota dimana *cluster* kedua memiliki rata-rata total variabel yang berada paling rendah dari ketiga *cluster* yaitu 135475.8 dan dapat dikelompokkan menjadi kelompok kabupaten/kota yang jumlah produksi tanaman biofarmakanya kategori sedikit. Pada *Cluster* 2 anggota kabupaten/kota

memproduksi semua tanaman biofarmaka tetapi hasil jumlah rata-rata produksinya masih tergolong sedikit. Maka *cluster 2* ini dikategorikan *cluster* rendah.

III. *Cluster 3*

Cluster 3 beranggotakan 1 kabupaten dimana *cluster* ketiga memiliki rata-rata total variabel yang berada paling tinggi dari ketiga *cluster* yaitu 2.476.562,9 dan dapat dikelompokkan menjadi kelompok kabupaten/kota yang jumlah produksi tanaman biofarmakanya kategori banyak. Pada *Cluster 3* anggota kabupaten memproduksi semua tanaman biofarmaka dengan hasil rata-rata jumlah produksinya yang tinggi. Maka *cluster 3* ini dikategorikan *cluster* tinggi.

5.4.2. Rasio Simpangan Baku dalam kelompok (S_w) dan Simpangan Baku antar kelompok (S_b) dalam *K-means*

Dari proses pengelompokan menggunakan metode *Cluster K-means* yang telah dijelaskan sebelumnya dengan jumlah *clusternya* yaitu 3 *cluster*. Kemudian selanjutnya akan dilampirkan hasil rasio simpangan baku dalam kelompok (s_w) dan simpangan baku antar kelompok (s_b). Berikut hasil rasio dari S_w dan S_b :

Tabel 5.5 Rasio S_w dan S_b *K-means*

<i>K-means</i>	Nilai
S_w	403.620,1
S_b	1.210.860,3
S_w/S_b	0,333333

Selanjutnya dari nilai rasio minimum S_w terhadap S_b , dalam hal ini metode yang baik dilihat dari nilai rasio yang minimum (Barakbah dan Arai.2007). Nilai rasio S_w/S_b dalam *Cluster K-means* sebesar 0,333333.

5.4.3. Analisis *Cluster K-medoids*

Pada metode *Cluster K-medoids* ini sama halnya dengan metode *Cluster K-means*, peneliti menggunakan 3 *cluster* untuk jumlah *clusternya* dan pengolahan analisis juga menggunakan bantuan program *Rstudio*. Disini hasil *output* dari *Cluster*

K-medoids sama dengan hasil dari *Cluster K-means*. Berikut hasil *output* dengan menggunakan program *Rstudio*:

```
Clustering vector:
 [1] 2 2 2 2 2 2 2 2 2 2 2 2 3 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2
[35] 2
```

Gambar 5.7 Hasil *Cluster K-medoids*

Pada **Gambar 5.7** dapat dilihat bahwa dari 3 *cluster* dihasilkan pada *cluster 1* terdapat 2 anggota, *cluster 2* terdapat 32 anggota dan *cluster 3* terdapat 1 anggota. Berikut tabel hasil *clusternya* beserta anggota-anggotanya :

Tabel 5.6 Anggota *Cluster K-medoids*

Cluster	Jumlah Anggota	Kabupaten/Kota
1	2	Karanganyar dan Semarang
2	32	Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan dan Kota Tegal
3	1	Wonogiri

Pada **Tabel 5.6** dapat diketahui bahwa hasil pengelompokan menggunakan metode *K-medoids* sama dengan hasil dari metode *K-means* yaitu pada *Cluster 1* terdiri dari 2 Kabupaten, *Cluster 2* terdiri dari 32 Kabupaten/Kota dan *Cluster 3* terdiri dari 1 Kabupaten.

