



**ANALISIS MULTIVARIAT *CLUSTERING K-MEANS***  
**UNTUK APBD KECAMATAN-KECAMATAN**  
**DI KOTA SEMARANG TAHUN 2015**

Tugas Akhir  
disusun sebagai salah satu syarat  
untuk memperoleh gelar Ahli Madya  
Program Studi Statistika Terapan dan Komputasi

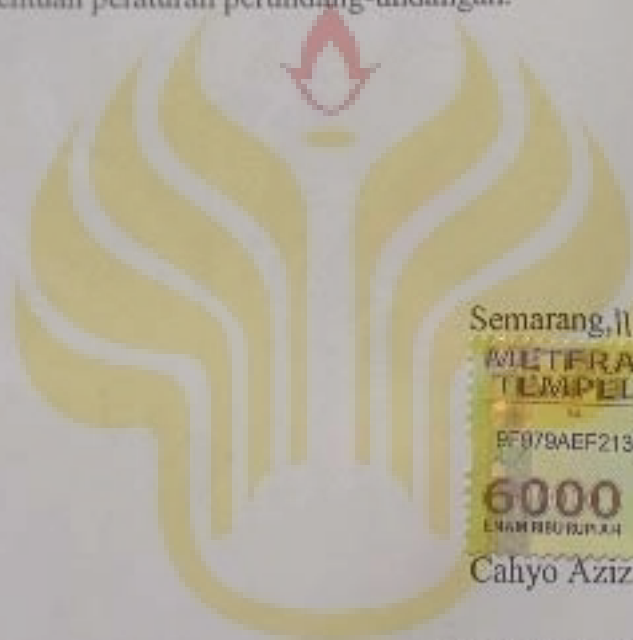
oleh:  
Cahyo Aziz Utomo  
4112314018

**JURUSAN MATEMATIKA**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS NEGERI SEMARANG**

**2017**

## PERNYATAAN

Saya menyatakan bahwa Tugas Akhir ini bebas plagiat, dan apabila di kemudian hari terbukti terdapat plagiat dalam Tugas Akhir ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.



Semarang, 11 Agustus 2017



Cahyo Aziz Utomo

4112314018

# UNNES

UNIVERSITAS NEGERI SEMARANG

## PENGESAHAN

Tugas Akhir yang berjudul

Analisis Multivariat *Clustering K-Means* untuk APBD Kecamatan-  
Kecamatan di Kota Semarang Tahun 2015

disusun oleh

Cahyo Aziz Utomo

4112314018

telah dipertahankan di hadapan sidang Panitia Ujian Tugas Akhir FMIPA UNNES  
pada tanggal 11 Agustus 2017.



Panitia  
Prof. Dr. Zaenuri S.E, M.Si, Akt  
196412231988031001

Sekretaris

Drs. Arief Agoestanto, M.Si  
196807221993031005

Penguji Utama/  
Pembimbing II

Dra. Sunarmi, M.Si  
195506241988032001

Penguji Pendamping/  
Pembimbing I

Dr. Scolastika Mariani, M.Si  
196502101991022001

## MOTTO

### *Kegigihan*

*Tidak ada satu halpun di dunia ini yang dapat mengganti ketekunan.*

*Bakat tidak akan bisa;*

*ada banyak sekali orang-orang berbakat yang gagal.*

*Kejeniusan tidak akan bisa;*

*orang jenius yang tidak dihargai adalah sesuatu yang klise.*

*Pendidikan tidak akan bisa;*

*dunia ini penuh dengan orang-orang berpendidikan yang kurang.*

*Ketekunan dan kebulatan tekad saja sudah cukup kuat.*

*Tunjukkan bahwa kamu tidak bisa dikalahkan oleh apa pun.*

*Bahwa kamu dapat memiliki ketenangan pikiran, peningkatan kesehatan dan aliran energi yang tidak pernah berhenti.*

*Jika kamu mencoba setiap hari untuk dapat mencapai hal-hal ini, hasilnya akan sangat jelas bagimu.*

*Walaupun mungkin terdengar seperti gagasan magis, tapi hal ini ada dalam dirimu untukmu menciptakan masa depanmu sendiri.*

*Penemuan terbesar dari generasiku adalah bahwa manusia dapat mengubah kehidupan mereka dengan mengubah cara berpikir mereka.*

*Atau seperti yang Ralph Waldo Emerson katakan, "Seseorang adalah apa yang dia pikirkan di sepanjang harinya."*

## PERSEMBAHAN

Tidak ada sesembahan yang berhak disembah kecuali Allah Subhanahu Wa Ta'ala.

## PRAKATA

Puji syukur penulis panjatkan kehadirat Allah Subhanahu Wa Ta'ala, karena atas rahmat, nikmat dan karunia-Nya penulis dapat menyelesaikan tugas akhir dengan judul “Analisis Multivariat *Clustering K-Means* untuk APBD Kecamatan-Kecamatan di Kota Semarang Tahun 2015”. Tugas akhir ini merupakan salah satu mata kuliah di Program Studi Statistika Terapan dan Komputasi jenjang D3 Universitas Negeri Semarang dan salah satu syarat untuk memperoleh gelar Ahli Madya.

Penulis menyadari dalam penulisan tugas akhir ini tidak lepas dari bantuan, motivasi, dan dukungan dari beberapa pihak. Oleh karena itu, penulis menyampaikan rasa hormat dan terima kasih kepada:

1. Prof. Dr. Zaenuri, S.E, M.Si, Akt, Dekan FMIPA Universitas Negeri Semarang.
2. Drs. Arief Agoestanto, M.Si, Ketua Jurusan Matematika FMIPA Universitas Negeri Semarang.
3. Dr. Wardono, M.Si, Ketua Prodi D3 Statistika Terapan dan Komputasi FMIPA Universitas Negeri Semarang.
4. Dr. Scolastika Mariani, M.Si, Dosen Pembimbing 1 Tugas Akhir yang telah memberikan bimbingan selama penyusunan Tugas Akhir.
5. Dra. Sunarmi, M.Si, Dosen Pembimbing 2 Tugas Akhir yang telah memberikan bimbingan selama penyusunan Tugas Akhir.

6. Perpustakaan Jurusan Matematika yang telah menyediakan fasilitas dan literatur untuk penyusunan Tugas Akhir ini.
7. Badan Perencanaan Pembangunan Daerah Kota Semarang yang telah menyediakan data untuk penyusunan Tugas Akhir ini.
8. Bapak dan Ibu tercinta serta keluarga yang telah memberikan motivasi dan dorongan semangat dalam mengerjakan Tugas Akhir ini.
9. Sahabat-sahabat seperjuangan Staterkom 2014 dan semua pihak yang telah membantu penulis dalam penyusunan Tugas Akhir ini.

Semoga bantuan, motivasi, dan dukungan yang telah diberikan kepada penulis dapat bermanfaat dan mendapat pahala dari Allah Subhanahu Wa Ta'ala.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini memiliki banyak kekurangan, oleh karena itu penulis mengharap saran dan kritik yang membangun. Akhir kata, penulis mengharapkan semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi penulis dan pembaca khususnya mahasiswa/i jurusan matematika Universitas Negeri Semarang.



Semarang, Agustus 2017

Penulis

## ABSTRAK

Utomo, C.A. 2017. *Analisis Multivariat Clustering K-Means untuk APBD Kecamatan-Kecamatan di Kota Semarang Tahun 2015*. Tugas Akhir. Prodi Statistika Terapan dan Komputasi, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, Pembimbing: Dr. Scolastika Mariani, M.Si, dan Dra. Sunarmi, M.Si.

**Kata kunci:** *K-Means*, Analisis Klaster, APBD.

Tujuan yang ingin dicapai melalui penelitian ini adalah (1) menerapkan metode *K-Means* dalam pembentukan klaster pada data penggunaan APBD kecamatan-kecamatan di Kota Semarang, (2) mengetahui karakteristik klaster yang terbentuk, dan (3) mengetahui variabel yang memberikan perbedaan paling besar pada klaster yang terbentuk.

Data yang digunakan pada penelitian ini adalah data sekunder dari BAPPEDA Kota Semarang. Populasi dan sampel dalam penelitian ini adalah penggunaan anggaran 9 program di 16 kecamatan di Kota Semarang tahun 2015. Variabel dalam penelitian ini adalah pengembangan kinerja pengelolaan persampahan ( $X_1$ ), pengendalian pencemaran dan kerusakan lingkungan hidup ( $X_2$ ), pelayanan administrasi perkantoran ( $X_3$ ), peningkatan sarana dan prasarana aparatur ( $X_4$ ), peningkatan pengembangan sistem pelaporan capaian kinerja dan keuangan ( $X_5$ ), peningkatan dan pengembangan pengelolaan keuangan daerah ( $X_6$ ), peningkatan sistem pengawasan internal dan pengendalian pelaksanaan kebijakan kdh ( $X_7$ ), peningkatan keberdayaan masyarakat pedesaan ( $X_8$ ), dan peningkatan partisipasi masyarakat dalam membangun desa ( $X_9$ ). Analisis data yang digunakan adalah analisis klaster *K-Means* dengan software SPSS 21.

Kesimpulan yang diperoleh dari penelitian adalah *K-Means* mengelompokkan penggunaan anggaran 16 kecamatan di Kota Semarang menjadi 3 kelompok berdasarkan penggunaan anggaran 9 program. Klaster 1 adalah kelompok kecamatan dengan penggunaan anggaran tinggi, yaitu Semarang Barat, Semarang Tengah, Genuk, Mijen, Banyumanik, Pedurungan, Gunungpati, Ngaliyan, dan Tembalang. Klaster 2 adalah kelompok kecamatan dengan penggunaan anggaran rendah, yaitu Gayamsari, Tugu, Gajah Mungkur, dan Candisari. Klaster 3 adalah kelompok kecamatan dengan penggunaan anggaran sedang, yaitu Semarang Selatan, Semarang Utara, dan Semarang Timur. Karakteristik klaster 1 berisi kecamatan yang menggunakan anggaran  $X_1, X_2$  di bawah rata-rata dan  $X_3, X_4, X_5, X_6, X_7, X_8, X_9$  di atas rata-rata. Klaster 2 berisi kecamatan yang menggunakan anggaran  $X_3$  di atas rata-rata dan  $X_1, X_2, X_4, X_5, X_6, X_7, X_8, X_9$  di bawah rata-rata. Klaster 3 berisi kecamatan yang menggunakan anggaran  $X_1, X_2, X_5$  di atas rata-rata dan  $X_3, X_4, X_6, X_7, X_8, X_9$  di bawah rata-rata. Variabel yang memberikan perbedaan paling besar pada ketiga klaster yang terbentuk adalah variabel program peningkatan partisipasi masyarakat dalam membangun desa. Sebaiknya anggaran program peningkatan partisipasi masyarakat dalam membangun desa ditingkatkan sesuai proporsi kebutuhan masing-masing klaster untuk meningkatkan partisipasi masyarakat dalam membangun desa dan meningkatkan pemberdayaan masyarakat.

## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN KEASLIAN .....	ii
HALAMAN PENGESAHAN.....	iii
MOTTO DAN PERSEMBAHAN .....	iv
PRAKATA.....	v
ABSTRAK .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN.....	xiv
<b>BAB</b>	
1. PENDAHULUAN .....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah .....	5
1.3 Pembatasan Masalah .....	5
1.4 Tujuan Penelitian .....	6
1.5 Manfaat Penelitian .....	6
1.6 Sistematika Penulisan Tugas Akhir .....	7
2. TINJAUAN PUSTAKA .....	9
2.1 Analisis Statistika.....	9
2.2 Klasifikasi Data.....	9
2.2.1 Skala Nominal .....	9
2.2.2 Skala Ordinal.....	10
2.2.3 Skala Kardinal .....	10
2.2.4 Skala Interval.....	10
2.2.5 Skala Rasio .....	10
2.3 Statistika Multivariat.....	11
2.3.1 Jenis-Jenis Analisis Statistika Multivariat.....	11



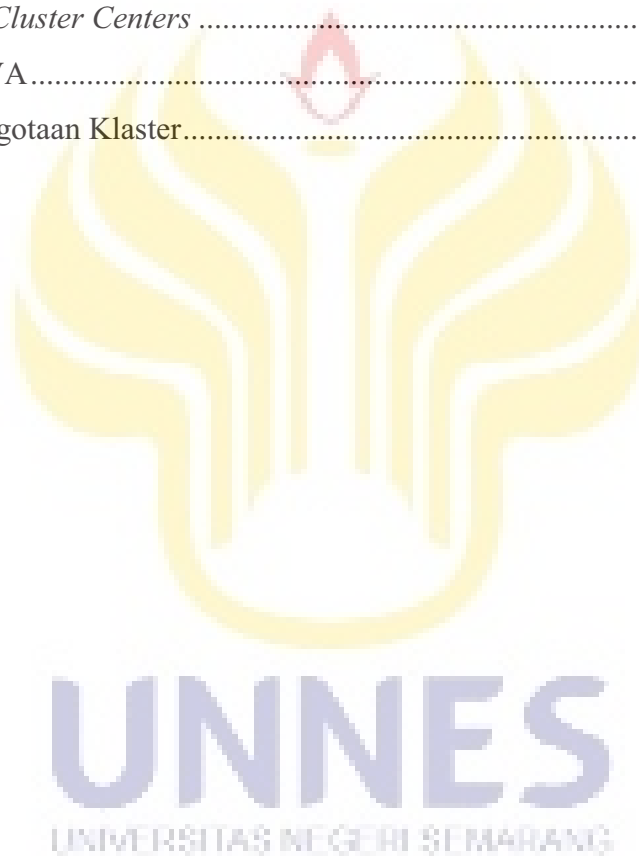
2.3.2.1 Metode Dependensi/Ketergantungan ( <i>Dependence Methods</i> ) .....	11
2.3.1.1.1 Regresi Linier Berganda .....	12
2.3.1.1.2 Analisis Diskriminan .....	13
2.3.1.1.3 Analisis Varian .....	13
2.3.1.1.4 Analisis Varian Multivariat .....	13
2.3.1.1.5 Korelasi Kanonikal .....	13
2.3.1.1.6 Analisis Konjoin .....	14
2.3.2.2 Metode Interependensi/Tidak Saling Ketergantungan ( <i>Dependence Methods</i> ) .....	14
2.3.1.2.1 Analisis Faktor .....	14
2.3.1.2.2 Analisis Kluster .....	15
2.3.1.2.3 Penskalaan Multidimensional .....	15
2.3.1.2.4 Analisis Korespondensi .....	16
2.3.2 Uji Data .....	16
2.3.2.1 <i>Missing Data</i> .....	16
2.3.2.2 <i>Outlier</i> .....	17
2.3.2.2.1 <i>Scatter Plot Data</i> .....	17
2.3.2.2.2 Standarisasi Data .....	17
2.3.2.2.3 <i>Box Plot</i> .....	18
2.3.2.3 Uji Normalitas Data .....	18
2.3.2.4 Uji Linieritas .....	19
2.3.2.5 Uji Homoskedastisitas Data .....	19
2.4 Analisis Kluster .....	19
2.4.1 Pengertian Analisis Kluster .....	19
2.4.2 Tujuan Analisis Kluster .....	20
2.4.3 Ciri dan Asumsi Analisis Kluster .....	20
2.4.4 Proses Analisis Kluster .....	21
2.4.4.1 Proses Standarisasi .....	21
2.4.4.2 Mengukur Kesamaan Objek ( <i>Similarity</i> ) .....	22
2.4.4.2.1 Ukuran Korelasi .....	22

2.4.4.2.2	Ukuran Jarak .....	23
2.4.4.2.3	Ukuran Asosiasi .....	25
2.4.4.3	Membuat Klaster (Clustering) .....	25
2.4.4.4	Interpretasi .....	25
2.4.4.5	Validasi dan <i>Profiling</i> Klaster .....	26
2.4.4.5.1	Validasi Solusi Klaster .....	26
2.4.4.5.2	<i>Profiling</i> Solusi Klaster .....	26
2.4.5	Metode Analisis Klaster .....	26
2.4.5.1	Metode Hierarki .....	27
2.4.5.1.1	Pautan Tunggal ( <i>Single Linkage/Nearest Neighbour Method</i> ) .....	28
2.4.5.1.2	Pautan Lengkap ( <i>Complete Linkage/Furthest Neighbour Method</i> ) .....	28
2.4.5.1.3	Pautan Rataan ( <i>Average Linkage/Between Groups Method</i> ) .....	28
2.4.5.1.4	Metode <i>Centroid</i> .....	28
2.4.5.1.5	<i>Ward's Error Sum of Square Method</i> .....	29
2.4.5.2	Metode Non-Hierarki .....	30
2.4.5.2.1	<i>Sequential Threshold</i> .....	30
2.4.5.2.2	<i>Parallel Threshold</i> .....	31
2.4.5.2.3	Optimalisasi .....	31
2.5	<i>K-Means Cluster</i> .....	32
2.5.1	Pengertian <i>K-Means Cluster</i> .....	32
2.5.2	Langkah-Langkah Analisis <i>K-Means Cluster</i> .....	32
2.5.2.1	Menentukan Jumlah Klaster .....	32
2.5.2.2	Menghitung Jarak Antara Objek dan Tiap <i>Centroid</i> .....	33
2.5.2.3	Pengalokasian Data ke <i>Centroid</i> Terdekat .....	33
2.5.2.4	Validasi .....	33
2.6	Anggaran Pendapatan dan Belanja Daerah .....	33
2.6.1	Urusan .....	34
2.6.2	Program .....	34

2.6.3	Kegiatan.....	34
2.6.4	Satuan Kerja Perangkat Daerah.....	35
2.6.5	Unit Kerja .....	35
2.7	Gambaran Umum SPSS.....	36
2.8	Kerangka Berpikir.....	37
3.	METODE PENELITIAN.....	42
3.1	Populasi dan Sampel .....	42
3.1.1	Populasi .....	42
3.1.2	Sampel .....	42
3.2	Variabel Penelitian dan Definisi Operasional.....	42
3.3	Pengumpulan Data .....	45
3.4	Analisis Data .....	45
4.	HASIL DAN PEMBAHASAN.....	52
4.1	Hasil Penelitian .....	52
4.1.1	Uji Multikolinieritas .....	52
4.1.2	Deskriptif.....	54
4.1.3	Analisis Kluster .....	56
4.2	Pembahasan.....	68
4.2.1	Tafsiran Kluster .....	68
4.2.2	Perbedaan Variabel.....	70
4.2.3	Komposisi Kluster .....	71
5.	PENUTUP.....	75
5.1	Simpulan.....	75
5.2	Saran .....	77
	DAFTAR PUSTAKA .....	78
	LAMPIRAN.....	80

## DAFTAR TABEL

Tabel	Halaman
4.1 Hasil Uji Multikolinieritas .....	53
4.2 Deskriptif .....	54
4.3 <i>Initial Cluster Center</i> .....	57
4.4 <i>Lteration History</i> .....	58
4.5 <i>Final Cluster Centers</i> .....	59
4.6 ANOVA .....	65
4.7 Keanggotaan Klaster .....	68



## DAFTAR GAMBAR

Gambar	Halaman
2.1 Kerangka Berpikir .....	41
3.1 Data Uji Multikolinieritas.....	46
3.2 Proses Uji Multikolinieritas.....	46
3.3 Kotak Dialog Uji Multikolinieritas <i>Linier Regression</i> .....	47
3.4 Kotak Dialog Variabel Uji Multikolinieritas <i>Linier Regression</i> .....	47
3.5 Kotak Dialog Uji Multikolinieritas <i>Linier Regression: Statistics</i> .....	48
3.6 Proses Analisis Deskriptif dan Standarisasi .....	48
3.7 Kotak Dialog Utama <i>Descriptives</i> .....	49
3.8 Kotak Dialog Proses <i>Descriptives</i> .....	49
3.9 Proses <i>K-Means</i> Klaster.....	50
3.10 Kotak Dialog Utama <i>K-Means Cluster Analysis</i> .....	50
3.11 Kotak Dialog <i>K-Means Cluster: Save</i> .....	51
3.12 Kotak Dialog <i>K-Means Cluster: Options</i> .....	51
4.1 Komposisi dan Jarak Anggota Klaster .....	71

## DAFTAR LAMPIRAN

Lampiran	Halaman
1 Data Penggunaan APBD 16 Kecamatan di Kota Semarang.....	81
2 <i>Output</i> SPSS Uji Multikolinieritas .....	83
3 <i>Output</i> SPSS Deskriptif .....	84
4 <i>Output</i> Analisis Klaster.....	85
5 Zscore .....	88
6 QCL1 dan QCL2.....	89
7 Perhitungan Tafsiran Angka Pada Klaster.....	90



# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Saat ini analisis multivariat mulai banyak digunakan dalam bidang ilmu, melengkapi analisis statistik univariat dan statistik bivariat yang dahulu menjadi satu-satunya alternatif dalam analisis data. Bahkan di masa mendatang, analisis multivariat dipastikan akan menggantikan peran pengolahan dan analisis data dari analisis univariat dan bivariat, walaupun tidak dapat menggantikannya secara total.

Analisis multivariat merupakan analisis multivariabel yang berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran pada individu atau objek (Santoso, 2015). Analisis multivariat dikelompokkan menjadi dua kelompok, yaitu analisis dependensi/ketergantungan (*dependence methods*) dan analisis interdependensi/saling tidak ketergantungan (*interdependence methods*). Analisis dependensi bertujuan untuk menjelaskan atau meramalkan nilai variabel tak bebas berdasarkan lebih dari satu variabel bebas yang mempengaruhinya. Contoh analisis dependensi, yaitu: analisis diskriminan, regresi linier berganda, analisis varian, analisis konjoin, analisis varian multivariat, dan korelasi kanonikal. Sedangkan analisis interdependensi bertujuan untuk memberikan arti (*meaning*) kepada suatu set variabel (kelompok variabel) atau menglompokkan suatu set variabel menjadi kelompok yang lebih sedikit jumlahnya dan masing-masing kelompok membentuk variabel baru yang disebut faktor (mereduksi jumlah variabel). Contoh analisis interdependensi, yaitu: analisis faktor,

analisis klaster, penskalaan multidimensional, dan analisis korespondensi (Supranto, 2004).

Analisis klaster, yaitu analisis untuk mengelompokkan elemen yang mirip sebagai objek penelitian menjadi kelompok (*cluster*) yang berbeda dan *mutually exclusive*. Berbeda dengan analisis diskriminan di mana, kelompok sudah ditentukan, kemudian suatu fungsi diskriminan bisa dipergunakan untuk menentukan suatu elemen (objek) harus masuk kelompok yang mana, misalnya seorang pelanggan termasuk yang loyal atau tidak loyal, nasabah bank peminta kredit masuk kelompok nasabah yang jujur dan yang tidak jujur. Sebaliknya, analisis klaster, dengan kriteria tertentu berdasarkan data yang ada yang ditunjukkan oleh nilai banyak variabel, membentuk kelompok (*cluster*) (Supranto, 2004).

*K-Means* adalah salah satu metode data *clustering* non-hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih klaster/kelompok. Metode ini mempartisi data ke dalam klaster/kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu klaster yang sama.

Penelitian yang bertemakan tentang analisis klaster *K-Means* telah banyak dilakukan. Penelitian terdahulu bertujuan membandingkan dan memperkuat atas hasil analisis yang dilakukan yang merujuk dari beberapa studi yang berkaitan langsung maupun tidak langsung.

