ANALISIS *MULTIVARIATE K-MEANS CLUSTER* UNTUK MENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI JAWA TENGAH BERDASARKAN IPM DAN KOMPONEN IPM TAHUN 2017

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

oleh:

Barbara Aurum Kusumawardani
4112315020

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG

2018

LEMBAR PENGESAHAN

Tugas Akhir yang berjudul

Analisis *Multivariate K-Means Cluster* untuk Mengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan IPM dan Komponen IPM Tahun 2017

disusun oleh

Nama : Barbara Aurum Kusumawardani

NIM : 4112315020

telah dipertahankan dihadapan sidang Panitia Ujian Tugas Akhir FMIPA Universitas Negeri Semarang pada tanggal 2 Agustus 2018.

Panitia:

Ketua

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

NIP. 196412231988031001

Sekretaris

Drs. Arief Agoestanto, M.Si

NIP. 196807221993031005

Pembimbing /Penguji II

Dra. Sunarmi M.Si

NIP. 195506241988032001

Penguji I

Dr. Scolastika Mariani M.Si

NIP. 196502101991022001

MOTTO

Kunci sebuah kesuksesan adalah jangan pernah berhenti,

Pun ketika keadaan tidak lagi seperti yang diinginkan, seperti layaknya masalah

Ketika jalan yang sedang kita tempuh terlihat sangat sulit

Ketika kita ingin tersenyum namun harus menyembunyikan kesedihan

Ketika semua keadaan membuat kita tertekan,

Istirahatlah jika memang perlu, namun jangan pernah berhenti.

Sukses memang berawal dari kegagalan yang pernah kita alami,

Memang kadang kita perlu merasakan keadaan paling rapuh diri kita,

untuk merasakan hal yang selalu kita impikan.

Takdir memang yang menggariskan kita dan keadaan

Namun usaha yang selalu kita lakukanlah yang akan menentukan takdir kita.

Jangan pernah berhenti bagaimanapun keadaan sekitar menekan,

Jangan pernah berhenti berdoa, berusaha, dan bermimpi.

Kesuksesan tidak akan datang jika kita hanya diam.

Kita tak akan pernah tau seberapa dekat kita dengan titik sukses kita.

Sukses mungkin telah sangat dekat dengan kita ketika terlihat jauh

Jangan sampai ada kata menyerah dalam perjalanan kita

Memang begitulah cara mendapat kebahagiaan,

Menikmati proses dari perjalanan

Dan jangan berhenti.

PERSEMBAHAN

Alhamdulillahirobilalamin, atas rahmat, nikmat dan karunia-Nya penulis dapat menyelesaikan tugas akhir yang berjudul "Analisis Multivariate *K-Means Cluster* untuk Mengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan IPM dan Komponen IPM Tahun 2017" ini dengan baik. Karena dalam penyusunan Tugas Akhir ini penulis banyak dibantu, dibimbing, dan didukung oleh berbagai pihak, maka Tugas Akhir ini penulis persembahkan untuk:

1. Allah Subhanahu Wa Ta'ala.
2. Kedua orang tua dan saudara penulis, yang telah memberi dukungan, semangat, dan motivasi dalam segala hal serta memberi kasih sayang yang teramat besar kepada penulis.
3. Keluarga Himatika dan keluarga organisasi/kepanitiaan event yang turut membersamai penulis.
4. Sahabat-sahabat penulis yang selalu sabar dengan segala keluh kesah penulis.
5. Teman-teman penulis yang telah memberi warna dan menemani perjalanan penulis.
6. Orang-orang terdekat penulis yang selalu ada serta tak henti-hentinya memberi semangat juga motivasi untuk penulis agar dapat menyelesaikan Tugas Akhir penulis.

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 Multivariate *K-Means Cluster* untuk Mengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan IPM dan Komponen IPM Tahun 2017". 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. Dra. Sunarmi, M.Si, Dosen Pembimbing 1 Tugas Akhir yang telah memberikan bimbingan selama penyusunan Tugas Akhir.
5. Perpustakaan Jurusan Matematika yang telah menyediakan fasilitas dan literatur untuk penyusunan Tugas Akhir ini.

6. Badan Pusat Statistika Provinsi Jawa Tengah yang telah menyediakan data untuk penyusunan Tugas Akhir ini.
7. Bapak, Ibu, dan adik tercinta serta keluarga yang telah memberikan motivasi dan dorongan semangat dalam mengerjakan Tugas Akhir ini.
8. Teman-teman Himatika angkatan 2015 yang selalu memberi dukungan semangat selama mengerjakan Tugas Akhir ini.
9. Sahabat-sahabat tercinta, teman-teman seperjuangan Staterkom 2015 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, 25 Juli 2018

Penulis

ABSTRAK

Kusumawardani, B.A. 2018. *Analisis Multivariate K-Means Cluster untuk Mengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan IPM dan Komponen IPM Tahun 2017*. Tugas Akhir. Prodi Statistika Terapan dan Komputasi, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, Pembimbing: Dra. Sunarmi, M.Si.