Kemudian untuk membedakan hasil pengelompokan yang terbentuk maka dihitung nilai rata-rata setiap variabelnya. Berikut hasil rata-rata setiap variabelnya:

Tabel 5.7 Nilai Rata-rata variabel *cluster K-medoids*

Variabel	Cluster 1	Cluster 2	Cluster 3
Jahe	8.793.952	244.523	8.604.656
Dringo	725	0	40.624
Kapulaga	2.042.795	435.678	17.816
Kejibeling	0	0	40.994
Kencur	55.400	6.245	1.115.858
Kunyit	1.241.534	123.036	17.935.420
Laos/Lengkuas	467.262	93.057	4.018.666
Lempuyang	20.000	586	972.336
Mengkudu	0	510	184.008
Sambiloto	0	0	283.769
Temuireng	21.790	1.210	983.513
Temukunci	34.613	405	517.088
Temulawak	208.337	28.670	2.196.177
Lidah Buaya	0	0	1.399
Mahkota Dewa	0	0	236.120
Rata-Rata Total	859.093,87	62.261,33	2.476.562,93

Setelah *cluster* terbentuk maka tahap selanjutnya adalah memberi ciri spesifik untuk menggambarkan isi *cluster* tersebut dan berdasarkan **tabel 5.7** rata-rata jumlah tanaman biofarmaka di Kabupaten/Kota Provinsi Jawa Tengah, maka diperoleh sebagai berikut:

I. *Cluster* 1

Cluster 1 beranggotakan 2 kabupaten dimana *cluster* pertama memiliki rata-rata total variabel yang lebih tinggi dari *cluster* 2 dan lebih rendah dari *cluster* 3 yaitu 859.093,87 dan dapat dikelompokkan menjadi kelompok kabupaten/kota yang jumlah produksi tanaman biofarmakanya kategori menengah. Maka *cluster* 1 ini dikategorikan *cluster* sedang.

II. *Cluster* 2

Cluster 2 beranggotakan 32 kabupaten/kota dimana *cluster* kedua memiliki rata-rata total variabel yang berada paling rendah dari ketiga *cluster* yaitu 62.261,33 dan dapat dikelompokkan menjadi kelompok kabupaten/kota yang jumlah produksi

tanaman biofarmakanya kategori sedikit. Maka *cluster* 2 ini dikategorikan *cluster* rendah.

III. *Cluster* 3

Cluster 3 beranggotakan 1 kabupaten dimana *cluster* ketiga memiliki rata-rata total variabel yang berada paling tinggi dari ketiga *cluster* yaitu 2.476.562,93 dan dapat dikelompokkan menjadi kelompok kabupaten/kota yang jumlah produksi tanaman biofarmakanya kategori banyak. Pada *Cluster* 3 anggota kabupaten memproduksi semua tanaman biofarmaka dengan hasil rata-rata jumlah produksinya yang tinggi. Maka *cluster* 3 ini dikategorikan *cluster* tinggi.

5.4.4. Rasio Simpangan Baku dalam kelompok (S_w) dan Simpangan Baku antar kelompok (S_b) dalam *K-medoids*

Dari proses pengelompokan menggunakan metode *Cluster K-medoids* yang telah dijelaskan sebelumnya dengan jumlah *clusternya* yaitu 3 *cluster*. Kemudian selanjutnya akan dilampirkan hasil rasio simpangan baku dalam kelompok (s_w) dan simpangan baku antar kelompok (s_b). Berikut hasil rasio dari S_w dan S_b :

Tabel 5.8 Rasio S_w dan S_b *K-medoids*

<i>K-medoids</i>	Nilai
S_w	410.058,72
S_b	1.230.176,2
S_w/S_b	0,333333

Selanjutnya dari nilai rasio minimum S_w terhadap S_b , dalam hal ini metode yang baik dilihat dari nilai rasio yang minimum (Barakbah dan Arai, 2007). Nilai rasio S_w/S_b dalam *Cluster K-medoids* sebesar 0,33333.

5.5. Perbandingan Metode Terbaik

Perbandingan metode terbaik penulis menggunakan hasil dari nilai Rasio simpangan baku, sebelumnya penulis membandingkan kedua metode dari hasil pembagian anggota *cluster* dalam bentuk tabel. Berikut tabel perbandingannya:

Tabel 5.9 Perbandingan *cluster K-means* dan *K-medoids*

Cluster	Kabupaten/Kota (<i>K-means</i>)	Kabupaten/Kota (<i>K-medoids</i>)
1	Karanganyar dan Semarang	Karanganyar dan Semarang
2	Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan dan Kota Tegal	Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan dan Kota Tegal
3	Wonogiri	Wonogiri

Pada **Tabel 5.9** dapat dilihat bahwa tidak ada perbedaan perbandingan antara *cluster K-means* dan *K-medoids* karena hasil *cluster* beserta anggota *clusternya* sama antara kedua metode yaitu pada metode *K-means cluster* 1 ada 2 Kabupaten begitu sebaliknya pada metode *K-medoids* juga ada 2 Kabupaten. Pada *cluster* 2 metode *K-means* ada 32 Kabupaten/Kota dan pada metode *K-medoids* juga ada 32 Kabupaten/Kota. Dan pada *cluster* 3 metode *K-means* ada 1 kabupaten, pada metode *K-medoids* juga ada 1 kabupaten.