Tahun 2013 Johan Oscar Ong melakukan penelitian mengenai implementasi algoritma *K-Means clustering* untuk menentukan strategi marketing President



University. Variabel yang diteliti adalah kota asal, jurusan, nilai IPK. Penelitiannya mengelompokkan marketing ke dalam 3 klaster.

Tahun 2013 Mahmudi melakukan penelitian mengenai *clustering* mahasiswa matematika berdasarkan karakteristik menggunakan metode *K-Means*. Variabel yang diteliti adalah uang saku, konsumsi, lama belajar, indeks prestasi kumulatif, lama menonton TV, lama tidur, banyak organisasi. Penelitiannya mengelompokkan mahasiswa matematika menjadi 3 golongan. Golongan 1 disebut golongan ekonomi tinggi dengan giat belajar tetapi IPK dibawah rata-rata. Golongan 2 disebut golongan ekonomi sedang tetapi giat belajar dan IPK diatas rata-rata. Golongan 3 disebut golongan ekonomi sedang dengan malas belajar dan IPK dibawah rata-rata.

Tahun 2015 Benri Melpa Metisen dan Herlina Latipa Sari melakukan penelitian mengenai analisis *clustering* menggunakan metode *K-Means* dalam pengelompokkan penjualan produk pada swalayan Fadhila. Penelitiannya mengelompokkan produk swalayan ke dalam 2 klaster. Klaster 1 adalah jenis barang yang laris terjual sedangkan klaster 2 adalah jenis barang yang tidak laris terjual.

Tahun 2015 Fina Nasari dan Surya Darma melakukan penelitian mengenai penerapan *K-Means clustering* pada data penerimaan mahasiswa baru. Penelitiannya mengelompokkan mahasiswa ke dalam 2 klaster.

Tahun 2015 Windha Mega Pradnya Duhita melakukan penelitian mengenai *clustering* menggunakan metode *K-Means* untuk menentukan status gizi balita. Variabel yang diteliti adalah tinggi badan dan berat badan. Penelitiannya

mengelompokkan balita ke dalam 5 klaster. Klaster 1 gizi buruk, klaster 2 gizi kurang, klaster 3 gizi baik, klaster 4 gizi lebih dan klaster 5 obesitas.

Badan Perencanaan Pembangunan Daerah (BAPPEDA) merupakan lembaga pemerintah yang bergerak pada bidang perencanaan pembangunan di suatu daerah. Salah satu hal yang dibahas oleh BAPPEDA adalah Anggaran Pendapatan dan Belanja Daerah (APBD). APBD adalah rencana keuangan tahunan pemerintah daerah yang dibahas dan disetujui bersama oleh pemerintah daerah dan dewan perwakilan rakyat daerah, dan ditetapkan dengan peraturan daerah. APBD disusun sesuai dengan kebutuhan penyelenggaraan pemerintahan dan kemampuan keuangan daerah.

Dalam penelitian ini akan dianalisis kasus *clustering* penggunaan APBD kecamatan-kecamatan di Kota Semarang tahun 2015, untuk mengetahui pengelompokan kecamatan-kecamatan di Kota Semarang tahun 2015 berdasarkan penggunaan anggaran pada program yang ada di kecamatan tersebut. APBD kecamatan-kecamatan di Kota Semarang terdiri atas beberapa program. Beberapa program tersebut antara lain dapat di lihat dari program pengembangan kinerja pengelolaan persampahan, program pengendalian pencemaran dan kerusakan lingkungan hidup, program pelayanan administrasi perkantoran, program peningkatan sarana dan prasarana aparatur, program peningkatan pengembangan sistem pelaporan capaian kinerja dan keuangan, program peningkatan dan pengembangan pengelolaan keuangan daerah, program peningkatan sistem pengawasan internal dan pengendalian pelaksanaan kebijakan Kdh, program peningkatan keberdayaan masyarakat pedesaan, program peningkatan partisipasi

masyarakat dalam membangun desa, dan berbagai program lainnya. Berdasarkan program-program tersebut akan di lakukan *clustering* pada kecamatan-kecamatan di Kota Semarang tahun 2015 dengan metode *K-Means* klaster.

Berdasarkan latar belakang di atas, peneliti tertarik untuk melakukan penelitian dengan judul “Analisis Multivariat *Clustering K-Means* untuk APBD Kecamatan-Kecamatan di Kota Semarang Tahun 2015”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas dapat dirumuskan masalah sebagai berikut.

1. Bagaimana penerapan metode *K-Means* dalam pembentukan klaster pada data penggunaan APBD kecamatan-kecamatan di Kota Semarang ?
2. Bagaimanakah karakteristik klaster yang terbentuk ?
3. Variabel apa yang memberikan perbedaan paling besar pada klaster yang terbentuk ?

## 1.3 Pembatasan Masalah

Untuk menghindari pembahasan yang terlalu melebar, terdapat beberapa batasan pada penelitian ini, sebagai berikut.

1. Dibatasi pada analisis klaster *K-Means* beserta teori-teori yang mendukung.
2. Studi kasus penelitian ini dibatasi pada penggunaan APBD 16 kecamatan di Kota Semarang tahun 2015.
3. Analisis data dilakukan dengan data yang diperoleh dari penggunaan anggaran 9 program di 16 kecamatan di Kota Semarang Tahun 2015 yang diperoleh dari BAPPEDA Kota Semarang.

4. Analisis data dilakukan dengan menggunakan *software* SPSS.

#### **1.4 Tujuan Penelitian**

Tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut.

1. Menerapkan metode *K-Means* dalam pembentukan kluster pada data penggunaan APBD kecamatan-kecamatan di Kota Semarang.
2. Mengetahui karakteristik kluster yang terbentuk.
3. Mengetahui variabel yang memberikan perbedaan paling besar pada kluster yang terbentuk.

#### **1.5 Manfaat Penelitian**

Adapun manfaat penelitian ini adalah sebagai berikut.

1. Bagi Mahasiswa
  - a. Mengaplikasikan ilmu yang telah diperoleh dibangku perkuliahan sehingga menunjang persiapan untuk terjun ke dunia kerja.
  - b. Menambah wawasan yang lebih luas tentang analisis kluster dan *software* SPSS.
2. Bagi Jurusan Matematika
  - a. Dapat dijadikan sebagai bahan studi kasus bagi pembaca dan acuan bagi mahasiswa.
  - b. Sebagai bahan referensi bagi pihak perpustakaan dan bahan bacaan yang dapat menambah ilmu pengetahuan bagi pembaca.
3. Bagi Instansi
  - a. Dapat dijadikan sebagai sebagai kajian dalam evaluasi APBD daerah Kota Semarang.

## 1.6 Sistematika Penulisan Tugas Akhir

Sistematika digunakan untuk mempermudah dalam memahami jalan pemikiran secara keseluruhan tugas akhir. Secara garis besar tugas akhir ini dibagi menjadi tiga bagian yaitu.

### 1. Bagian Awal

Bagian awal tugas akhir ini meliputi halaman judul, halaman pernyataan keaslian tulisan, halaman pengesahan, halaman motto dan pembahasan, prakata, abstrak, daftar isi, daftar tabel, daftar gambar, dan daftar lampiran.

### 2. Bagian Isi

Bagian isi tugas akhir ini terdiri atas lima bab. Adapun lima bab tersebut adalah sebagai berikut.

#### BAB 1 PENDAHULUAN

Pada bab ini dikemukakan latar belakang masalah, rumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan tugas akhir.

#### BAB 2 TINJAUAN PUSTAKA

Pada bab ini dikemukakan konsep-konsep yang berisi tentang teori-teori yang menjadi kerangka pikir untuk penyelesaian permasalahan dalam penelitian ini.

#### BAB 3 METODE PENELITIAN

Pada bab ini dikemukakan populasi dan sampel, variabel dan definisi operasional, pengumpulan data, dan analisis data.

#### BAB 4 HASIL DAN PEMBAHASAN

Pada bab ini dikemukakan tentang hasil penelitian dan pembahasan, sebagai jawaban dari permasalahan.

#### BAB 5 PENUTUP

Pada bab ini dikemukakan simpulan yang diperoleh dari hasil penelitian dan saran yang berkaitan dengan simpulan.

#### 3. Bagian Akhir

Bagian akhir tugas akhir ini meliputi daftar pustaka dan lampiran-lampiran yang mendukung.



## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Analisis Statistika

Berdasarkan jumlah variabel yang dianalisis, analisis statistika dapat dikelompokkan menjadi analisis univariat, bivariat dan multivariat. Analisis univariat berasal dari kata *uni* dan *variate* yang berarti analisis satu variabel. Maksud dari analisis satu variabel adalah analisis statistika yang hanya melibatkan satu variabel. Analisis bivariat berasal dari kata *bi* dan *variate* berarti analisis dua variabel. Maksud dari analisis dua variabel adalah analisis statistika yang melibatkan dua variabel. Apabila analisis lebih dari dua variabel dinamakan analisis multivariat. Analisis multivariat adalah perluasan dari analisis univariat dan bivariate (Santoso, 2014).

#### 2.2 Klasifikasi data

Menurut Sukestiyarno (2013: 2-4), berdasarkan jenis perolehannya atau pengumpulannya data diklasifikasikan menjadi 5 macam tipe atau skala. Hal tersebut adalah:

##### 2.2.1 Skala Nominal

Data nominal termasuk jenis data kualitatif. Karena, data diperoleh dari hasil pengamatan atau observasi. Apabila data disimbolkan menjadi data numerik (kuantitatif) maka bilangan yang digunakan bersifat diskrit dan tidak mengenal urutan. Untuk menyatakan kategorisasi jenis gender, misalnya kode 1 = laki-laki, 2

= perempuan. Data nominal tidak mengenal urutan, sehingga apabila dilakukan pertukaran kode tidak mempengaruhi urutan skalanya.

### **2.2.2 Skala Ordinal**

Data ordinal juga masuk dalam jenis data kualitatif, sama halnya dengan data nominal. Data diperoleh dari observasi, pengamatan atau angket berskala dari suatu variabel. Bedanya dengan data nominal adalah data ordinal mengenal suatu urutan menurut kualitas atributnya. Contoh: Sangat tidak setuju = 1, tidak setuju = 2, netral = 3, setuju = 4, sangat setuju = 5.

### **2.2.3 Skala Kardinal**

Data ini berasal dari hasil membilang atau menghitung suatu variabel. Data berbentuk kuantitatif bilangan diskrit, umumnya dinyatakan dalam bilangan kardinal. Contohnya data dari variabel jumlah kursi di setiap ruang kelas.

### **2.2.4 Skala Interval**

Data interval masuk dalam jenis data kuantitatif. Data diperoleh dari hasil mengukur, berbentuk bilangan kontinu, dan tidak memiliki nilai mutlak. Contohnya yang paling jelas adalah temperatur sebuah ruangan. Saat sebuah ruangan dinyatakan memiliki suhu  $0^{\circ}\text{C}$ , bukan berarti ruangan tersebut tidak memiliki temperatur sama sekali.

### **2.2.5 Skala Rasio**

Data rasio masuk dalam jenis data kuantitatif. Seperti halnya dengan data interval, data rasio diperoleh dari hasil mengukur, berbentuk bilangan kontinu. Perbedaannya dengan data interval adalah pada data rasio memiliki nilai nol



mutlak. Contohnya berat suatu benda, apabila benda massanya  $0\text{ kg}$  berarti tidak ada bendanya. Massa  $10\text{ kg}$  berarti 5 kali lipat dari massa  $2\text{ kg}$ .

