

*ANALISIS HYBRID MUTUAL CLUSTERING
MENGUNAKAN JARAK SQUARE EUCLIDEAN*

Skripsi

Disusun untuk melengkapi syarat-syarat
guna memperoleh gelar Sarjana Sains



ASTRID ALFIRA

3125121985

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA

2017

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

ANALISIS *HYBRID MUTUAL CLUSTERING* MENGUNAKAN JARAK *SQUARE EUCLIDEAN*

Nama : Astrid Alfira

No. Registrasi : 3125121985

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: Prof. Dr. Suyono, M.Si. NIP. 19671218 199303 1 005
Wakil Penanggung Jawab			
Wakil Dekan I	: Dr. Muktiningsih N., M.Si. NIP. 19640511 198903 2 001
Ketua	: Dr. Lukita Ambarwati, S.Pd., M.Si. NIP. 19721026 200112 2 001
Sekretaris	: Med Irzal, M.Kom. NIP. 19770615 200312 1 001
Penguji	: Dra. Widyanti Rahayu, M.Si. NIP. 19661103 200112 2 001
Pembimbing I	: Ir. Fariani Hermin, MT. NIP. 19600211 198703 2 001
Pembimbing II	: Dr. Eti Dwi Wiraningsih, S.Pd., M.Si. NIP. 19810203 200604 2 001

Dinyatakan lulus ujian skripsi tanggal: 14 Februari 2017

ABSTRACT

ASTRID ALFIRA, 3125121985. Hybrid Mutual Clustering Analysis Using Square Euclidean Distance. Thesis. Faculty of Mathematics and Natural Science Jakarta State University. 2017.

Cluster analysis is use to agglomerate objects based on similarity measure, which the based concept of cluster analysis is distance and similarity measuring. Objects agglomeration inside the analysis cluster can be done with bottom-up method, top-down method, and Hybrid Mutual Clustering. Objects agglomeration with bottom-up using clustering method which start from a little cluster to a bigger cluster, objects agglomeration with top-down using the opposite method which start from breaking a big cluster to a smaller cluster. Hybrid Mutual Clustering method introduced on 2006 by Hugh Chipman and Robert Tibshirani, this method is combining the advantages of bottom-up and top-down method. Hybrid Mutual Clustering method in this thesis is using Square Euclidean distance as a method of calculation the distance of an object to other objects. Best cluster is chosen by proportion value on variable and characteristic differences between variable. In this thesis the biggest proportion gets from age under 14 variable at Nusa Tenggara Barat and Papua city.

Keywords : *Cluster, Hybrid Mutual Clustering, Bottom-Up, Top-Down, Square Euclidean distance.*

ABSTRAK

ASTRID ALFIRA, 3125121985. Analisis Hybrid Mutual Clustering menggunakan Jarak Square Euclidean. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2017.

Analisis kelompok berguna untuk mengelompokkan objek berdasarkan ukuran kemiripan, dimana konsep dasar dari analisis kelompok adalah pengukuran jarak dan kesamaan. Pengelompokan objek di dalam analisis kelompok dapat dilakukan dengan metode *bottom-up*, *top-down*, dan *Hybrid Mutual Clustering*. Pengelompokan objek dengan *bottom-up* menggunakan metode pengelompokan yang dimulai dari kelompok kecil menjadi kelompok yang lebih besar, pengelompokan objek dengan *top-down* menggunakan metode sebaliknya yaitu pengelompokan dengan memecah kelompok besar menjadi kelompok yang lebih kecil. Metode *Hybrid Mutual Clustering* baru diperkenalkan pada tahun 2006 oleh Hugh Chipman dan Robert Tibshirani, dimana metode ini mengkombinasikan kelebihan metode *bottom-up* dan *top-down*. Metode *Hybrid Mutual Clustering* yang digunakan dalam skripsi ini adalah metode pengelompokan *hybrid* menggunakan jarak *Square Euclidean* sebagai metode perhitungan jarak objek satu ke objek lainnya. Pemilihan kelompok terbaik dipilih berdasarkan nilai proporsi terbesar pada variabel dan perbedaan karakteristik antar variabel. Pada skripsi ini proporsi terbesar didapat dari variabel umur di bawah 14 tahun di kota Nusa Tenggara Barat dan Papua.