Kata kunci: *K-Means*, Analisis Klaster, IPM, Komponen IPM.

Tujuan yang ingin dicapai melalui penelitian ini adalah (1) menerapkan metode *K-Means* dalam pembentukkan klaster pada data IPM dan Komponen IPM di Kabupaten/Kota Provinsi Jawa Tengah, (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 BPS Provinsi Jawa Tengah. Populasi dan sampel dalam penelitian ini adalah IPM dan komponen pembangunnya pada 29 kabupaten dan 6 kota di Provinsi Jawa Tengah. Variabel dalam penelitian ini adalah Angka Harapan Hidup saat Lahir Jawa Tengah (X_1), Harapan Lama Sekolah Jawa Tengah (X_2), Pengeluaran per Kapita disesuaikan Jawa Tengah (X_3), dan Rata-rata Lama Sekolah Jawa Tengah (X_4). Analisis data yang digunakan adalah analisis klaster *K-Means* dengan software SPSS 16.0.

Kesimpulan yang diperoleh dari penelitian adalah *K-Means* mengelompokkan indeks pembangunan pada 29 kabupaten dan 6 kota menjadi 3 kelompok berdasarkan 4 komponen pembangun IPM. Klaster 1 adalah kelompok wilayah dengan IPM tinggi, yaitu Kota Magelang, Kota Salatiga, Kota Semarang, dan Kota Surakarta. Klaster 2 adalah kelompok wilayah dengan IPM rendah, yaitu Kabupaten Cilacap, Kabupaten Purbalingga, Kabupaten Banjarnegara, Kabupaten Kebumen, Kabupaten Wonosobo, Kabupaten Magelang, Kabupaten Grobogan, Kabupaten Blora, Kabupaten Rembang, Kabupaten Temanggung, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Pemalang, Kabupaten Tegal, dan Kabupaten Brebes. Klaster 3 adalah kelompok wilayah dengan IPM sedang, yaitu Kabupaten Banyumas, Kabupaten, Purworejo, Kabupaten Boyolali, Kabupaten Klaten, Kabupaten Sukoharjo, Kabupaten Wonogiri, Kabupaten Karanganyar, Kabupaten Sragen, Kabupaten Pati, Kabupaten Kudus, Kabupaten Jepara, Kabpaten Demak, Kabupaten Semarang, Kabupaten Kendal, Kota Pekalongan, dan Kota Tegal. Karakteristik klaster 1 berisi wilayah yang indeks pembangunan X_1, X_2, X_3, dan X_4 di atas rata-rata. Klaster 2 berisi wilayah yang indeks pembangunan X_1, X_2, X_3, dan X_4 di bawah rata-rata. Klaster 3 berisi wilayah yang indeks pembangunan X_1, X_2, X_3, dan X_4 di atas rata-rata. Variabel yang memberikan perbedaan paling besar pada ketiga klaster yang terbentuk adalah variabel Rata-Rata Lama Sekolah dalam membangun Indeks Pembangunan Manusia.

DAFTAR ISI

Halaman

HALAMAN JUDUL i
HALAMAN PENGESAHAN ii
MOTTO iii
PERSEMBAHAN iv
PRAKATA v
ABSTRAK vii
DAFTAR ISI viii
DAFTAR TABEL xi
DAFTAR GAMBAR xii
DAFTAR LAMPIRAN xiii
BAB I PENDAHULUAN
1. PENDAHULUAN 1
1.1 Latar Belakang Masalah 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 Tugas Akhir 6
2. TINJAUAN PUSTAKA 8
2.1 Analisis Statistika 8
2.2 Klasifikasi Data 8
2.3 Statistika Multivariat 10
2.3.1 Jenis-Jenis Analisis Statistika Multivariat 10
2.3.1.1 Metode Dependensi/Ketergantungan (*Dependence Methods*) 11
2.3.1.2 Metode Interependensi/Tidak Saling Ketergantungan (*Dependence Methods*) 13

2.3.2 Uji Data 15
2.3.2.1 *Missing* Data 15
2.3.2.2 Outlier 16
2.3.2.3 Uji Normalitas Data 18
2.3.2.4 Uji Linearitas Data 18
2.3.2.5 Uji Homoskedastisitas Data 18
2.4 Analisis Klaster 19
2.4.1 Proses Analisis Klaster 21
2.4.2 Metode Analisis Klaster 26
2.5 *K-Means Cluster* 31
2.6 Indeks Pembangunan Manusia (IPM) 33
2.6.1 Angka Harapan Hidup saat Lahir 33
2.6.2 Rata-Rata Lama Sekolah 34
2.6.3 Harapan Lama Sekolah 34
2.6.4 Pengeluaran per Kapita Disesuaikan 34
2.7 Gambaran Umum SPSS 35
2.8 Kerangka Berpikir 36
3. METODE PENELITIAN 42
3.1.Populasi dan Sampel 42
3.2.Variabel Penelitian 43
3.3.Pengumpulan Data 44
3.4.Analisis Data 44
4. METODE PENELITIAN 52
4.1.Hasil Penelitian 52
4.1.1 Uji Multikolinearitas 52
4.1.2 Analisis Deskriptif 53
4.1.3 Analisis Klaster 54
4.2.Pembahasan 60
4.2.1 Tafsiran Klaster 60
4.2.2 Perbedaan Variabel 63
4.2.3 Komposisi Klaster 64