## 2.3 Statistika Multivariat

Analisis multivariat merupakan perluasan dari analisis univariat atau bivariat. Analisis multivariat adalah analisis multi variabel dalam satu atau lebih hubungan. Analisis ini berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran pada individu atau objek (Santoso, 2014: 7). Analisis multivariat berasal dari kata *multi* dan *variate*, dua kata ini dapat diartikan analisis lebih dari dua variabel. Analisis multivariat merupakan salah satu analisis statistika yang berkaitan dengan analisis banyak variabel (Widarjono, 2015: 1). Menurut Hair, sebagaimana dikutip oleh Sarwono (2013: 1), analisis multivariat mengacu pada semua teknik statistik yang secara bersamaan menganalisis beberapa pengukuran terhadap individual atau objek dalam suatu riset. Menurut Johnson, sebagaimana dikutip oleh Sarwono (2013: 1), analisis multivariat mencakup analisis data penelitian yang menggunakan banyak variabel yang dikenakan pengukuran secara bersamaan.

### 2.3.1 Jenis-Jenis Analisis Statistika Multivariat

#### 2.3.1.1 Metode Dependensi/Ketergantungan (*Dependence Methods*)

Apabila dalam suatu analisis dapat membedakan antara variabel bebas dengan variabel terikat, maka metode ini disebut dengan metode dependensi. Karakteristik pada metode dependensi adalah variabel yang satu dengan variabel yang lain saling bergantung. Pengelompokan jenis analisis dengan metode dependensi dapat didasarkan oleh dua hal yaitu jumlah variabel terikat (*dependent*

variable) dan tipe data pada variabel bebas dan variabel terikat. Berdasarkan jumlah variabel dependen, maka analisis multivariat dikelompokkan apakah mempunyai satu variabel dependen, dua variabel dependen, ataukah beberapa variabel dependen. Selanjutnya, setelah diketahui jumlah variabel dependen maka pengelompokan selanjutnya berdasarkan jenis data variabel dependen maupun variabel independennya (Widarjono, 2015: 4).

Apabila variabel terikat berjumlah satu, terdapat tiga analisis multivariat, yaitu: regresi berganda, *conjoint analysis*, analisis diskriminan. Tipe data variabel terikat pada analisis regresi berganda dan *conjoint analysis* adalah interval atau rasio. Sedangkan tipe data variabel terikat pada analisis diskriminan adalah nominal atau ordinal.

Apabila variabel terikat berjumlah dua atau lebih, terdapat dua analisis multivariat, yaitu: analisis korelasi kanonikal, analisis varian multivariat. Tipe data variabel terikat pada analisis korelasi kanonikal dan analisis varian multivariat adalah interval atau rasio. Tipe data untuk variabel bebas pada analisis kanonikal adalah interval atau rasio sedangkan pada analisis varian multivariat tipe data untuk variabel bebas adalah nominal atau ordinal (Supranto, 2004).

#### 2.3.1.1.1 Regresi Linier Berganda

Metode analisis yang tepat dipergunakan kalau masalah penelitian (*research problem*) melibatkan satu variabel tak bebas  $Y$  yang metrik yang dipengaruhi atau terkait dengan lebih dari satu variabel bebas  $X$  yang metrik atau non-metrik. Tujuan analisis ini untuk memperkirakan/meramalkan nilai  $Y$ , kalau semua variabel bebas  $X$  sudah diketahui nilainya, dengan menggunakan persamaan

regresi linier berganda yang dibentuk dengan menggunakan metode kuadrat terkecil (*least square method*). Di samping itu juga mengetahui besarnya pengaruh setiap variabel bebas yang terdapat dalam persamaan.

#### 2.3.1.1.2 Analisis Diskriminan

Analisis yang bertujuan untuk memahami perbedaan kelompok (*group differences*) dan meramalkan *probability* bahwa suatu objek penelitian (pelanggan, karyawan, mahasiswa, barang) akan masuk/menjadi anggota kelompok tertentu, berdasarkan pada beberapa variabel bebas yang metrik. Kelompok (*group*) merupakan variabel tak bebas non-metrik. Analisis diskriminan tepat/cocok dipergunakan kalau variabel tak bebasnya non-metrik.

#### 2.3.1.1.3 Analisis Varian

Analisis untuk mengetahui dampak beberapa variabel bebas yang non-metrik (berupa kelompok) disebut perlakuan atau *treatment*, terhadap variabel tak bebas yang metrik. Pengaruh ditunjukkan dengan adanya perbedaan rata-rata.

#### 2.3.1.1.4 Analisis Varian Multivariat

Analisis untuk mengetahui dampak beberapa variabel bebas yang non-metrik (berupa kelompok) disebut perlakuan atau *treatment*, terhadap variabel tak bebas yang metrik. Dimana variabel tak bebas yang terkena pengaruh (dampak) lebih dari satu.

#### 2.3.1.1.5 Korelasi Kanonikal

Korelasi kanonikal bisa dipandang sebagai suatu perluasan yang logis dari analisis regresi berganda. Korelasi kanonikal bertujuan untuk mengkorelasikan

secara simultan (bersama-sama) beberapa variabel tak bebas  $Y$  dengan beberapa variabel bebas  $X$ .

#### 2.3.1.1.6 Analisis Konjoin

Analisis konjoin, yang disebut juga multi-atribut compositional model atau dinyatakan analisis preferensi, adalah teknik statistik yang berasal dari matematika psikologi. Analisis konjoin memberikan suatu ukuran kuantitatif mengenai kepentingan relatif (*relative importance*) suatu atribut terhadap atribut lain dari suatu produk (barang/jasa).

#### 2.3.1.2 *Metode Interdependensi/Tidak Saling Ketergantungan (Independence Methods)*

Apabila dalam suatu analisis kesulitan untuk memisahkan antara variabel bebas dengan variabel terikat, maka metode ini disebut dengan metode interdependensi, karena semua variabel dianggap independen. Karakteristik pada metode interdependensi adalah variabel yang satu dengan variabel yang lain tidak saling bergantung. Pembagian metode interdependensi dilihat dari pengolahan data berdasarkan objek atau berdasarkan variabel. Apabila pengolahan data berdasarkan variabel, maka analisis yang digunakan adalah analisis faktor. Sedangkan apabila pengolahan data berdasarkan objek, maka analisis yang digunakan adalah analisis kluster, MDS atau CA (Supranto, 2004).

#### 2.3.1.2.1 Analisis Faktor

Analisis untuk menemukan variabel baru yang disebut faktor yang jumlahnya lebih sedikit dibandingkan dengan jumlah variabel asli. Variabel baru tersebut memuat sebanyak mungkin informasi yang terkandung di dalam variabel asli. Di

dalam proses mereduksi jumlah variabel, informasi yang hilang harus seminimum mungkin. Variabel baru yang disebut faktor, bisa dipergunakan untuk melakukan analisis regresi linier berganda, sebagai variabel bebas yang tidak lagi mengalami multikolinieritas.

#### 2.3.1.2.2 Analisis Klaster

Analisis klaster, yaitu analisis untuk mengelompokkan elemen yang mirip sebagai objek penelitian menjadi kelompok (*cluster*) yang berbeda dan *mutually exclusive*. Berbeda dengan analisis diskriminan di mana, kelompok sudah ditentukan, kemudian suatu fungsi diskriminan bisa dipergunakan untuk menentukan suatu elemen (objek) harus masuk kelompok yang mana, misalnya seorang pelanggan termasuk yang loyal atau tidak loyal, nasabah bank peminta kredit masuk kelompok nasabah yang jujur dan yang tidak jujur. Sebaliknya, analisis klaster, dengan kriteria tertentu berdasarkan data yang ada yang ditunjukkan oleh nilai banyak variabel, membentuk kelompok (*cluster*).

#### 2.3.1.2.3 Penskalaan Multidimensional

*Multidimensional scaling* (MDS) adalah kumpulan teknik statistik terkait yang sering digunakan dalam visualisasi untuk menjelajahi informasi atau kesamaan *dissimilarities* dalam data. MDS bertujuan untuk membentuk pertimbangan atau penilaian pelanggaran mengenai kemiripan (*similarity*) atau preferensi (perasaan lebih suka) ke dalam jarak (*distance*) yang diwakili dalam ruang multidimensional.

#### 2.3.1.2.4 Analisis Korespondensi

Berbeda dengan teknik interdependen lainnya, yang telah dibahas di atas di dalam kemampuannya mengakomodasi dua hal yaitu data non-metrik (kualitatif, nominal dan ordinal) dan hubungna non-linier. Analisis korespondensi menggunakan suatu tabel *contingency*, yaitu tabulasi silang dari dua variabel kategori. Kemudian mengubah data non-metrik menjadi data metrik dan melakukan reduksi dimensional dan *perceptual mapping*.

### 2.3.2 Uji Data

Uji data pada prinsipnya bertujuan untuk memastikan bahwa berbagai metode multivariat (*cluster analysis*, *factor analysis* dan lainnya) bisa digunakan pada data tertentu. Dengan demikian, hasil proses multivariat bisa diinterpretasikan dengan tepat (Santoso, 2014:15). Berikut ini merupakan uji yang biasanya dilakukan pada berbagai metode statistik multivariat.

#### 2.3.2.1 *Missing Data*

*Missing data* terjadi karena data mengenai suatu kasus hilang, tidak diberikan atau memang tidak ada akan data tersebut. Hal ini menyebabkan ada sel yang kosong, contoh apabila data disajikan pada SPSS. Sebenarnya *missing data* tidak bermasalah apabila jumlah *missing data* sedikit, berkisar 1% dari seluruh data yang diteliti. Namun, jika presentase data yang hilang tersebut cukup besar, maka perlu dilakukan pengujian apakah data yang mengandung banyak *missing* tersebut masih layak diproses lebih lanjut ataukah tidak (Santoso, 2014:16). Berikut merupakan perlakuan terhadap data yang mengalami *missing data*.

1. Menghilangkan atau menghapus baris atau kolom yang mengalami *missing value*.
2. Sel yang mengalami *missing value* diisi dengan nilai tertentu, biasanya diisi dengan nilai rata-rata keseluruhan data, karena nilai rata-rata dari keseluruhan data dianggap mendekati nilai yang sebenarnya.

### 2.3.2.2 *Outlier*

*Outlier* adalah kasus atau data yang memiliki karakteristik unik yang terlihat sangat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik untuk sebuah variabel tunggal atau variabel kombinasi (Ghozali, 2011: 41). Berikut merupakan penyebab adanya data *outlier*.

1. Terjadi kesalahan dalam pengambilan sampel.
2. Terjadi kesalahan dalam mengentri data.
3. Memang terdapat data yang ekstrim.

Berikut merupakan uji yang dapat dilakukan untuk mendeteksi adanya data *outlier*.

#### 2.3.2.2.1 *Scatter Plot Data*

*Scatter plot* adalah sajian data dalam bentuk koordinat titik dari dua variabel yang satu pada sumbu  $x$  dan yang lain pada sumbu  $y$  (Sukestiyarno, 2013: 31). Sehingga *scatter plot* yang ditampilkan hanya dari dua variabel dalam suatu grafik, ini adalah kelemahan dari *scatter plot*.

#### 2.3.2.2.2 Standardisasi Data

Uji keberadaan *outlier* dapat dilakukan dengan cara standardisasi data yaitu data yang diteliti diubah ke dalam bentuk  $z$ . Berikut merupakan rumus standardisasi dengan nilai  $z$ .

(2.1)

$$z = \frac{X - \bar{X}}{\sigma}$$

di mana:

$X$  = nilai data

$\bar{X}$  = nilai rata-rata

$\sigma$  = standard deviasi

(Santoso, 2014:34)

Data *outlier* dapat dilihat dari nilai  $z$ . *Outlier* terjadi apabila nilai  $z$  berada di luar rentang antara  $-2,5$  dengan  $+2,5$ .

#### 2.3.2.2.3 *Box Plot*

*Box plot* adalah sajian data yang menggambarkan hubungan antara median ( $Q_2$ ), kuartil atas ( $Q_3$ ), dan kuartil bawah ( $Q_1$ ), termasuk pencilan data (Sukestiyarno, 2013: 35). Berikut merupakan penanganan apabila terdapat data *outlier*.