Kata kunci : Klaster, Hybrid Mutual Clustering, Bottom-Up, Top-Down, Jarak Square Euclidean.

PERSEMBAHANKU...

" You're future is created by what you do today, not tommorrow."

" If people don't believe that mathematics is simple, it's only because they don't realize how complicated life is. "

-John von Neumann

Skripsi ini kupersembahkan untuk Mamah dan Papah. *"Terima kasih atas dukungan, do'a, serta kasih sayang kalian".*

KATA PENGANTAR

Puji syukur kepada kehadirat Allah SWT atas pengetahuan dan kemampuan sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis *Hybrid Hierarchical Clustering* menggunakan Jarak *Square Euclidean*" yang merupakan salah satu syarat dalam memperoleh gelar Sarjana Jurusan Matematika Universitas Negeri Jakarta.

Skripsi ini berhasil diselesaikan tidak terlepas dari adanya bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan terima kasih terutama kepada:

1. Ibu Ir. Fariani Hermin, MT. selaku Dosen Pembimbing I dan Ibu Dr. Eti Dwi Wiraningsih, S.Pd., M.Si. selaku Dosen Pembimbing II, yang telah meluangkan waktunya dalam memberikan bimbingan, saran, nasehat serta arahan sehingga skripsi ini dapat selesai dan menjadi lebih baik dan terarah.
2. Ibu Ratna Widyati, S.Si, M.Kom., selaku Pembimbing Akademik dan Ibu Lukita Ambarwati, S.Pd., M.Si., selaku Koordinator Program Studi Matematika FMIPA UNJ. Terima kasih atas segala bantuan dan kerja sama Ibu selama perkuliahan dan pengerjaan skripsi ini.
3. Seluruh Bapak/Ibu dosen Matematika FMIPA UNJ. Terima kasih atas pengajaran dan kerja sama yang telah diberikan.
4. Mamah, Papah, Mba Tita, Kak Buce, dan Gendhis tersayang. Terima kasih atas segala bentuk dukungan, doa, motivasi, perhatian, dan kasih sayang kepada dede selama ini. Maaf telat satu semester, *but finally I made it :')*

5. Kaniggia Augusta Syahril, *partner in goods and crimes*. Terima kasih banyak atas segalanya. Maaf acit selalu ngerepotin. Semoga kamu cepat kelarin kuliah kamu, supaya kita bisa wujudin cita-cita kita bareng.
6. Teman Kesayangan penulis, Anggi, Vinna, Mega, Yuli, Ibeth, Jen, Dedy, Atun, khususnya Faralita yang telah banyak membantu penulis menyelesaikan skripsi ini. Terima kasih telah menjadi dosen pembimbing ketiga. Semoga kita tetap menjadi kesayangan.
7. Cindy Almas dan Urie Ayam. Terima kasih banyak untuk segala bentuk dukungan kalian. Akhirnya kelar juga nih jadi mahasiswa!
8. Teman-teman di Markas G7: Adit, Nico, Pitut, Gina, Fahrian, Fathi, Faiz, Jodi, Bom-bom, Tante Na, dan lain-lain. Terima kasih selalu memberikan hiburan dikala Acit pusing mengerjakan skripsi ini. Abis ini kita *party* ya!:p
9. Teman-teman di Prodi Matematika UNJ 2012. Terima kasih kebersamaan selama 4,5 tahunnya. Semoga kita tetap terus bisa bersilaturahmi.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Masukan dan kritikan akan sangat berarti. Semoga skripsi ini dapat bermanfaat bagi pembaca sekalian.

Jakarta, Februari 2017

Astrid Alfira

DAFTAR ISI

ABSTRACT	i
ABSTRAK	ii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	3
1.3 Pembatasan Masalah	3
1.4 Tujuan Penulisan	4
1.5 Manfaat Penulisan	4
1.6 Metode Penelitian	4
II LANDASAN TEORI	5
2.1 Analisis Kelompok	5
2.2 Analisis Korelasi	7
2.3 Analisis Komponen Utama	8
2.4 Mutual Cluster	13
2.5 Metode Pengelompokan Objek	14
2.5.1 Metode Pengelompokan Hirarki	14
2.5.2 Metode Pengelompokan non-Hirarki	16