5. PENUTUP 67
5.1.Simpulan 67
5.2.Saran 69

DAFTAR PUSTAKA 70
LAMPIRAN 72

DAFTAR TABEL

Halaman

Tabel 1 Hasil Uji Multikolinearitas 52
Tabel 2 Deskriptif 53
Tabel 3 *Initial Cluster Center* 55
Tabel 4 *Iteration History* 55
Tabel 5 *Final Cluster Centers* 56
Tabel 6 ANOVA 58
Tabel 7 Keanggotaan Klaster 60

DAFTAR GAMBAR

Halaman

Gambar 1 Kerangka Berpikir 41
Gambar 2 Data Uji Multikolinearitas 45
Gambar 3 Proses Uji Multikolinearitas 45
Gambar 4 Kotak Dialog Uji Multikolinearitas 46
Gambar 5 Kotak Dialog Variabel Uji Multikolinearitas *Linier Regression* 46
Gambar 6 Kotak Dialog Variabel Uji Multikolinearitas *Linier Regression: Statistics* 47
Gambar 7 Proses Analisis Deskriptif dan Standardisasi 47
Gambar 8 Kotak Dialog Utama *Descriptives* 48
Gambar 9 Kotak Dialog Proses *Descriptives* 48
Gambar 10 Proses *K-Means* Klaster 49
Gambar 11 Kotak Dialog Utama *K-Means Cluster Analysis* 49
Gambar 12 Kotak Dialog *K-Means Cluster: Save* 50
Gambar 13 Kotak Dialog *K-Means Cluster: Options* 51
Gambar 14 Komposisi dan Jarak Anggota Klaster 65

DAFTAR LAMPIRAN

Halaman

Lampiran 1 Data IPM dan Komponen Indeks Pembangunan Manusia di Provinsi Jawa Tengah tahun 2017 73

Lampiran 2 *Output* SPSS Uji Multikolinieritas 75

Lampiran 3 *Output* SPSS Deskriptif 76

Lampiran 4 *Output* SPSS Analisis Klaster 77

Lampiran 5 Zscore 80

Lampiran 6 QCL1 dan QCL2 82

BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Seiring dengan berkembangnya pengetahuan, ilmu statistika semakin banyak diterapkan dalam berbagai bidang dikehidupan manusia. Statistika, baik ilmu atau penerapannya telah banyak dijumpai pada masyarakat luas, di bidang ekonomi, sosial, pendidikan, kependudukan, dan berbagai bidang kehidupan lainnya. Lalu pada bidang sosial dan kependudukan misalnya, ilmu statistika sangat bermanfaat dan banyak digunakan sebagai tolak ukur untuk menganalisis suatu data yang diperoleh dari lapangan, juga dapat pula dipergunakan dalam mengambil suatu keputusan mengenai penelitian. Pada beberapa bidang tertentu, ilmu statistika juga digunakan untuk menghitung capaian pembangunan atau target keberhasilan pemerintah lainnya.

Selain itu ilmu statistika sering diterapkan pula pada data yang terdiri pada banyak variabel, yaitu untuk mengetahui hubungan atau keterkaitan antar variabel tersebut. Tidak sedikit data yang dapat terdiri dari banyak variabel, pada data bidang sosial dan kependudukan misalnya. Pada data kejadian ini analisis statistik yang biasa digunakan ialah analisis multivariat.

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 dikelompokan ke dalam satu klaster yang sama.

Pembangunan manusia didefinisikan sebagai proses perluasan pilihan bagi penduduk *(a process of enlarging the choices of people)*. Tujuan utama pembangunan adalah menciptakan lingkungan yang memungkinkan rakyat untuk menikmati umur pajang, sehat, dan menjalankan kehidupan yang produktif (*United Nation Development Programme* – UNDP).

Indeks Pembangunan Manusia (IPM) mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi dasar. Dimensi tersebut mencangkup umur panjang dan sehat; pengetahuan, dan kehidupan yang layak. Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Untuk mengukur dimensi kesehatan digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator rata-rata lama sekolah dan harapan lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok makanan dan bukan makanan, yang dilihat dari rata-rata besarnya pengeluaran per kapita sebagai pendekatan pendapatan yang mewakili capaian pembangunan untuk hidup layak.

Indeks Pembangunan Manusia (IPM) merupakan merupakan suatu indeks komposit yang mencangkup tiga bidang pembangunan manusia yang dianggap sangat mendasar, yaitu kemampuan untuk bertahan hidup lebih lama yang mewakili bidang kesehatan; angka melek huruf, partisipasi sekolah dan rata-rata lamanya sekolah mengukur kinerja pembangunan bidang pendidikan; dan kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita. Sehingga, untuk mengetahui pengelompokan IPM dapat digunakan analisis cluster (Indriani, 2013).