1. Data yang mengalami *outlier* dihapus atau dihilangkan.
2. Mempertahankan adanya data *outlier*, dikarenakan memang ada data *outlier*.

#### 2.3.2.3 *Uji Normalitas Data*

Tujuan uji normalitas adalah ingin mengetahui apakah distribusi sebuah data mengikuti atau mendekati distribusi normal, yakni distribusi data dengan bentuk lonceng (*bell shaped*) (Santoso, 2014:43). Berikut merupakan beberapa cara yang dapat dilakukan apabila sebaran data tidak normal.

1. Menghapus atau menghilangkan data yang dianggap menjadi penyebab data tidak normal.
2. Menambah data.



3. Melakukan transformasi data.
4. Menerima suatu data apa adanya.

#### **2.3.2.4 Uji Linearitas**

Uji linearitas digunakan untuk melihat apakah spesifikasi model yang digunakan sudah benar atau tidak (Ghozali, 2011: 166). Linieritas adalah keadaan di mana hubungan antara variabel dependen dengan variabel independen bersifat linier (garis lurus) dalam range variabel independen tertentu (Santoso, 2014: 51).

#### **2.3.2.5 Uji Homoskedastisitas Data**

Uji homoskedastisitas pada prinsipnya ingin menguji apakah sebuah grup (data kategori) mempunyai varians yang sama di antara anggota grup tersebut. Jika varians sama, dan ini yang seharusnya terjadi, maka dikatakan ada homoskedastisitas. Sedangkan jika varians tidak sama, dikatakan terjadi heteroskedastisitas (Santoso, 2014: 49).

## **2.4 Analisis Klaster**

### **2.4.1 Pengertian Analisis Klaster**

Analisis klaster, yaitu analisis untuk mengelompokkan elemen yang mirip sebagai objek penelitian menjadi kelompok (*cluster*) yang berbeda dan *mutually exclusive*. Berbeda dengan analisis diskriminan di mana, kelompok sudah ditentukan, kemudian suatu fungsi diskriminan bisa dipergunakan untuk menentukan suatu elemen (objek) harus masuk kelompok yang mana, misalnya seorang pelanggan termasuk yang loyal atau tidak loyal, nasabah bank peminta kredit masuk kelompok nasabah yang jujur dan yang tidak jujur. Sebaliknya, analisis klaster, dengan kriteria tertentu berdasarkan data yang ada yang

ditunjukkan oleh nilai banyak variabel, membentuk kelompok (*cluster*) (Supranto, 2004).

#### **2.4.2 Tujuan Analisis Klaster**

Tujuan utama analisis klaster adalah mengelompokkan objek-objek berdasarkan kesamaan karakteristik diantara objek-objek tersebut. Objek bisa berupa produk, benda, serta orang. Objek akan di klasifikasikan ke dalam dua atau lebih klaster sehingga objek-objek yang berada dalam satu klaster akan mempunyai kemiripan satu dengan yang lain. Sebuah klaster adalah koleksi dari data objek yang mirip dengan objek yang lain dalam satu klaster yang sama dan berbeda dengan objek yang berada di klaster yang berlainan. Klaster dari suatu objek data dapat diperlakukan secara kolektif sebagai satu grup sehingga di anggap sebagai bentuk data *compression* (Supranto, 2004).

Seperti yang telah dibahas sebelumnya, analisis klaster merupakan kelompok sehingga pengelompokan yang akan dilakukan bertujuan agar data yang terdapat di dalam kelompok yang sama relatif lebih homogen daripada data yang berada pada kelompok yang berbeda. Diharapkan dengan terbentuknya kelompok tersebut akan lebih mudah untuk menganalisa dan lebih tepat dalam pengambilan keputusan sehubungan dengan masalah yang ada.

#### **2.4.3 Ciri dan Asumsi Analisis Klaster**

Menurut Santoso (2014), klaster yang baik adalah klaster yang memiliki ciri:

1. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster (*within cluster*).

2. Heterogenitas (perbedaan) yang tinggi antar kluster yang satu dengan yang lainnya (*between cluster*).

Menurut Santoso (2014), asumsi yang harus di penuhi pada analisis kluster yaitu:

1. Sampel yang diambil harus benar-benar bisa mewakili populasi yang ada.
2. Multikolinieritas, yakni kemungkinan adanya korelasi antar objek. Namun sebaiknya tidak ada, bila ada besar multikolinieritas tersebut tidaklah tinggi (misal di atas 0,5).

#### 2.4.4 Proses Analisis Kluster

##### 2.4.4.1 Proses Standarisasi

Menurut Santoso (2014), bila terdapat data (variabel) yang mempunyai perbedaan ukuran satuan yang besar perlu dilakukan proses standarisasi. Perbedaan satuan yang mencolok akan menyebabkan bias dalam analisis kluster, sehingga data asli harus ditransformasi (standarisasi) sebelum bisa di analisis. Transformasi dilakukan terhadap variabel yang relevan ke dalam bentuk  $Z$  skor, seperti berikut:

(2.2)

$$Z = \frac{X - \bar{X}}{\sigma}$$

di mana:

$X$  : Nilai data

$\bar{X}$  : Nilai rata-rata

$\sigma$  : Standar deviasi

#### 2.4.4.2 *Mengukur Kesamaan Antar Objek (Similarity)*

Konsep kesamaan adalah hal yang fundamental dalam analisis kluster. Kesamaan antar objek merupakan ukuran korespondensi antar objek. Ada tiga metode yang dapat diterapkan, yaitu ukuran korelasi, ukuran jarak, dan ukuran asosiasi. Sesuai prinsip dasar kluster yaitu mengelompokkan objek yang mempunyai kemiripan (Supranto, 2004).

##### 2.4.4.2.1 Ukuran Korelasi

Ukuran ini dapat diterapkan pada data dengan skala metrik, namun jarang digunakan karena titik beratnya pada nilai suatu pola tertentu, padahal fisik berat analisis kluster adalah besarnya objek. Kesamaan antar objek dapat dilihat dari koefisien korelasi antar pasangan objek yang diukur dengan bebraa variabel (Supranto, 2004).

Ukuran kesamaan ini jarang digunakan karena titik beratnya pada nilai suatu pola tertentu (korelasi). Padahal titik berat analisis kluster adalah pada besarnya korelasinya, sehingga proses analisis kluster akan terkendala jika terdapat data yang saling tumpang tindih (Supranto, 2004). Yang termasuk jenis ukuran kesamaan adalah:

##### 1. Ukuran Mutlak Korelasi

Dalam ukuran kesamaan ini tidak mempertimbangkan nilai korelasi yang positif ataupun negatif dari objek yang dianalisis. Ukuran ini menggunakan nilai mutlak dari korelasinya sehingga ukuran kemiripan objek  $i$  terhadap  $j$  atau dinotasikan  $d_{ij}$  dapat dirumuskan sebagai berikut:

(2.3)

$$d_{ij} = 1 - |r_{ij}|$$

Dengan  $r_{ij}$  adalah koefisien korelasi antar objek ke- $i$  dan objek ke- $j$  dan  $|r_{ij}|$  merupakan nilai mutlaknya (Mahmudi, 2013).

## 2. Ukuran Jarak Korelasi

Dalam ukuran kesamaan ini nilai korelasi yang positif ataupun negatif dari objek yang dianalisis mempengaruhi pengukuran kesamaannya. Ukuran kemiripan objek  $i$  dengan objek  $j$  dapat dirumuskan sebagai berikut:

(2.4)

$$d_{ij} = 1 - r_{ij}$$

Dengan  $r_{ij}$  adalah koefisien korelasi objek ke- $i$  dan objek ke- $j$  (Mahmudi, 2013).

### 2.4.4.2.2 Ukuran Jarak

Merupakan ukuran yang paling sering digunakan. Ditetapkan untuk data berskala metrik. Sebenarnya merupakan ukuran ketidakmiripan, dimana jarak yang besar menunjukkan sedikit kesamaan sebaliknya jarak yang pendek/kecil menunjukkan bahwa suatu objek makin mirip dengan objek lain. Bedanya dengan ukuran korelasi adalah bahwa ukuran jarak fokusnya pada besarnya nilai. Klaster berdasarkan ukuran korelasi bisa saja tidak memiliki kesamaan nilai tapi memiliki kesamaan pola, sedangkan klaster berdasarkan ukuran jarak lebih memiliki kesamaan nilai meskipun polanya berbeda. Mengukur seberapa jauh ada kesamaan antar objek. Metode yang dapat digunakan untuk mengukur kesamaan adalah dengan mengukur jarak (*distance*) antar dua objek (Mahmudi, 2013).

Ukuran ketidakmiripan antara objek ke- $r$  dengan ke- $s$  disebut sebagai jarak yang disimbolkan dengan  $d_{rs}$ . Semakin kecil nilai dari  $d_{rs}$  semaik mirip kedua objek tersebut. Ada beberapa cara menentukan  $d_{rs}$  antara lain

*Euclidean Distance*

(2.5)

$$d_{rs} = \sqrt{\sum_{i=1}^p (X_{ri} - X_{si})^2}$$

*Minkowski Metric*

(2.6)

$$d_{rs} = \sqrt[m]{\sum_{i=1}^p (X_{ri} - X_{si})^m}$$

Untuk  $m = 1$ ,  $d_{rs}$  menjadi *city block distance* antar dua titik dengan  $p$  dimensi. Untuk  $m = 2$ ,  $d_{rs}$  menjadi *euclidean distance*. Secara umum untuk  $m$  makin besar, perbedaannya semakin kecil.

*Cambera Metric*

(2.7)

$$d_{rs} = \sum_{i=1}^p \frac{|X_{ri} - X_{si}|}{(X_{ri} + X_{si})}$$

*Czekanowski Coefisient*

(2.8)

$$d_{rs} = 1 - \frac{2 \sum_{i=1}^p \min(X_{ri}, X_{si})}{\sum_{i=1}^p (X_{ri} + X_{si})}$$

Diantara metode di atas, yang paling populer adalah metode *euclidean distance*. Untuk data non-metrik yaitu data yang didapat dengan menghitung, tidak mempunyai desimal serta dilakukan dengan kategori diperlukan pengukuran asosiasi antar objek (Mahmudi, 2013).

#### 2.4.4.2.3 Ukuran Asosiasi

Dalam analisis data berskala non-matrik, ukuran kesamaan atau kemiripan digunakan ukuran asosiasi karena pada data berskala non-metrik bentuk data lebih menggerombol. Ukuran kesamaan ini adalah mengamil bentuk-bentuk dari koefisien tiap sampel/objeknya dengan korelasi-korelasi yang bernilai negatif diganti dengan memutlakkkan nilainya (Mahmudi, 2013).

#### 2.4.4.3 *Membuat Klaster (clustering)*

Ada dua proses penting yaitu algoritma klaster dalam pembentukan klaster dan menentukan jumlah klaster yang akan dibentuk. Keduanya mempunyai implikasi substansial tidak hanya pada hasil yang diperoleh tetapi juga pada interpretasi yang akan dilakukan terhadap hasil tersebut (Supranto, 2004).

#### 2.4.4.4 *Interpretasi*

Setelah klaster terbentuk, langkah selanjutnya melakukan interpretasi. Tahap interpretasi meliputi pengujian tiap klaster dalam term untuk menamai dan menandai dengan suatu label secara akurat dapat menjelaskan kealamian klaster. Proses ini dimulai dengan suatu ukuran yang sering digunakan yaitu *centroid cluster*. Membuat profil dan interpretasi klaster tidak hanya untuk memperoleh suatu gambaran saja melainkan pertama, menyediakan suatu rata-rata untuk menilai korespondensi pada klaster yang terbentuk, kedua, profil klaster memberikan arah

bagi penelitian terhadap signifikansi praktis. Pada intinya memberi nama spesifik untuk menggambarkan isi kluster tersebut (Supranto, 2004).