2.6	Jarak <i>Square Euclidean</i>	17
2.7	Pengelompokan Metode <i>Bottom-Up</i>	18
2.8	Pengelompokan Metode <i>Top-Down</i>	21
III PEMBAHASAN		26
3.1	Pengelompokan Menggunakan Metode <i>Hybrid Mutual Clustering</i> dengan Jarak <i>Square Euclidean</i>	26
3.2	Penerapan Metode <i>Hybrid Mutual Clustering</i> dengan Jarak <i>Square Euclidean</i> ke dalam contoh kasus	29
IV PENUTUP		40
4.1	Kesimpulan	40
4.2	Saran	41
DAFTAR PUSTAKA		43
LAMPIRAN-LAMPIRAN		45

DAFTAR TABEL

2.1	Hasil nilai Tes A dan Tes B	11
2.2	Hasil nilai Tes A dan Tes B	12
2.3	Matriks Jarak Hasil Nilai Renang Mahasiswa	19
2.4	Matriks Jarak Hasil Nilai Renang Mahasiswa	20
2.5	Matriks Setelah Penggabungan M1 dan M3	20
2.6	Matriks Setelah Penggabungan M1/M3 dan M2/M4	20
2.7	Matriks Setelah Penggabungan M1/M3/M2/M4	21
2.8	Data Produksi dan Luas Lahan di Provinsi Jawa Timur Tahun 2013	23
2.9	Data Hasil Akhir pengelompokan Menggunakan <i>K-means Cluster</i>	23
2.10	Data Hasil Pengelompokan Kota dengan Menggunakan <i>K-means Cluster</i>	24

DAFTAR GAMBAR

2.1	Diagram alir pengelompokan metode <i>bottom-up (agglomerative)</i> .	19
2.2	Dendogram Data Hasil Nilai Mahasiswa.	21
2.3	Diagram alir pengelompokan metode <i>top-down (k-means)</i>	22
2.4	Dendogram Data Hasil Pengelompokan Kota.	25
3.1	Diagram alir pengelompokan menggunakan metode <i>Hybrid Mutual Clustering</i> dengan jarak <i>Square Euclidean</i>	27
3.2	Tabel <i>Descriptive Z-Score</i>	30
3.3	Tabel <i>Descriptive Statistics</i>	31
3.4	Tabel <i>Case Processing Summary</i>	31
3.5	Tabel <i>Agglomeration Schedule</i>	32
3.6	Tabel <i>Cluster Membership</i>	34
3.7	Dendogram Analisis Klaster Hierarki	37
3.8	Tabel <i>Initial Cluster Centers</i>	38
3.9	Tabel <i>Iteration History</i>	38
3.10	Tabel <i>Final Cluster Centers</i>	38
4.1	Tabel Data Persentase Penduduk Buta Aksara menurut Provinsi di Indonesia Tahun 2015 (sumber: Badan Pusat Statistik Tahun 2015).	45
4.2	Tabel Matriks Prioritas Hasil <i>Output SPSS</i>	46
4.3	Tabel Matriks Prioritas Hasil <i>Output SPSS</i> (lanjutan).	47
4.4	Tabel QCL Hasil <i>Output SPSS</i>	48

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Cluster atau 'klaster' dapat diartikan kelompok; dengan demikian, pada dasarnya analisis klaster akan menghasilkan sejumlah klaster (kelompok). Analisis ini diawali dengan pemahaman bahwa sejumlah data tertentu sebenarnya mempunyai kemiripan diantara anggotanya. Karena itu, dimungkinkan untuk mengelompokkan anggota-anggota yang 'mirip' atau mempunyai karakteristik yang serupa tersebut dalam satu atau lebih dari satu klaster.

Salah satu cara kerja dan tujuan analisis klaster adalah mengelompokkan objek-objek berdasarkan persamaan karakteristik di antara objek-objek tersebut. Objek yang akan diklaster bisa berupa produk (barang dan jasa), makhluk hidup (tumbuhan dan binatang), atau manusia (disebut responden, konsumen, partisipan dalam kegiatan eksperimen, atau yang lain). Objek tersebut akan diklasifikasikan ke dalam satu atau lebih klaster (kelompok) sehingga objek-objek yang berada dalam satu klaster akan mempunyai kemiripan satu sama lain.