Dalam penelitian ini akan dianalisis kasus *clustering* Indeks Pembangunan Manusia dan Komponen IPM tahun 2017, untuk mengetahui pengelompokan kabupaten/ kota di Jawa Tengah berdasarkan Indeks Pembangunan Manusia dan Komponen IPM tahun 2017 pada daerah tersebut. Indeks Pembangunan Manusia memiliki beberapa komponen yang mempengaruhi nilai IPM sendiri. Beberapa komponen tersebut antara lain dapat di lihat dari Angka Harapan Hidup Saat Lahir Jawa Tengah Menurut Kabupaten/ Kota, Harapan Lama Sekolah Jawa Tengah Menurut Kabupaten/ Kota, Pengeluaran per Kapita Disesuaikan Jawa Tengah Menurut Kabupaten/ Kota, dan Rata-rata Lama Sekolah Menurut Kabupaten/ Kota. Berdasarkan Indeks Pembangunan Manusia dan Komponen IPM tersebut akan di lakukan *clustering* pada Kabupaten/ Kota di Jawa Tengah tahun 2017 dengan metode *K-Means* klaster.

Oleh karena itu penulis mengangkat judul tentang "Analisis *Multivariate K-Means Cluster* untuk Mengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan IPM dan Komponen IPM Tahun 2017".

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas dapat dirumuskan masalah sebagai berikut.

1. Bagaimana penerapan metode *K-Means* dalam pembentukkan klaster pada data IPM dan Komponen IPM kabupaten/kota di Provinsi Jawa Tengah ?
2. Bagaimanakah karakteristik klaster yang terbentuk pada tiap-tiap klaster ?
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 IPM dan komponen IPM 29 kabupaten dan 6 kota di Provinsi Jawa Tengah tahun 2017.
3. Analisis data dilakukan dengan data yang diperoleh dari IPM dan komponen yang terdapat dalam IPM di 29 kabupaten dan 6 kota di Provinsi Jawa Tengah Tahun 2017 yang diperoleh dari BPS Provinsi Jawa Tengah.
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 pembentukkan klaster pada data data IPM dan komponen IPM kabupaten/kota di Provinsi Jawa Tengah.
2. Mengetahui karakteristik klaster yang terbentuk.

3. Mengetahui variabel yang memberikan perbedaan paling besar pada klaster yang terbentuk.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut.

1 Bagi mahasiswa, menerapkan ilmu pengetahuan yang telah diperoleh selama di bangku perkuliahan dan memberi pengetahuan kepada mahasiswa mengenai penerapan ilmu statistika itu sangat luas.

2 Bagi Jurusan Matematika, sebagai bahan studi kasus bagi pembaca dan acuan bagi mahasiswa. Selain itu dapat pula dijadikan sebagai bahan referensi bagi pihak perpustakaan dan bahan bacaan yang dapat menambah ilmu pengetahuan bagi pembaca.

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 teoriteori 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 PENELITIAN 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 dikelompokan 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:

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. 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.

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 disetiap ruang kelas.

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 $0°C$, bukan berarti ruangan tersebut tidak memiliki temperatur sama sekali.

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 $0\ kg$ berarti tidak ada bendanya. Massa $10\ kg$ berarti 5 kali lipat dari massa $2\ kg$.

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). Analisis multivariat berasal dari kata *multi* dan *variate*, dua kata ini dapat diartikan analisis lebih dari dua variabel. Menurut Hair, sebagaimana dikutip oleh Utomo (2017), 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 Utomo (2017), analisis multivariat mencakup analisis data penelitian yang menggunakan banyak variabel yang dikenakan pengukuran secara bersamaan.

2.3.1 Jenis-Jenis Analisis Statistika Multivariat

Dalam analisis statistika multivariat terdapat dua jenis atau metode, yaitu metode dependensi dan metode interdependensi yang masing-masing metode memiliki beberapa tipe atau analisis. Berikut adalah jenis analisis statistika multivariate.

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).

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. 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.

3. Analisis Varian

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

4. Analisis Varian Multivariat

Analisis untuk mengetahui dampak beberapa variabel bebas yang nonmetrik (berupa kelompok) disebut perlakuan atau *treatment*, terhadap variabel tak bebas

yang metrik. Dimana variabel tak bebas yang terkena pengaruh (dampak) lebih dari satu.

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.

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 klaster, MDS atau CA (Supranto, 2004).

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. 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*).

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 prefensi (perasaan lebih suka) ke dalam jarak (*distance*) yang diwakili dalam ruang multidimensional.

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*.

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. 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

σ = 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 $+2{,}5$.

3. *Box Plot*

Box plot adalah sajian data yang menggambarkan hubungan antara median (Q_2), quartil 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

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).

Tujuan utama analisis klaster adalah mengelompokan 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 suatau 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.

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 klaster yang satu dengan yang lainnya (*between cluster*).