#### **2.4.4.5 Validasi dan Profiling Kluster**

##### **2.4.4.5.1 Validasi Solusi Kluster**

Proses validasi bertujuan menjamin bahwa solusi yang dihasilkan dari analisis kluster dapat mewakili populasi dan dapat digeneralisasi untuk objek lain. Pendekatan ini membandingkan solusi kluster dan menilai korespondensi hasil. Terkadang tidak dapat dipraktekkan karena adanya kendala waktu dan biaya atau ketidakterediaan objek untuk analisis kluster ganda (Supranto, 2004).

##### **2.4.4.5.2 Profiling Solusi Kluster**

Tahap ini menggambarkan karakteristik tiap kluster untuk menjelaskan kluster-kluster tersebut dapat berbeda pada dimensi yang relevan. Titik beratnya pada karakteristik yang secara signifikan berbeda antar kluster dan memprediksi anggota dalam suatu kluster khusus (Supranto, 2004).

Secara keseluruhan proses analisis kluster berakhir setelah kelima tahap ini dilalui. Hasil analisis kluster dapat digunakan untuk berbagai kepentingan sesuai dengan materi yang dianalisis.

#### **2.4.5 Metode Analisis Kluster**

Dua metode paling umum dalam algoritma kluster adalah metode hierarki dan metode non-hierarki. Keduanya memiliki kelebihan sendiri-sendiri. Keuntungan metode hierarki adalah cepat dalam proses pengolahan sehingga menghemat waktu, namun kelemahannya metode ini dapat menimbulkan kesalahan. Selain itu tidak baik diterapkan untuk menganalisis sampel dengan



ukuran besar. Metode non-hierarki memiliki keuntungan lebih daripada metode hierarki. Hasilnya memiliki sedikit kelemahan pada data *outlier*, ukuran jarak yang digunakan, dan termasuk variabel tak relevan atau variabel yang tidak tepat. Keuntungannya hanya dengan menggunakan titik bakal non *random*, penggunaan metode non-hierarki untuk titik bakal *random* secara nyata lebih buruk dari pada metode hierarki. Penentuan metode ini yang akan dipakai tergantung kepada peneliti dan konteks penelitian dengan tidak mengabaikan substansi, teori dan konsep yang berlaku. Alternatif lain adalah dengan mengkombinasikan kedua metode ini. Pertama gunakan metode hierarki kemudian dilanjutkan dengan metode non-hierarki (Supranto, 2004).

#### 2.4.5.1 *Metode Hierarki*

Pada metode hierarki pengelompokan dimulai dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses dilanjutkan ke-objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga klaster membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah klaster. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Supranto, 2004).

Andaikan  $d_{ij}$  merupakan ukuran ketakmiripan antara klaster ke- $i$  dengan klaster ke- $j$  dan  $d_{k(i,j)}$  merupakan ukuran ketakmiripan antara klaster ke- $k$  dengan klaster  $(i,j)$  yang merupakan penggabungan antara klaster ke- $i$  dan klaster ke- $j$ . Berikut ini diberikan ukuran ketakmiripan antar klaster.

#### 2.4.5.1.1 Pautan Tunggal (*Single Linkage/Nearest Neighbour Method*)

Metode ini akan mengelompokkan dua objek yang mempunyai jarak terdekat terlebih dahulu (Santoso, 2014). Ukuran ketakmiripan yang digunakan adalah

(2.9)

$$d_{k(i,j)} = \min\{d_{ki}, d_{kj}\}$$

#### 2.4.5.1.2 Pautan Lengkap (*Complete Linkage/Furthest Neighbour Method*)

Metode ini akan mengelompokkan dua objek yang mempunyai jarak terjauh terlebih dahulu (Santoso, 2014). Ukuran ketakmiripan yang digunakan adalah

(2.10)

$$d_{k(i,j)} = \max\{d_{ki}, d_{kj}\}$$

#### 2.4.5.1.3 Pautan Rataan (*Average Linkage/Between Groups Method*)

Metode ini akan mengelompokkan objek berdasar jarak rata-rata yang didapat dengan melakukan rata-rata semua jarak antar objek terlebih dahulu (Santoso, 2014). Ukuran ketakmiripan yang digunakan adalah

(2.11)

$$d_{k(i,j)} = \frac{n_i}{n_i + n_j} d_{ki} + \frac{n_j}{n_i + n_j} d_{kj}$$

Dengan  $n_i$  adalah jumlah item pada kluster  $i$ .

#### 2.4.5.1.4 Metode *Centroid*

Pada metode ini, jarak antara dua kluster adalah jarak di antara dua *centroid* kluster-kluster tersebut. *Centroid* adalah rata-rata jarak yang ada pada sebuah kluster, yang didapat dengan melakukan rata-rata pada semua anggota suatu kluster tertentu. Dengan metode ini, setiap terjadi kluster baru, segera terjadi perhitungan

ulang *centroid*, sampai terbentuk kluster yang tetap. Ukuran ketakmiripan yang digunakan adalah

(2.12)

$$d_{k(i,j)} = \frac{n_i}{n_i + n_j} d_{ki} + \frac{n_j}{n_i + n_j} d_{kj} - \frac{n_i + n_j}{(n_i + n_j)^2} d_{ij}$$

#### 2.4.5.1.5 Ward's Error Sum of Square Method

Metode *ward's* ditemukan oleh Ward (1963) tidak menghitung jarak antar kluster atau objek, namun metode ini membentuk kluster-kluster dengan memaksimalkan kehomogenan dalam kluster. Jumlah kuadrat dalam kluster digunakan sebagai ukuran kehomogenan. Metode *ward's* terus mencoba meminimalkan total jumlah kuadrat dalam kluster. Kluster-kluster dibentuk pada masing-masing tahap seolah-olah telah menjadi solusi analisis kluster dan solusi terbaik adalah kombinasi kluster yang mempunyai jumlah kuadrat dalam kluster terkecil. Jumlah kuadrat dalam kluster (jarak kuadrat *euclidean* masing-masing objek terhadap *mean* kluster yang memuat objek kluster tersebut) yang meminimalkan sering disebut

(2.13)

$$ESS = \sum_{k=1}^k \left\{ \sum_{i=1}^{n_k} \sum_{j=1}^p x_{ijk}^2 - \frac{1}{n_k} \left( \sum_{i=1}^{n_k} x_{ijk} \right)^2 \right\}$$

Ukuran ketidakmiripan di atas dapat ditulis secara umum sebagai

(2.14)

$$d_{k(i,j)} = a_i d_{kj} + a_j d_{ki} + \beta a_{ij} + d^3 d_{ki} - d^3 d_{kj}$$

Yang akan menjadi ukuran ketidakmiripan pautan tunggal bila  $a_i = a_j = \frac{1}{2}$ ,  $\beta = 0$ , dan  $d = -\frac{1}{2}$ . Akan menjadi pautan lengkap bila  $a_i = a_j = \frac{1}{2}$ ,  $\beta = 0$ , dan  $d = \frac{1}{2}$ . Akan menjadi pautan rata-rata bila  $a_i = \frac{n_i}{n_i+n_j}$ ,  $a_j = \frac{n_j}{n_i+n_j}$ ,  $\beta = -a_i a_j$  dan  $d = 0$ .

#### 2.4.5.2 Metode Non-Hierarki

Berbeda dengan metode hierarki, metode ini dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan. Setelah jumlah kluster diketahui, baru proses kluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut sebagai *K-Means cluster* (Supranto, 2004). Menurut Santoso (2014), ada tiga prosedur dalam metode non-hierarki, yaitu:

##### 2.4.5.2.1 Sequential Threshold

Proses *clustering* dengan menggunakan metode *sequential threshold* dimulai dengan memilih sebuah kluster *seed* dan kemudian menggabungkan setiap objek yang ada dalam jarak yang telah ditentukan sebelumnya. Kluster tersebut akan disebut sebagai kluster pertama. Setelah kluster pertama terbentuk, maka kluster *seed* kedua digunakan dan kemudian objek-objek yang mempunyai jarak terdekat akan digabungkan. Jika telah selesai, maka kluster berikutnya akan dibentuk dengan cara yang sama.

Metode ini disebut dengan *sequential threshold* karena proses *clustering* dilakukan berurutan dari kluster pertama, kedua, ketiga dan berikutnya. Metode ini tidak memperbolehkan suatu objek pindah ke kluster lain.

#### 2.4.5.2.2 *Parallel Threshold*

Proses *clustering* dengan menggunakan metode *parallel threshold* dimulai dengan memilih klaster *seed* yang akan dijadikan patokan pembuatan klaster. Setiap objek akan diukur terhadap klaster *seed* tersebut. Sebuah objek akan masuk ke suatu klaster jika mempunyai nilai jarak terhadap suatu klaster *seed* lebih dekat daripada klaster *seed* yang lain. Langkah ini mengakibatkan penentuan klaster tidak berurutan.

Metode ini disebut dengan *parallel threshold* karena proses *clustering* dilakukan tidak berurutan. Metode ini tidak memperbolehkan suatu objek yang sudah menjadi anggota suatu klaster berpindah ke klaster lain.

#### 2.4.5.2.3 Optimalisasi

Proses *clustering* dengan menggunakan metode ini mirip dengan *sequential threshold* atau *parallel threshold clustering*, perbedaannya adalah diperbolehkannya suatu objek berpindah dari suatu klaster ke klaster lain. Ini terjadi ketika jarak ke suatu klaster ternyata lebih dekat daripada jarak objek tersebut ke klaster yang sekarang.

Proses optimalisasi dilakukan dengan menghitung ulang setiap objek terhadap semua *centroid* klaster yang ada. Jika objek tersebut mempunyai jarak yang lebih kecil terhadap suatu *centroid* klaster lain daripada *centroid* klaster sekarang, maka objek tersebut direlokasi ke klaster terdekat tersebut.

## 2.5 *K-Means Cluster*

### 2.5.1 *Pengertian K-Means Cluster*

Metode *K-Means* merupakan salah satu metode analisis kluster non-hierarki yang di golongkan sebagai metode pengklasifikasian yang bersifat *unsupervised* (tanpa arahan) karena data yang di analisis tidak mempunyai label kelas yang berarti dalam proses pengelompokkannya, analisis ini tidak mempunyai anggota kluster yang pasti. Melainkan data yang sudah masuk kedalam kluster bisa saja berpindah ke kluster yang lain. Akan tetapi, karena peneliti sering menentukan sendiri jumlah kluster awal, baik dengan menggunakan metode tertentu atau berdasarkan pengalaman, maka metode *K-Means* ini disebut sebagai metode *semi-supervised classification* (Agusta, 2007).

Dalam metode ini di asumsikan terlebih dahulu jumlah kelompok atau kluster yang diinginkan pada akhir suatu pengelompokkan. Metode kluster jenis ini di mungkinkan bahwa yang telah berada ke suatu kluster tertentu dapat pindah ke kelompok lain. Istilah *K-Means* sendiri di temukan untuk mendeskripsikan bahwa algoritma ini memadai setiap objek masuk ke dalam kelompok yang mempunyai rata-rata (pusat kluster terdekat) (Agusta, 2007).

### 2.5.2 *Langkah-Langkah Analisis K-Means Cluster*

Menurut Agusta (2007), data *clustering* menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

#### 2.5.2.1 *Menentukan Jumlah Kluster*

Tahap awal dalam analisis kluster dengan metode *K-Means* adalah dengan menentukan besarnya  $k$ , yaitu banyaknya kluster dan menentukan *centroid* atau

pusat kluster pada tiap kluster. Dalam menentukan jumlah kluster dan *centroid* pada tiap kluster bersifat objektif.

#### **2.5.2.2 Menghitung Jarak Antara Objek dan Tiap Centroid**

Untuk menghitung jarak tiap objek dengan setiap *centroid* menggunakan ukuran jarak pada ukuran kesamaan objek dalam kluster. Ukuran jarak yang digunakan dalam metode *K-Means* adalah jarak *euclidean*.