Selanjutnya peneliti membandingkan kedua metode *K-means* dan *K-medoids* berdasarkan nilai rasio simpangan baku (S_w/S_b).

Tabel 5.10 Perbandingan Nilai Rasio Simpangan Baku

<i>Rasio</i>	<i>K-means</i>	<i>K-medoids</i>
S_w	403.620,1	410.058,72
S_b	1.210.860,3	1.230.176,2
S_w/S_b	0,333333	0,333333

Berdasarkan **Tabel 5.10**, nilai rasio simpangan baku (S_w/S_b) menunjukkan bahwa metode *K-means* dan metode *K-medoids* memiliki nilai perbandingan rasio S_w/S_b yang sama yaitu 0,333333. Artinya, pada penelitian ini dengan data tanaman biofarmaka di Provinsi Jawa Tengah tahun 2018 dapat dilakukan dengan metode *cluster K-means* dan metode *cluster K-medoids*, karena kedua metode tersebut sama-sama kuat untuk melakukan analisis pada penelitian ini.

5.6. Perbedaan Penelitian ini dengan Penelitian Wicaksono (2019)

Terdapat perbedaan antara penelitian ini dengan penelitian yang dilakukan oleh Wicaksono (2019) yaitu pada metode penelitian, tahun penelitian dan variabel penelitian yang dilakukan. Penelitian ini menggunakan metode perbandingan antara *Cluster K-Means* dan *K-Medoids*. Kemudian pada penelitian Wicaksono (2019) hanya menggunakan metode *Cluster K-Means*. Penelitian ini menggunakan data tanaman biofarmaka pada tahun 2018, sedangkan pada penelitian Wicaksono (2019) menggunakan data tanaman biofarmaka tahun 2015-2017. Selanjutnya, variabel yang digunakan pada penelitian ini menggunakan 15 variabel jumlah produksi tanaman biofarmaka, sedangkan penelitian wicaksono hanya menggunakan 4 variabel jumlah produksi tanaman biofarmaka. Untuk hasilnya juga terdapat perbedaan, karena pada penelitian ini menggunakan jumlah *clusternya* sebanyak 3, sedangkan pada penelitian Wicaksono (2019) menggunakan jumlah *cluster* sebanyak 4. Tetapi untuk daerah yang ada pada *cluster* tertinggi sama yaitu Kabupaten Wonogiri.

BAB VI

KESIMPULAN DAN SARAN

6.1. Kesimpulan

Berdasarkan dari hasil analisis dan pembahasan, peneliti mengambil kesimpulan sebagai berikut.

1. Produksi terbanyak tanaman Jahe ada di Kabupaten Semarang sebesar 8.793.952 kg, produksi tanaman Dringo terbanyak ada di Kabupaten Wonogiri sebesar 40.624 kg, produksi tanaman Kapulaga terbanyak terdapat di Kabupaten Magelang sebanyak 4.960.835 kg, produksi terbanyak tanaman Kejibeling ada di Kabupaten Wonogiri sebesar 40.994 kg, produksi tanaman Kencur terbanyak ada di Kabupaten Boyolali sebesar 1.811.480 kg, produksi tanaman Kunyit terbanyak terdapat di Kabupaten Wonogiri sebanyak 17.935.420 kg, produksi terbanyak tanaman Laos/Lengkuas ada di Kabupaten Wonogiri sebesar 4.018.666 kg, produksi tanaman Lempuyang terbanyak ada di Kabupaten Wonogiri sebesar 972.336 kg, produksi tanaman Mengkudu terbanyak juga terdapat di Kabupaten Wonogiri sebanyak 184.008 kg, produksi terbanyak tanaman Sambiloto ada di Kabupaten Wonogiri sebesar 283.769 kg, produksi tanaman Temuireng terbanyak ada di Kabupaten Wonogiri sebesar 983.513 kg, produksi tanaman Temukunci terbanyak juga terdapat di Kabupaten Wonogiri sebanyak 517.088 kg, produksi terbanyak tanaman Temulawak ada di Kabupaten Wonogiri sebesar 2.196.177 kg, produksi tanaman Lidah Buaya terbanyak ada di Kabupaten Pemalang sebesar 43.225 kg, produksi tanaman Temukunci terbanyak terdapat di Kabupaten Wonogiri sebanyak 236.120 kg dengan nilai rata-ratanya sebesar 26.950 kg. Dapat disimpulkan bahwa Kabupaten/Kota di Provinsi Jawa Tengah yang menghasilkan produksi tanaman Biofarmaka terbanyak terdapat di Kabupaten Wonogiri.