Menurut Santoso (2014), asumsi yang harus di penuhi pada analisis klaster 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).

Uji multikolinieritas bertujuan untuk menguji apakah ditemukan adanya korelasi antara variabel independen. Cara mengetahui adanya gejala multikolinieritas sebagai berikut.

1. Melihat nilai *tolerance*

   Tidak terjadi multikolinieritas, jika nilai *tolerance* lebih besar dari 0,10.

   Terjadi multikolinieritas, jika nilai *tolerance* lebih kecil atau sama dengan 0,10.
2. Melihat nilai VIF (*Variance Inflation Factor*)

   Tidak terjadi Multikolinieritas, jika nilai VIF lebih kecil 10.

   Terjadi Multikolinieritas, jika nilai VIF lebih besar atau sama dengan 10.

2.4.1 Proses Analisis Klaster

Proses *clustering* terhadap suatu data secara sekaligus diperlukan untuk mendapatkan data yang terkelompok atau terbagi menjadi beberapa berdasarkan tingkat kemiripannya. Adapun beberapa proses untuk melakukan analisis klaster.

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 klaster, 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

σ : Standar deviasi

Terkait dengan proses standarisasi data sebelumnya, yang mengacu pada angka *z*, yang memiliki ketentuan arti angka hasil output pada proses angka negatif berarti data di bawah rata-rata total sedangkan angka positif berarti data di atas rata-rata total.

(2.3)

$$X = \mu + z\sigma$$

di mana:

X = rata-rata sampel

μ = rata-rata populasi

z = nilai standarisasi yang dapat dilihat pada tabel 4.5

σ = standar deviasi

2. Mengukur Kesamaan Antar Objek (*Similarity*)

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

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 klaster 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 klaster adalah pada besarnya korelasinya, sehingga proses analisis klaster 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 kermiripan objek i terhadap j atau dinotasikan d_{ij} dapat dirumuskan sebagai berikut:

(2.4)

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

Dengan r_{ij} adalah koefisian 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:

(2.5)

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

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

2. Ukuran Jarak

Pengukuran ada bermacam-macam, yang paling populer adalah metode *Euclidean Distance*. Pada dasarnya, cara ini akan memasukkan sebuah data kedalam *cluster* tertentu dengan mengukur ‘jarak’ data tersebut dengan pusat

cluster. Jika data ada dalam jarak yang masih ada dalam batas tertentu, data tersebut dapat dimasukkan pada *cluster* tersebut (Santoso, 2014 : 128).

Merupakan ukuran yang paling sering digunakan. Ditetapkan untuk data berskala metrik. Sebenanrnya 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).

3. Ukuran Asosiasi

Pada dasarnya, cara ini akan mengasosiasikan sebuah data dengan cluster tertentu; dalam praktek, cara ini tidak sepopuler kedua cara sebelumnya (Santoso, 2014 : 128).

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 memutlakkan nilainya (Mahmudi, 2013).

3. Membuat Klaster (*clustering*)

Ada dua proses penting yaitu algoritma klaster dalam pembentukkan 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).

4. Interpretasi

Setelah klaster tebentuk, 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 klaster tersebut (Supranto, 2004).

5. Validasi dan *Profiling* Klaster

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

Tahap *profiling* klaster ini menggambarkan karakteristik tiap klaster untuk menjelaskan klaster-klaster tersebut dapat berbeda pada dimensi yang relevan. Titik

beratnya pada karakteristik yang secara signifikan berbeda antar klaster dan memprediksi anggota dalam suatu klaster khusus (Supranto, 2004).

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

2.4.2 Metode Analisis Klaster

Dua metode paling umum dalam algoritma klaster 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). Berikut adalah metode yang terdapat pada analisis klaster.

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.

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

(2.6)

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

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

(2.7)

$$_{(i,j)} = \text{maks}\{d_{ki}, d_{kj}\}$$

3. Pautan Rataan (*Average Linkage/Between Groups Method*)

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

(2.8)

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

Dengan n_i adalah jumlah item pada klaster i.

4. Metode *Centroid*

Pada metode ini, jarak antara dua klaster adalah jarak di antara dua *centroid* klaster-klaster tersebut. *Centroid* adalah rata-rata jarak yang ada pada sebuah klaster, yang didapat dengan melakukan rata-rata pada semua anggota suatu klaster tertentu. Dengan metode ini, setiap terjadi klaster baru, segera terjadi perhitungan ulang *centroid*, sampai terbentuk klaster yang tetap. Ukuran ketakmiripan yang digunakan adalah

(2.9)

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

5. *Ward's Error Sum of Square Method*

Metode *ward's* ditemukan oleh Ward (1963) tidak menghitung jarak antar klaster atau objek, namun metode ini membentuk klaster-klaster dengan memaksimalkan kehomogenan dalam klaster. Jumlah kuadrat dalam klaster digunakan sebagai ukuran kehomogenan. Metode *ward's* terus mencoba meminimalkan total jumlah kuadrat dalam klaster. Klaster-klaster dibentuk pada

masing-masing tahap seolah-olah telah menjadi solusi analisis klaster dan solusi terbaik adalah kombinasi klaster yang mempunyai jumlah kuadrat dalam klaster terkecil. Jumlah kuadrat dalam klaster (jarak kuadrat *euclidean* masing-masing objek terhadap *mean* klaster yang memuat objek klaster tesebut) yang meminimalkan sering disebut