#### **2.5.2.3 Pengalokasian Data ke Centroid Terdekat**

Cari data yang lebih dekat dengan pusat kluster. Tandai titik data tersebut di pusat kluster yang terdekat dan posisi pusat kluster dihitung kembali dengan rata-rata anggota dari setiap kluster.

#### **2.5.2.4 Validasi**

Cek semua data sekali lagi dan letakan setiap data yang terdekat dengan pusat kluster (pusat kluster tidak dihitung lagi). Jika anggota dari tiap pusat kluster tidak berubah, berhenti dan jika masih berubah kembali ke langkah tiga.

### **2.6 Anggaran Pendapatan dan Belanja Daerah**

Anggaran Pendapatan dan Belanja Daerah, selanjutnya disingkat APBD adalah rencana keuangan tahunan pemerintahan daerah yang dibahas dan disetujui bersama oleh pemerintah daerah dan DPRD, dan ditetapkan dengan peraturan daerah (Peraturan Menteri Dalam Negeri Nomor 25 Tahun 2009). Anggaran Pendapatan dan Belanja Daerah, yang selanjutnya disingkat APBD adalah APBD Kota Semarang yang merupakan rencana keuangan tahunan daerah yang ditetapkan dengan Perda (Peraturan Walikota Semarang Nomor 32 Tahun 2015).

### **2.6.1 Urusan**

Urusan pemerintahan adalah fungsi-fungsi pemerintahan yang menjadi hak dan kewajiban setiap tingkatan dan/atau susunan pemerintahan untuk mengatur dan mengurus fungsi-fungsi tersebut yang menjadi kewenangannya dalam rangka melindungi, melayani, memberdayakan, dan mensejahterakan masyarakat (Peraturan Menteri Dalam Negeri Nomor 13 Tahun 2006). Urusan Pemerintahan Wajib adalah Urusan Pemerintahan yang wajib diselenggarakan oleh Daerah sesuai dengan potensi yang dimiliki Daerah. Urusan Pemerintahan Pilihan adalah Urusan Pemerintahan yang wajib diselenggarakan oleh Daerah sesuai dengan potensi yang dimiliki daerah (Peraturan Walikota Semarang Nomor 32 Tahun 2015).

### **2.6.2 Program**

Program adalah penjabaran kebijakan SKPD dalam bentuk upaya yang berisi satu atau lebih kegiatan dengan menggunakan sumber daya yang disediakan untuk mencapai hasil yang terukur sesuai dengan misi SKPD (Peraturan Menteri Dalam Negeri Nomor 13 Tahun 2006). Program adalah bentuk instrumen kebijakan yang berisi 1 (satu) atau lebih kegiatan yang dilaksanakan oleh SKPD atau masyarakat, yang dikoordinasikan oleh pemerintah daerah untuk mencapai sasaran dan tujuan pembangunan daerah (Peraturan Walikota Semarang Nomor 32 Tahun 2015).

### **2.6.3 Kegiatan**

Kegiatan adalah bagian dari program yang dilaksanakan oleh satu atau lebih unit kerja pada SKPD sebagai bagian dari pencapaian sasaran terukur pada suatu program dan terdiri dari sekumpulan tindakan pengerahan sumber daya baik yang



berupa personil (sumber daya manusia), barang modal termasuk peralatan dan teknologi, dana, atau kombinasi dari beberapa atau kesemua jenis sumber daya tersebut sebagai masukan (*input*) untuk menghasilkan keluaran (*output*) dalam bentuk barang/jasa (Peraturan Menteri Dalam Negeri Nomor 13 Tahun 2006). Kegiatan adalah bagian dari program yang dilaksanakan oleh 1 (satu) atau beberapa SKPD sebagai bagian dari pencapaian sasaran terukur pada suatu program, dan terdiri dari sekumpulan tindakan pengalokasian sumber daya baik yang berupa personil (sumber daya manusia), barang modal termasuk peralatan dan teknologi, dana, atau kombinasi dari beberapa atau kesemua jenis sumber daya tersebut, sebagai masukan (*input*) untuk menghasilkan keluaran (*output*) dalam bentuk barang/jasa (Peraturan Walikota Semarang Nomor 32 Tahun 2015).

#### **2.6.4 Satuan Kerja Perangkat Daerah**

Satuan Kerja Perangkat Daerah yang selanjutnya disingkat SKPD adalah perangkat daerah pada pemerintah daerah selaku pengguna anggaran/pengguna barang (Peraturan Menteri Dalam Negeri Nomor 13 Tahun 2006). Satuan Kerja Perangkat Daerah, yang selanjutnya disingkat SKPD adalah perangkat daerah yang terdiri dari Sekretariat Daerah, Sekretariat DPRD, Badan, Dinas, Kantor, dan Kecamatan di lingkungan Pemerintah Kota Semarang yang mempunyai tugas mengelola anggaran dan barang daerah (Peraturan Walikota Semarang Nomor 32 Tahun 2015).

#### **2.6.5 Unit Kerja**

Unit kerja adalah bagian dari SKPD yang melaksanakan satu atau beberapa program (Peraturan Menteri Dalam Negeri Nomor 13 Tahun 2006). Kecamatan

adalah wilayah dari daerah kabupaten/kota yang dipimpin oleh camat yang dibentuk dengan Perda Kabupaten/Kota yang berpedoman pada peraturan pemerintah. Kelurahan adalah wilayah dari kecamatan yang dipimpin oleh Lurah sebagai perangkat kecamatan dan bertanggungjawab kepada camat (Peraturan Walikota Semarang Nomor 32 Tahun 2015).

## 2.7 Gambaran Umum SPSS

Perkembangan teknologi pengolahan data yang berkaitan dengan penelitian telah meningkat sedemikian pesatnya sehingga disadari atau tidak, memaksa perusahaan-perusahaan pembuat *software* pengolahan data untuk berlomba-lomba menyediakan *software* yang dapat memenuhi kebutuhan tersebut. Karena pentingnya dalam pengolahan data tersebut maka diperlukan suatu pengaplikasian data dengan komputerisasi. Komputerisasi merupakan suatu alternatif yang tepat dalam memenuhi kebutuhan dalam pengolahan data secara cepat, tepat dan mudah. Untuk mendapatkan nilai data yang valid dari penggunaan komputer tersebut maka kita perlu menggunakan program yang dapat membantu kita untuk menyelesaikan suatu masalah tertentu. Ada beberapa program yang dapat digunakan dalam pengolahan data, namun salah satu yang paling populer adalah SPSS.

SPSS kepanjangan dari *Statistical Package for Social Science*. SPSS merupakan *software* statistik terpopuler di dunia, termasuk di Indonesia. SPSS dapat digunakan untuk menganalisis dalam berbagai bidang, berbagai pilihan metode analisis, fasilitas yang disediakan lengkap, menu-menu sangat *user friendly*, mudah dimengerti. SPSS berkembang pada tahun 1960-an oleh Norman H. Nie, C. Hadlai Hull dan Dale H. Brent. Mereka bertujuan untuk mengembangkan sebuah

*system software* yang berdasarkan pada ide dari penggunaan statistik untuk menganalisis data menjadi informasi yang penting untuk pengambilan keputusan. Untuk pengoperasian SPSS pengguna tidak perlu khawatir, karena SPSS telah dilengkapi fitur cara untuk mengoperasikan SPSS. SPSS memiliki versi yang banyak karena hampir setiap tahun merilis versi yang baru, tetapi antara versi satu dengan yang lain tidak memiliki perbedaan yang signifikan.

Program SPSS *For windows* pada umumnya digunakan untuk memecahkan suatu permasalahan riset atau bisnis dalam hal statistika atau manajemen data, khususnya dalam penelitian dan analisis. Cara kerjanya adalah dengan membandingkan suatu data ke dalam suatu paket analisis. Keunggulan SPSS antara lain lebih mudah dalam penggunaan dan mudah dipahami. Selain itu SPSS merupakan suatu bagian *integral* tentang proses analisis, menyediakan kemampuan untuk akses data, persiapan dan manajemen data, analisis data, serta dalam hal pelaporan.

## **2.8 Kerangka Berpikir**

Berdasarkan jumlah variabel yang dianalisis, analisis statistika dapat dikelompokkan menjadi analisis univariat, bivariat dan multivariat. Analisis univariat adalah analisis statistika yang hanya melibatkan satu variabel. Analisis bivariat adalah analisis statistika yang melibatkan dua variabel. Analisis lebih dari dua variabel dinamakan analisis multivariat. Menurut Hair, sebagaimana dikutip oleh Sarwono (2013: 1), analisis multivariat mengacu pada semua teknik statistik yang secara bersamaan menganalisis beberapa pengukuran terhadap individual atau objek dalam suatu riset. Berdasarkan jenis analisis, statistik multivariat dapat

dikelompokkan menjadi dua metode, yaitu metode dependensi dan metode interdependensi.

Apabila dalam suatu analisis dapat membedakan antara variabel bebas dengan variabel terikat, maka metode ini disebut dengan metode dependensi. Karakteristik pada metode dependensi adalah variabel yang satu dengan variabel yang lain saling bergantung (Widarjono, 2015: 4). Apabila variabel terikat berjumlah satu, terdapat tiga analisis multivariat, yaitu: regresi berganda, *conjoint analysis*, analisis diskriminan. Tipe data variabel terikat pada analisis regresi berganda dan *conjoint analysis* adalah interval atau rasio. Sedangkan tipe data variabel terikat pada analisis diskriminan adalah nominal atau ordinal. Apabila variabel terikat berjumlah dua atau lebih, terdapat dua analisis multivariat, yaitu: analisis korelasi kanonikal, analisis varian multivariat. Tipe data variabel terikat pada analisis korelasi kanonikal dan analisis varian multivariat adalah interval atau rasio. Tipe data untuk variabel bebas pada analisis kanonikal adalah interval atau rasio sedangkan pada analisis varian multivariat tipe data untuk variabel bebas adalah nominal atau ordinal (Supranto, 2004).

Apabila dalam suatu analisis kesulitan untuk memisahkan antara variabel bebas dengan variabel terikat, maka metode ini disebut dengan metode interdependensi, karena semua variabel dianggap independen. Karakteristik pada metode interdependensi adalah variabel yang satu dengan variabel yang lain tidak saling bergantung. Pembagian metode interdependensi dilihat dari pengolahan data berdasarkan objek atau berdasarkan variabel. Apabila pengolahan data berdasarkan variabel, maka analisis yang digunakan adalah analisis faktor. Sedangkan apabila

pengolahan data berdasarkan objek, maka analisis yang digunakan adalah analisis kluster, MDS atau CA (Spuranto, 2004).

Analisis kluster, yaitu analisis untuk mengelompokkan elemen yang mirip sebagai objek penelitian menjadi kelompok (*cluster*) yang berbeda dan *mutually exclusive*. Tujuan utama analisis kluster adalah mengelompokkan objek-objek berdasarkan kesamaan karakteristik diantara objek-objek tersebut. Objek bisa berupa produk, benda, serta orang. Dua metode paling umum dalam algoritma kluster adalah metode hierarki dan metode non-hierarki (Supranto, 2004).