2. Hasil *cluster* yang terbentuk dengan metode *Cluster K-means* dan metode *Cluster K-medoids* adalah sama. Diperoleh hasil *clusternya* yaitu *cluster* 1 ada 2 kabupaten, *cluster* 2 ada 32 kabupaten/kota dan *cluster* 3 ada 1 kabupaten. Dari *cluster* yang terbentuk diperoleh urutan kelompok kabupaten/kota dengan jumlah produksi yang paling banyak hingga kelompok kabupaten/kota yang jumlah produksi paling sedikit berturut-turut adalah *cluster* 3, *cluster* 1, *cluster* 2. *Cluster* 3 adalah kategori *cluster* tinggi yang artinya kelompok kabupaten yang berada di *cluster* 3 memproduksi semua tanaman biofarmaka dengan hasil rata-rata jumlah produksinya yang tinggi dengan nilai rata-ratanya yaitu 2.476.562,9. Daerah yang berada di *cluster* 3 adalah Kabupaten Wonogiri.
3. Jika dilihat dari nilai rasio simpangan S_w terhadap S_b , menunjukkan bahwa nilai rasio simpangan baku (S_w/S_b) pada metode *Cluster K-means* dan metode *Cluster K-medoids* yaitu 0,333333. Artinya, pada penelitian ini dengan data tanaman biofarmaka di Provinsi Jawa Tengah tahun 2018 dapat dilakukan dengan metode *cluster K-means* dan metode *cluster K-medoids*, karena kedua metode tersebut sama-sama kuat untuk melakukan analisis pada penelitian ini.

6.2. Saran

Berdasarkan hasil penelitian, terdapat saran-saran yang diajukan yaitu sebagai berikut.

1. berdasarkan hasil *cluster* yang terbentuk pemerintah Jawa Tengah dapat mengetahui kabupaten dan kota yang memiliki potensi terbaik dalam menghasilkan produksi tanaman biofarmaka.
2. Bagi pengambil kebijakan informasi ini dapat bermanfaat dalam membuat program untuk meningkatkan potensi tanaman biofarmaka di kabupaten/kota di Jawa Tengah.
3. Penelitian ini dapat dilanjutkan dalam beberapa tahun ke depan agar dapat menambah jumlah data yang digunakan sehingga memperoleh hasil yang lebih baik dan lebih akurat.

DAFTAR PUSTAKA

- Aryuni, Mediana., Madyatmadja, E. Didik., & Miranda, Eka. 2018. *Penerapan K-Means dan K-Medoids Clustering Pada Data Internet Banking Di Bank XYZ*. Retrieved from <https://docplayer.info/85027360-Penerapan-k-means-dan-k-medoids-clustering-pada-data-internet-banking-di-bank-xyz.html>. Diakses pada tanggal 20 April 2020.
- Barakbah, Ali & Arai, Kohei. 2004. *Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering*. Retrieved from http://ridho.lecturer.pens.ac.id/papers/Barakbah_IES_2004.pdf. Diakses pada tanggal 7 Juni 2020 jam 17.00 WIB.
- Flowrensia, Yanne. 2010. *Perbandingan Penggerombolan K-Means dan K-Medoids pada Data Yang Mengandung Pencilan*. Retrieved from <https://repository.ipb.ac.id/handle/123456789/62185>. Diakses pada tanggal 20 April 2020.
- Hair, J.F.Jr., R.E. Anderson, R.L. Thatham dan Black, W.C. 2010. *Multivariat Data Analysis Seventh Edition*. New Jersey: Pearson Education, Inc.
- Hilda, Yunita. 2015. *Perbandingan K-Means dan K-Medoids Clustering Terhadap Kelayakan Puskesmas di DIY Tahun 2015*. Retrieved from <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/47>. Diakses pada tanggal 20 April 2020.
- Irwanto, et. al (2012). *Optimasi Kinerja Algoritma Klasterisasi K-Means untuk kuantisasi Warna Citra*. Jurnal Teknik ITS. Vol 1 No. 1: 197-202.
- Kamila, Insanul., Khairunnisa, Ulya., & Mustakim. 2019. *Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau*. Retrieved from <http://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/7381>. Diakses pada tanggal 20 April 2020.