(2.10)

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

Ukuran ketidakmiripan di atas dapat ditulis secara umum sebagai

(2.11)

$$_{(i,j)} = a_i d_{kj} + \alpha_j d_{kj} + \beta a_{ij} + d_3 d_{ki} - d_3 d_{kj}$$

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 rataan bila $a_i = \frac{n_i}{n_i+n_j}, a_i = \frac{n_j}{n_i+n_j}, \beta = -a_i a_j$ dan $d = 0$.

2. Metode Non-Hierarki

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

1. *Sequential Threshold*

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

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

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.

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*

Metode *K-Means* merupakan salah satu metode analisis klaster 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 klaster yang pasti. Melainkan data yang sudah masuk kedalam klaster bisa saja berpindah ke klaster yang lain. Akan tetapi, karena peneliti sering menentukan sendiri jumlah klaster 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 klaster yang diinginkan pada akhir suatu pengelompokkan. Metode klaster jenis ini di mungkinkan bahwa yang telah berada ke suatu klaster 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 rataan (pusat klaster terdekat) (Agusta, 2007).

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

1. Menentukan Jumlah Klaster

Tahap awal dalam analisis klaster dengan metode *K-Means* adalah dengan menentukan besarnya k, yaitu banyaknya klaster dan menentukan *centroid* atau pusat klaster pada tiap klaster. Dalam menentukan jumlah klaster dan *centroid* pada tiap klaster bersifat objektif.

2. Menghitung Jarak Antara Objek dan Tiap Centroid

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

3. Pengalokasian Data ke Centroid Terdekat

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

4. Validasi

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

2.6 Indeks Pembangunan Manusia (IPM)

Pembangunan manusia didefinisikan sebagai proses perluasan pilihan bagi penduduk *(a process of enlarging the choices of people)*. Tujuan utama pembangunan adalah menciptakan lingkungan yang memungkinkan rakyat untuk menikmati umur pajang, sehat, dan menjalankan kehidupan yang produktif (*United Nation Development Programme* – UNDP).

Indeks Pembangunan Manusia (IPM) mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi dasar. Dimensi tersebut mencakup umur panjang dan sehat; pengetahuan, dan kehidupan yang layak. Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Untuk mengukur dimensi kesehatan digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan dihunakan gabungan indicator rata-rata lama sekolah dan harapan lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indicator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok makanan dan bukan makanan, yang dilihat dari rata-rata besarnya pengeluaran per kapita sebagai pendekatan pendapatan yang mewakili capaian pembangunan untuk hidup layak.

2.6.1 Angka Harapan Hidup Saat Lahir

Angka Harapan Hidup saat lahir didefinisikan sebagai rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir.

2.6.2 Rata-rata Lama Sekolah

Rata-rata Lama Sekolah didefinisikan sebagai jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal.

Cakupan pendududk dihitung RLS adalah penduduk berusia 25 tahun keatas. RLS dihitung untuk usia 25 tahun keatas dengan asumsi pada umur 25 tahun proses pendidikan sudah berakhir. Perhitunagn RLS pada usia 25 tahun ke atas juga mengikuti standard internasional oleh UNDP.

2.6.3 Harapan Lama Sekolah

Angka Harapan Lama Sekolah didefinisikan sebagai lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang.

HLS dapat digunakan untuk mengetahui kondisi pembangunan sistem pendidikan di berbagai jenjang. HLS dihitung pada usia 7 tahun ke atas karena mengikuti kebijakan pemerintah yaitu program wajib belajar. Untuk Mengakomodir penduduk yang tidak tercakup dalam Susenas, HLS dikoreksi dengan siswa yang bersekolah di pesantren yaitu dari Direktorat Pendidikan Islam.

2.6.4 Pengeluaran per Kapita Disesuaikan

Pengeluaran per Kapita disesuaikan ditentukan dari nilai pengeluaran per kapita dan paritas daya beli. Rta-rata pengeluaran per kapita setahun diperoleh dari Susenas Modul, dihitung dari level kab/kota. Rata-rata pengeluaran per kapita dibuat konstan/riil dengan tahun dasar 2012=100.

Perhitungan paritas daya beli pada metode baru menggunakan 96 komoditas dimana 66 kooditas merupakan komoditas makanan dan sisanya merupakan komoditas non makanan. Metode perhitungannya menggunakan metode Rao.

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 popular 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 mebandingkan 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

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

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 komulatif, 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 Windha Mega Pradnya Dhuhita 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.