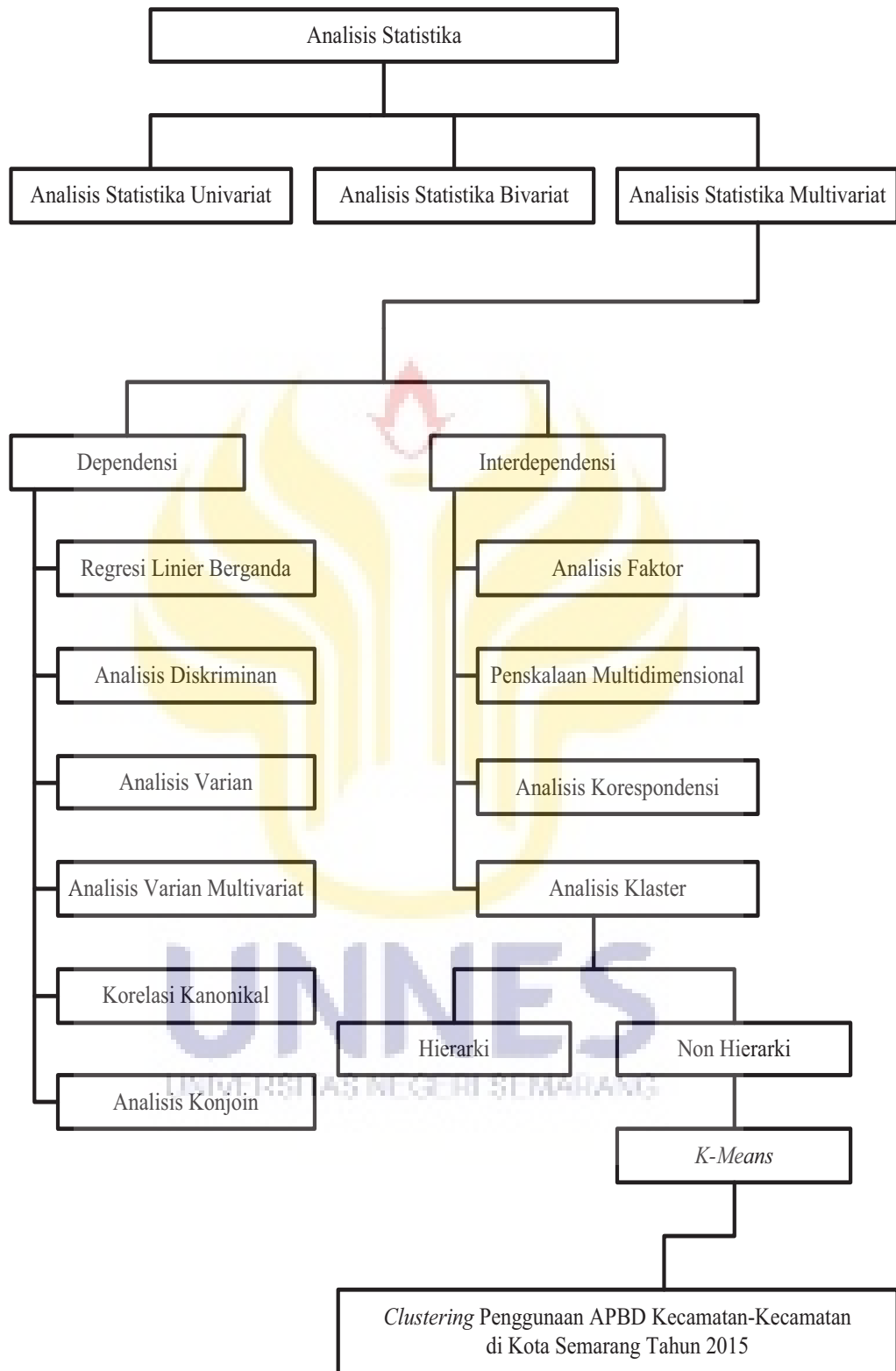
Pada metode hierarki pengelompokan dimulai dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses dilanjutkan ke-objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga kluster membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah kluster (Supranto, 2004).

Berbeda dengan metode hierarki, metode non hierarki ini dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan. Setelah jumlah kluster diketahui, baru proses kluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa di sebut sebagai *K-Means cluster* (Supranto, 2004).

Dalam penelitian ini digunakan metode *K-Means*. Metode *K-Means* merupakan salah satu metode yang di golongkan sebagai metode pengklasifikasian yang bersifat *unsupervised* (tanpa arahan) karena data yang di analisis tidak mempunyai label kelas yang berarti dalam proses pengelompokkannya, analisis ini

tidak mempunyai anggota klaster yang pasti. Melainkan data yang sudah masuk kedalam klaster bisa saja berpindah ke klaster yang lain.





Gambar 2.1 Kerangka Berpikir

## **BAB 5**

### **PENUTUP**

#### **5.1 Simpulan**

1. Penerapan metode K-Means dalam pembentukkan klaster berdasarkan penggunaan APBD mengelompokkan 16 kecamatan di Kota Semarang menjadi 3 kelompok dengan klaster 1 adalah kelompok kecamatan dengan penggunaan anggaran tinggi, klaster 2 adalah kelompok kecamatan dengan penggunaan anggaran rendah, dan klaster 3 adalah kelompok kecamatan dengan penggunaan anggaran sedang. Kecamatan dengan penggunaan anggaran tinggi adalah Semarang Barat, Semarang Tengah, Genuk, Mijen, Banyumanik, Pedurungan, Gunungpati, Ngaliyan, dan Tembalang. Kecamatan dengan penggunaan anggaran sedang adalah Semarang Selatan, Semarang Utara, dan Semarang Timur. Sedangkan kecamatan dengan penggunaan anggaran rendah adalah Gayamsari, Tugu, Gajah Mungkur, dan Candisari.
2. Karakteristik klaster-klaster yang terbentuk adalah sebagai berikut.  
Klaster 1 berisi kecamatan yang menggunakan anggaran untuk program pengembangan kinerja pengelolaan persampahan di bawah rata-rata, program pengendalian pencemaran dan perusakan lingkungan hidup di bawah rata-rata, program pelayanan administrasi perkantoran di atas rata-rata, program peningkatan sarana dan prasarana aparatur di atas rata-rata, program peningkatan pengembangan sistem pelaporan capaian kinerja dan



keuangan di atas rata-rata, program peningkatan dan pengembangan pengelolaan keuangan daerah di atas rata-rata, program peningkatan sistem pengawasan internal dan pengendalian pelaksanaan kebijakan kdh di atas rata-rata, program peningkatan keberdayaan masyarakat pedesaan di atas rata-rata, program peningkatan partisipasi masyarakat dalam membangun desa di atas rata-rata.

Klaster 2 berisi kecamatan yang menggunakan anggaran untuk program pengembangan kinerja pengelolaan persampahan di bawah rata-rata, program pengendalian pencemaran dan kerusakan lingkungan hidup di bawah rata-rata, program pelayanan administrasi perkantoran di atas rata-rata, program peningkatan sarana dan prasarana aparatur di bawah rata-rata, program peningkatan pengembangan sistem pelaporan capaian kinerja dan keuangan di bawah rata-rata, program peningkatan dan pengembangan pengelolaan keuangan daerah di bawah rata-rata, program peningkatan sistem pengawasan internal dan pengendalian pelaksanaan kebijakan kdh di bawah rata-rata, program peningkatan keberdayaan masyarakat pedesaan di bawah rata-rata, program peningkatan partisipasi masyarakat dalam membangun desa di bawah rata-rata.

Klaster 3 berisi kecamatan yang menggunakan anggaran untuk program pengembangan kinerja pengelolaan persampahan di atas rata-rata, program pengendalian pencemaran dan kerusakan lingkungan hidup di atas rata-rata, program pelayanan administrasi perkantoran di bawah rata-rata, program peningkatan sarana dan prasarana aparatur di bawah rata-rata, program

peningkatan pengembangan sistem pelaporan capaian kinerja dan keuangan di atas rata-rata, program peningkatan dan pengembangan pengelolaan keuangan daerah di bawah rata-rata, program peningkatan sistem pengawasan internal dan pengendalian pelaksanaan kebijakan kdh di bawah rata-rata, program peningkatan keberdayaan masyarakat pedesaan di bawah rata-rata, program peningkatan partisipasi masyarakat dalam membangun desa di bawah rata-rata.

3. Variabel yang memberikan perbedaan paling besar pada ketiga klaster yang terbentuk adalah variabel program peningkatan partisipasi masyarakat dalam membangun desa dengan nilai  $F$  30,136 dan signifikansi 0,000.

## 5.2 Saran

Dari kesimpulan diatas terlihat bahwa variabel yang memberikan perbedaan paling besar pada ketiga klaster adalah program peningkatan partisipasi masyarakat dalam membangun desa. Sebaiknya anggaran program peningkatan partisipasi masyarakat dalam membangun desa ditingkatkan sesuai proporsi kebutuhan masing-masing klaster dengan prioritas utama adalah kecamatan dengan penggunaan anggaran tinggi, yaitu Semarang Barat, Semarang Tengah, Genuk, Mijen, Banyumanik, Pedurungan, Gunungpati, Ngaliyan, dan Tembalang untuk meningkatkan partisipasi masyarakat dalam membangun desa dan meningkatkan pemberdayaan masyarakat di kecamatan tersebut.

## DAFTAR PUSTAKA

- , *Peraturan Menteri Dalam Negeri Nomor 13 Tahun 2006 tentang Pedoman Pengelolaan Keuangan Daerah.*
- , *Peraturan Menteri Dalam Negeri Nomor 25 Tahun 2009 tentang Pedoman Penyusunan Anggaran Pendapatan dan Belanja Daerah Tahun Anggaran 2010.*
- , *Peraturan Walikota Semarang Nomor 32 Tahun 2015 tentang Pedoman Pelaksanaan Rembug Warga dan Musyawarah Perencanaan Pembangunan Kelurahan dan Kecamatan Dalam Rangka Penyusunan Rencana Kerja Pemerintah Daerah Kota Semarang Tahun 2017.*
- , *Rencana Strategis Kecamatan Pedurungan Kota Semarang Tahun 2010-2015.* Semarang, 2010.
- Agusta, Yudi. 2007. *K-Means – Penerapan, Permasalahan dan Metode Terakait.* *Jurnal Sistem dan Informatika*, Vol. 3, 47-60.
- Dhuhita, W. M. P. 2015. *Clustering Menggunakan Metode K-Means untuk Menentukan Status Gizi Balita.* *Jurnal Informatika*, No. 2, Vol. 15, 160-174.
- Ghozali, Imam. 2011. *Aplikasi Analisis Multivariate dengan Program IBM SPSS 19.* Semarang: Undip.
- Hair, J.F., Anderson, dkk. 1998. *Multivariate Data Analysis, Fifth Edition,* Prentice-Hall International, Inc.
- Johnson, R.A., D.W. Wichern. 1982. *Applied Multivariate Statistical Analysis.* Prentice-Hall, Inc.
- Laeli, Sofya. 2014. *Analisis Cluster dengan Average Linkage Method dan Ward's Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link.* Skripsi. Yogyakarta: Universitas Negeri Semarang.

- Mahmudi. 2013. *Clustering Mahasiswa Matematika Berdasarkan Karakteristik Menggunakan Metode K-Means*. Tugas Akhir. Semarang: FMIPA Universitas Negeri Semarang.
- Metisen, Benri M. & Sari, Herlina L. 2015. *Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila*. *Jurnal Media Infotama*, No. 2, Vol. 11, 110-118.
- Nasari, Fina & Darma, Surya. 2015. *Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru*. Seminar Nasional Teknologi Informasi dan Multimedia 2015, STMIK AMIKOM Yogyakarta, 6-8 februari 2015.
- Ong, Johan Oscar. 2013. *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*. *Jurnal Ilmiah Teknik Industri*, No. 2, Vol. 12, 10-20.
- Santoso, Singgih. 2014. *Statistik Multivariat*. Jakarta: PT Elex Media Komputtindo.
- Sarwono, Jonathan. 2013. *Statistik Multivariat Aplikasi untuk Riset Skripsi*. Yogyakarta: C.V Andi Offset.
- Sukestiyarno. 2013. *Olah Data Penelitian Berbantuan SPSS*. Semarang: Unnes.
- Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Jakarta: PT Rineka Cipta.
- Widarjono, Agus. 2015. *Analisis Multivariat Terapan*. Yogyakarta: UPP STIM YKPN.