- Machfudhoh, S., Wahyuningsih, N. 2013. *Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur*. Jurnal Sains dan Seni Ponmits. Vol. 2 No. 1: 1 – 8.
- Madhulatha, T.S., 2012. *An Overview On Clustering Methods*. IOSR Journal of Engineering. Vol 2 No.4: 719-725.
- Pratiwi, Nurkholida. 2016. *Implementasi K-Means dan K-Medoids Clustering dalam Pengelompokan Unit Usaha Koperasi (Studi Kasus: Unit Usaha Koperasi Terdaftar di Kabupaten Sleman per Tahun Buku 2014)*. Skripsi. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
- Rivani, Edmira. 2010. *Aplikasi K-Means Cluster Untuk Pengelompokan Provinsi Berdasarkan Produksi Padi, Jagung, Kedelai, dan Kacang hijau Tahun 2009*. Jurnal Matematika Statistika.
- Ramadhan, Aditya., Efendi, Zuliar., & Mustakim. 2017. *Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling*. Retrieved from <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/3268>. Diakses pada tanggal 20 April 2020.
- Sarwono, Jonathan. 2007. *Teori Analisis Multivariat*. Diakses dari <http://www.jonathansarwono.info/mvariat/multivariat.htm>. Pada tanggal 7 Juni 2020, jam 16.00 WIB.
- Setyawati, Astri Widiastuti. 2017. *Implementasi Algoritma Partitioning Around Medoid (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY Berdasarkan Nilai Daya Serap Ujian Nasional*. Skripsi. Yogyakarta: Fakultas Sains dan Teknologi Universitas Sanata Dharma.
- Sitepu, R., Irmeilyana, dan Gultom, B. 2011. *Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan*. Jurnal Penelitian Sains. Vol. 14 No. 3A Juli 2011, 11 – 17.

- Supranto, J. (2004). *Analisis Multivariat Arti dan Interpretasi*. Jakarta: Rineka Cipta.
- Triyanto, Wiwit Agus. 2015. *Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk*. Jurnal SIMETRIS. Vol. 6 No. 1 April 2015, 183 – 188.
- Wicaksono, Dhimas Satria. 2019. *Analisis K-Means Cluster Perbandingan 3 Tahun Produksi Tanaman Biofarmaka di Provinsi Jawa Tengah Tahun 2015-2017*. Retrieved from <https://medium.com/@16611093/analisis-k-means-cluster-perbandingan-3-tahun-produksi-tanaman-biofarmaka-di-provinsi-jawa-tengah-8f728c9dd6c9>. Diakses pada tanggal 20 April 2020.