Tahun 2017 Cahyo Aziz Utomo melakukan penelitian mengenai *clustering* menggunakan *K-Means* untuk mengelompokkan APBD kecamatan-kecamatan di Kota Semarang tahun 2015. Penelitiannya 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, klaster 2 adalah kelompok kecamatan dengan penggunaan anggaran rendah, dan klaster 3 adalah kelompok kecamatan dengan penggunaan anggaran sedang.

Berdasarkan jumlah variabel yang dianalisis, analisis statistika dapat dikelompokan 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 klaster, MDS atau CA (Spuranto, 2004).

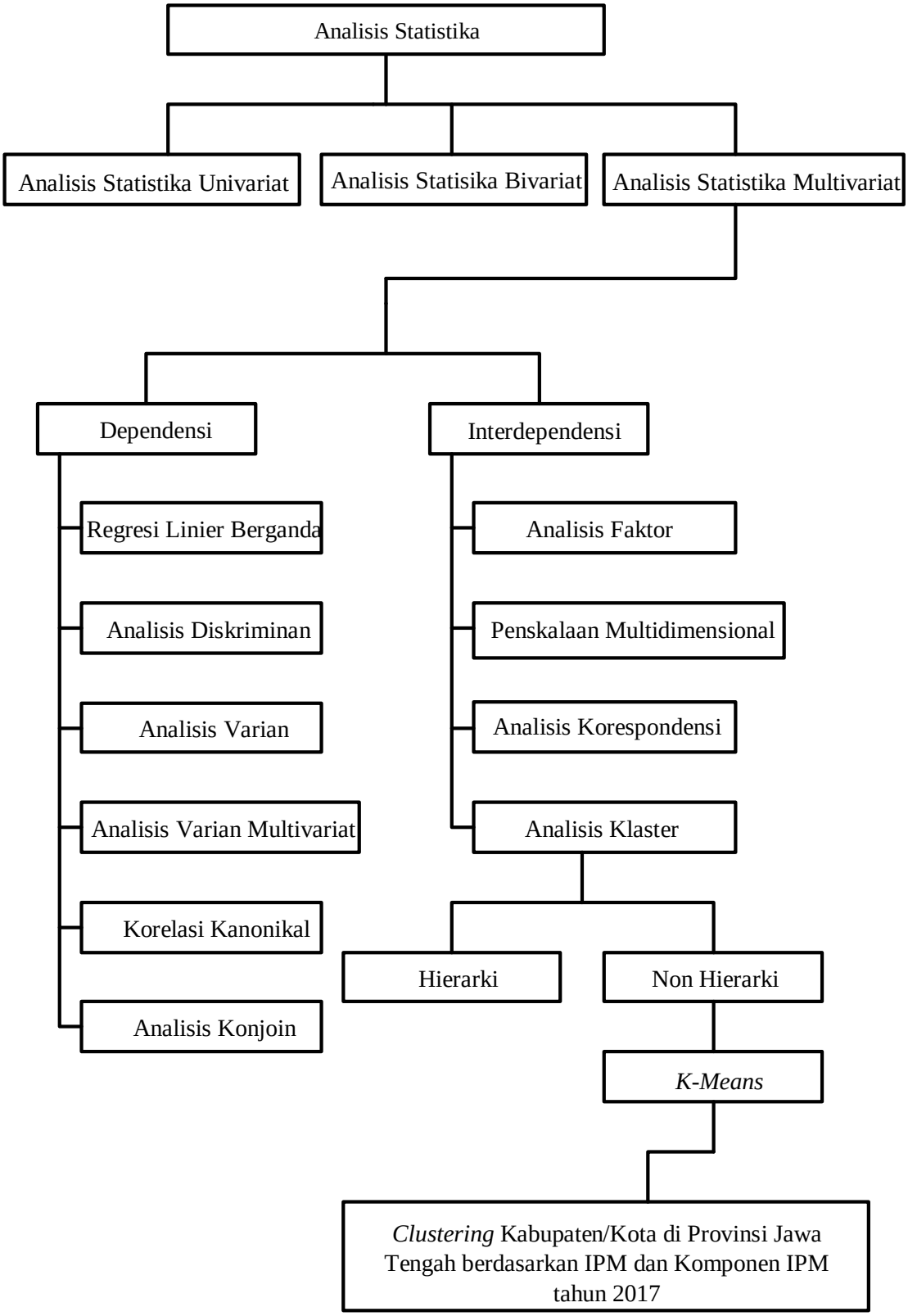
Analisis klaster, yaitu analisis untuk mengelompokkan elemen yang mirip sebagai objek penelitian menjadi kelompok (*cluster*) yang berbeda dan *mutually exclusive*. Tujuan utama analisis klaster adalah mengelompokan objek-objek berdasarkan kesamaan karakteristik diantara objek-objek tersebut. Objek bisa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

Berbeda dengan metode hierarki, metode non hierarki ini dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan. Setelah jumlah klaster

diketahui, baru proses klaster 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 dianalisis 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 1 Kerangka Berpikir