Dalam praktek, analisis klaster bisa diterapkan pada banyak bidang ilmu; seperti psikologi, biologi, kedokteran, dan manajemen. Berikut contoh kegunaan analisis klaster pada bidang ilmu manajemen pemasaran. Sebuah perusahaan *real estate* ingin mengetahui perilaku konsumen yang membeli rumah di lingkungan tertentu. Untuk itu, sekian responden diminta memberi pendapat mereka tentang berbagai variabel pembelian sebuah rumah, seperti lingkung-

an yang aman, lingkungan yang nyaman dan bebas banjir, fasilitas umum dan sosial yang memadai, letak perumahan yang strategis, harga yang terjangkau, sistem pembayaran yang fleksibel dan sebagainya. Dengan analisis klaster diketahui bahwa ada tiga kelompok konsumen pembeli rumah tersebut, yakni klaster yang memperhatikan lingkungan, klaster yang mengutamakan sistem kredit dan klaster yang mengutamakan mutu rumah tersebut.

Analisis kelompok berguna untuk mengelompokkan objek berdasarkan ukuran kemiripan, dimana konsep dasar dari analisis kelompok adalah pengukuran jarak dan kesamaan. Pengelompokan objek di dalam analisis kelompok dapat dilakukan dengan metode *bottom-up*, *top-down*, dan *Hybrid Mutual Clustering*. Pengelompokan objek dengan *bottom-up* menggunakan metode pengelompokan yang dimulai dari kelompok kecil menjadi kelompok yang lebih besar, pengelompokan objek dengan *top-down* menggunakan metode sebaliknya yaitu pengelompokan dengan memecah kelompok besar menjadi kelompok yang lebih kecil. Metode *Hybrid Mutual Clustering* baru diperkenalkan pada tahun 2006 oleh Hugh Chipman dan Robert Tibshirani, dimana metode ini mengkombinasikan kelebihan metode *bottom-up* dan *top-down*. Algoritma *bottom-up* baik dalam mengelompokkan ukuran sampel kecil dan sebaliknya, algoritma *top-down* baik dalam mengelompokkan ukuran sampel besar. Metode *Hybrid Mutual Clustering* yang digunakan adalah metode pengelompokan *hybrid* melalui *mutual cluster*. *Mutual cluster* adalah pengelompokan yang menggunakan jarak terbesar antara pasangan dalam kelompok yang lebih kecil dari jarak terpendek ke setiap titik di luar kelompok.

Penelitian tentang *Hybrid Mutual Clustering* ini sebelumnya pernah dilakukan oleh Agustina (2013) dengan judul "Pemilihan Metode Pengelompokan Terbaik Kabupaten/Kota Berdasarkan Indikator Pendidikan Menggunakan *Hybrid Melalui Mutual Cluster, Bottom-Up dan Top-Down*". Dalam penu-

lisannya, penulis membandingkan antara 3 metode yaitu *Hybrid*, *Bottom-up*, dan *Top-down*. Selain itu, dilakukan penelitian oleh Madani (2014) dengan judul Analisis *Hybrid Mutual Clustering* Melalui *Mutual Cluster*, *Bottom-Up* dan *Top-Down* Menggunakan Jarak *Euclidean* dan *Mahalanobis*. Dalam penulisannya, penulis juga membandingkan ketiga metode serta menggunakan jarak *Euclidean* dan *Mahalanobis*.

Dalam skripsi ini, metode yang dipakai menggunakan jarak *Square Euclidean* yang merupakan pengembangan dari jarak *Euclidean*. Sebagaimana namanya, *Square Euclidean* adalah ukuran jarak dengan mengkuadratkan selisih antara dua objek yang sama pada kelompok yang berbeda. Relatif untuk beberapa persoalan terutama menyangkut persoalan lokasi objek diselesaikan dengan penerapan *Square Euclidean*.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, permasalahan yang akan dikaji pada skripsi ini yaitu bagaimana cara pengelompokan menggunakan metode *Hybrid Mutual Clustering* dengan jarak *Square Euclidean*.

1.3 Pembatasan Masalah

Pembatasan ruang lingkup permasalahan yang dibahas dalam skripsi ini diperlukan untuk membuat skripsi ini memiliki arah yang tepat dan selaras, berikut ini merupakan pembatasan masalah yang dibuat antara lain:

1. Pemilihan kelompok terbaik menggunakan nilai proporsi terbesar.
2. Metode *Hybrid Mutual Clustering* menggunakan penggabungan antara metode *Bottom-Up* dan *Top-Down*.