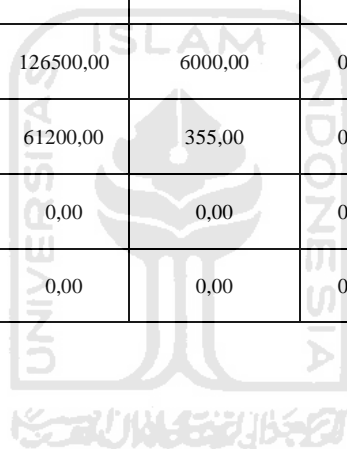
LAMPIRAN

Lampiran 1. Data Penelitian

Kab /Kota	Jahe	Dringo	Kapulaga	Kejibeling	Kencur	Kunyit	Laos/Lengkuas	Lempuyang	Mengkudu	Sam bilot o	Tem uireng	Temukun ci	Temula wak	Lida h Buaya	Mahko ta Dewa
Cilaca p	528495,00	103,00	506673,00	0,00	329894,00	455435,00	315749,00	235366,00	21900,00	0,00	271947,00	670,00	276275,00	107,00	133957,00
Banyu mas	316620,00	42,00	949851,00	32,00	184627,00	441501,00	474056,00	5995,00	38677,00	0,00	3305,00	0,00	610247,00	345,00	36815,00
Purbali ngga	119525,00	17,00	1818691,00	9,00	93810,00	49395,00	86105,00	124,00	9386,00	0,00	0,00	0,00	437,00	3,00	19456,00
Banjar negara	265855,00	50,00	265801,00	1170,00	40859,00	74877,00	550253,00	25943,00	43771,00	1170,00	0,00	90,00	2708,00	22320,00	0,00
Kebum en	80754,00	420,00	1459092,00	0,00	87982,00	98163,00	52344,00	7418,00	23095,00	173,00	9836,00	1501,00	21871,00	0,00	146884,00
Purwor ejo	846187,00	1670,00	1078730,00	600,00	109500,00	232205,00	142194,00	88612,00	78713,00	0,00	66005,00	16770,00	1197833,00	0,00	33384,00
Wonos obo	1553079,00	0,00	3423780,00	0,00	503,00	24126,00	0,00	0,00	0,00	0,00	0,00	0,00	455,00	0,00	0,00
Magela ng	2104683,00	1283,00	4960835,00	0,00	95000,00	870200,00	441000,00	113400,00	15142,00	4800,00	5690,00	16000,00	408000,00	0,00	3465,00
Boyola li	2511300,00	1200,00	3200,00	308,00	1811480,00	4870,00	8710,00	2150,00	17707,00	144,00	1150,00	550,00	11895,00	64,00	0,00
Klaten	12221,00	0,00	0,00	4,00	300,00	920,00	1361,00	2695,00	167,00	6,00	3600,00	0,00	4100,00	201,00	888,00
Sukoha rjo	288360,00	0,00	0,00	0,00	100240,00	662394,00	135740,00	170134,00	7850,00	4950,00	56071,00	1380,00	181430,00	1200,00	73628,00
Wonog iri	8604656,00	40624,00	17816,00	40994,00	1115858,00	17935420,00	4018666,00	972336,00	184008,00	283769,00	983513,00	517088,00	2196177,00	1399,00	236120,00

Karang anyar	5554792,00	0,00	17200,00	0,00	122705,00	952963,00	3000064,00	80295,00	1275,00	0,00	158884,00	47400,00	258035,00	0,00	0,00
Sragen	79507,00	500,00	0,00	0,00	85379,00	43938,00	48046,00	6850,00	496,00	0,00	8431,00	0,00	21223,00	0,00	455,00
Grobo gan	68676,00	139,00	2211,00	419,00	18960,00	54150,00	56515,00	20402,00	13102,00	416,00	5832,00	17066,00	13356,00	252,00	1887,00
Blora	97416,00	7406,00	0,00	549,00	49107,00	82144,00	117604,00	90961,00	6443,00	614,00	26149,00	13233,00	60946,00	0,00	9369,00
Remba ng	2925152,00	0,00	9,00	80,00	538814,00	179882,00	576987,00	44751,00	15038,00	10,00	88107,00	35772,00	69222,00	654,00	68471,00
Pati	176709,00	0,00	71649,00	0,00	239784,00	129689,00	14709,00	1275,00	15788,00	0,00	4550,00	4190,00	10834,00	0,00	5778,00
Kudus	804410,00	0,00	0,00	0,00	966780,00	465612,00	2864003,00	0,00	13400,00	0,00	0,00	16330,00	0,00	0,00	0,00
Jepara	140520,00	24,00	626,00	39,00	1128520,00	5410,00	729351,00	2540,00	22293,00	0,00	1839,00	2304,00	1507,00	92,00	1370,00
Demak	131015,00	375,00	0,00	0,00	8944,00	342591,00	50602,00	4186,00	32175,00	0,00	3368,00	3391,00	7295,00	339,00	1222,00
Semara ng	8793952,00	725,00	2042795,00	0,00	55400,00	1241534,00	467262,00	20000,00	0,00	0,00	21790,00	34613,00	208337,00	0,00	0,00
Teman ggung	1510850,00	0,00	795270,00	0,00	12700,00	55000,00	217474,00	73200,00	0,00	0,00	40450,00	0,00	27500,00	0,00	0,00
Kendal	99685,00	0,00	1260921,00	20,00	1586,00	101504,00	12648,00	7954,00	685,00	12,00	0,00	310,00	3060,00	304,00	832,00
Batang	87270,00	0,00	348221,00	0,00	42500,00	129915,00	1086514,00	37059,00	0,00	0,00	0,00	0,00	0,00	0,00	4353,00
Pekalo ngan	143707,00	0,00	96785,00	0,00	2115,00	38632,00	2828530,00	1661,00	0,00	0,00	993,00	53,00	5680,00	0,00	0,00
Pemala ng	312951,00	23985,00	4015,00	197,00	14950,00	62109,00	129282,00	44106,00	71057,00	0,00	34388,00	23927,00	29072,00	43225,00	42391,00