BAB 5

PENUTUP

5.1 Simpulan

1. Penerapan metode K-Means dalam pembentukkan klaster berdasarkan Indeks Pembangunan Manusia dan Komponen Indeks Pembangunan Manusia untuk mengelompokkan 29 Kabupaten dan 6 Kota di Provinsi Jawa Tengah menjadi 3 kelompok yaitu pertama dengan melakukan uji asumsi pada data tersebut menggunakan SPSS 16.0 yang jika uji asumsi terpenuhi maka dapat melanjutkan melakukan analisis K-Means Cluster. Setelah dilakukan uji asumsi pada data tersebut, uji asumsi terpenuhi. Maka setelah itu dilanjutkan dengan melakukan analisis K-Means Cluster yang menghasilkan, klaster 1 kelompok wilayah dengan Indeks Pembangunan Manusia tinggi, klaster 2 adalah kelompok wilayah dengan Indeks Pembangunan Manusia rendah, dan klaster 3 adalah kelompok wilayah dengan Indeks Pembangunan Manusia sedang. Wilayah dengan Indeks Pembangunan Manusia tinggi adalah Kota Magelang, Kota Salatiga, Kota Semarang, dan Kota Surakarta.

   Wilayah dengan Indeks Pembangunan Manusia sedang adalah Kabupaten Banyumas, Kabupaten, Purworejo, Kabupaten Boyolali, Kabupaten Klaten, Kabupaten Sukoharjo, Kabupaten Wonogiri, Kabupaten

Karanganyar, Kabupaten Sragen, Kabupaten Pati, Kabupaten Kudus, Kabupaten Jepara, Kabpaten Demak, Kabupaten Semarang, Kabupaten Kendal, Kota Pekalongan, dan Kota Tegal.

Sedangkan wilayah dengan Indeks Pembangunan Manusia rendah adalah Kabupaten Cilacap, Kabupaten Purbalingga, Kabupaten Banjarnegara, Kabupaten Kebumen, Kabupaten Wonosobo, Kabupaten Magelang, Kabupaten Grobogan, Kabupaten Blora, Kabupaten Rembang, Kabupaten Temanggung, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Pemalang, Kabupaten Tegal, dan Kabupaten Brebes.

2. Karakteristik klaster-klaster yang terbentuk adalah sebagai berikut. Klaster 1 berisi wilayah yang memiliki angka harapan hidup saat lahir di atas rata-rata, harapan lama sekolah di atas rata-rata, pengeluaran per kapita disesuaikan di atas rata-rata, dan rata-rata lama sekolah di atas rata-rata.

   Klaster 2 berisi wilayah yang memiliki angka harapan hidup saat lahir di bawah rata-rata, harapan lama sekolah di bawah rata-rata, pengeluaran per kapita disesuaikan di bawah rata-rata, dan rata-rata lama sekolah di bawah rata-rata.

   Klaster 3 berisi wilayah yang memiliki angka harapan hidup saat lahir di atas rata-rata, harapan lama sekolah di atas rata-rata, pengeluaran per kapita disesuaikan di atas rata-rata, dan rata-rata lama sekolah di atas rata-rata.

3. Variabel yang memberikan perbedaan paling besar pada ketiga klaster yang terbentuk adalah variabel Rata-Rata Lama Sekolah dalam membangun Indeks Pembangunan Manusia dengan nilai F 82,241 dan signifikansi 0,000.

5.2 Saran

Dari kesimpulan diatas terlihat bahwa variabel yang memberikan perbedaan paling besar pada ketiga klaster adalah rata-rata lama sekolah. Sebaiknya indeks pembangunan pada rata-rata lama sekolah ditingkatkan sesuai standart terhadap seluruh klaster dengan prioritas utama adalah wilayah dengan indeks pembangunan manusia rendah, yaitu Kabupaten Cilacap, Kabupaten Purbalingga, Kabupaten Banjarnegara, Kabupaten Kebumen, Kabupaten Wonosobo, Kabupaten Magelang, Kabupaten Grobogan, Kabupaten Blora, Kabupaten Rembang, Kabupaten Temanggung, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Pemalang, Kabupaten Tegal, dan Kabupaten Brebes untuk meningkatkan partisipasi masyarakat dalam membangun indeks pembangunan manusia tersebut.

DAFTAR PUSTAKA

Badan Pusat Statistik Provinsi Jawa Tengah. 2017. *Indeks Pembangunan Manusia*. Semarang: Badan Pusat Statistik.

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.

Indriani, Y.A. 2013. *Perbandingan Jumlah Kelompok Optimal pada Metode Single Lingkage dan Complete Linkage dengan Indeks Validitas Silhoutte*. Malang: Universitas Negeri Malang.

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 Pengelompokkan 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.

Utomo, C.A. 20017. *Analisis Multivariat Clustering K-Means untuk APBD Kecamatan-Kecamatan di Kota Semarang Tahun 2015*. Tugas Akhir. Semarang: FMIPA Universitas Negeri Semarang.

Widarjono, Agus. 2015. *Analisis Multivariat Terapan*. Yogyakarta: UPP STIM YKPN.