Tegal	244523,00	0,00	435678,00	0,00	6245,00	123036,00	93057,00	586,00	510,00	0,00	1210,00	405,00	28670,00	0,00	0,00
Brebes	583105,00	0,00	1662650,00	0,00	26648,00	702551,00	240449,00	44407,00	88368,00	0,00	134841,00	1026,00	128818,00	112,00	115994,00
Kota Magelang	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kota Surakarta	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kota Salatiga	173000,00	0,00	20482,00	0,00	0,00	126500,00	6000,00	0,00	0,00	0,00	0,00	0,00	20000,00	0,00	0,00
Kota Semarang	39478,00	0,00	75100,00	0,00	8160,00	61200,00	355,00	0,00	1672,00	0,00	2600,00	0,00	56000,00	0,00	6549,00
Kota Pekalongan	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kota Tegal	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00



Lampiran 2. Syntax PCA

```
#####Principal Component Analysis#####
arfa<-read.delim('clipboard')
View(arfa)
summary(arfa)
attach(arfa)
cor(arfa[,2:16])
##fungsi cek multiko
#uji bartlett
uji_bart <- function(x)
{
  method <- "Bartlett's test of sphericity"
  data.name <- deparse(substitute(x))
  x <- subset(x, complete.cases(x))
  n <- nrow(x)
  p <- ncol(x)
  chisq <- (1-n+(2*p+5)/6)*log(det(cor(x)))
  df <- p*(p-1)/2
  p.value <- pchisq(chisq, df, lower.tail=FALSE)
  names(chisq) <- "Khi-squared"
  names(df) <- "df"
  return(structure(list(statistic=chisq, parameter=df, p.value=p.value,
                       method=method, data.name=data.name), class="htest"))
}

uji_bart(arfa[,2:16]) #H0 :p11 =p12 =...=p78 = 0

##Uji KMO
library(psych)
kmo <- function(x)
{
  x <- subset(x, complete.cases(x)) # menghilangkan data kosong (NA)
  r <- cor(x) # Membuat matrix korelasi
  r2 <- r^2 # nilai koefisien untuk r squared
  i <- solve(r) # Inverse matrix dari matrix korelasi
  d <- diag(i) # element diagonal dari inverse matrix
  p2 <- (-i/sqrt(outer(d, d)))^2 # koefisien korelasi Parsial kuadrat
  diag(r2) <- diag(p2) <- 0 # menghapus element diagonal
  KMO <- sum(r2)/(sum(r2)+sum(p2))
  MSA <- colSums(r2)/(colSums(r2)+colSums(p2))
  return(list(KMO=KMO, MSA=MSA))
}

kmo(arfa[,2:15]) #sampel cukup untuk digunakan jika kmo>0.5

###Analisis PCA###
library(factoextra)
pca_arfa<- prcomp(arfa[,2:16], scale. = TRUE)
pca_arfa
round(pca_arfa$rotation,2)
#bayaknya jumlah faktor dilihat dari nilai eigen value > 1 atau nilai varians > 80%
round(pca_arfa$sdev^2,2)
fviz_eig(pca_arfa, addlabels = TRUE, ylim = c(0, 80))
```