3. Data pada contoh kasus diasumsikan berdistribusi normal.

1.4 Tujuan Penulisan

Berdasarkan rumusan masalah dan pembatasan masalah yang dijelaskan sebelumnya, maka didapat tujuan dari penulisan skripsi ini yaitu untuk mengetahui cara pengelompokan menggunakan analisis *Hybrid Mutual Clustering* dengan jarak *Square Euclidean*.

1.5 Manfaat Penulisan

Adapun manfaat dari penulisan ini yaitu:

1. Sumber pengetahuan mengenai dunia matematika di bidang statistika khususnya *clustering*.
2. Alternatif sumber bacaan dalam pengembangan suatu metode generalisasi secara induktif, yaitu pengambilan kesimpulan secara umum dengan berdasarkan fakta-fakta khusus.
3. Pengetahuan dalam mendeskripsikan sifat-sifat atau karakteristik dalam masing-masing kelompok klaster.

1.6 Metode Penelitian

Skripsi ini merupakan kajian teori dalam bidang analisis peubah ganda dan analisis runtun waktu yang didasarkan pada buku-buku dan jurnal-jurnal tentang teori permasalahan di bidang statistik. Referensi utama yang digunakan yaitu Santoso (2010), Johnson & Wichern (2002).

BAB II

LANDASAN TEORI

Pada bab ini akan dibahas teori-teori mengenai analisis klaster (kelompok), analisis korelasi, analisis komponen utama, lalu akan dijelaskan juga mengenai jarak *Square Euclidean*, pengelompokan menggunakan metode *bottom-up* dan *top-down*, serta metode *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean*. Sebagai awalan, akan dijelaskan mengenai analisis kelompok.

2.1 Analisis Kelompok

Analisis kelompok merupakan suatu analisis multivariat yang digunakan untuk mengelompokkan objek pengamatan menjadi beberapa kelompok berdasarkan ukuran kemiripan antar objek, sehingga objek-objek yang berada dalam satu kelompok memiliki kemiripan yang lebih homogen dibandingkan objek dari kelompok yang berbeda (Johnson & Wichern, 2002). Seperti diketahui, analisis kelompok akan membagi sejumlah data pada satu atau beberapa *cluster* tertentu. Sebuah *cluster* yang baik adalah *cluster* yang mempunyai ciri sebagai berikut:

1. Homogenitas (kesamaan) yang tinggi antara anggota dalam satu *cluster* atau biasa disebut *within cluster*. Sebagai contoh, konsumen restoran yang mengutamakan interior ruangan tentu terdiri dari orang-orang yang mengutamakan kebersihan dan kenyamanan restoran. Mereka yang mengutamakan harga makanan yang murah tidak dapat digabungkan menjadi 'anggota' *cluster* tersebut.

2. Heterogenitas (perbedaan) yang tinggi antara *cluster* satu dengan *cluster* yang lain atau biasa disebut *between cluster*. Dalam contoh sebelumnya, anggota dari *cluster* konsumen restoran yang mengutamakan interior ruangan tentu mempunyai pendapat yang jelas berbeda dengan anggota *cluster* konsumen yang mengutamakan harga makanan yang murah.

Dari dua hal di atas dapat disimpulkan bahwa ciri sebuah *cluster* yang baik adalah *cluster* yang mempunyai anggota yang semirip mungkin satu sama lain, tetapi sangat tidak mirip dengan anggota *cluster* yang lain. Kata 'mirip' diartikan sebagai tingkat kesamaan karakteristiknya.

Proses *clustering* pada dasarnya mencari dan mengelompokkan data yang mirip satu dengan yang lain, maka kriteria mirip (*similarity*) adalah dasar dari metode *clustering*. Proses pengolahan data sehingga sekumpulan data mentah dapat dikelompokkan menjadi satu atau beberapa *cluster* adalah sebagai berikut:

1. Menetapkan ukuran jarak antar data. Mengukur kesamaan antar objek (*similarity*). Sesuai prinsip dasar *cluster* yang mengelompokkan objek yang mempunyai kemiripan, maka proses pertama adalah mengukur seberapa jauh ada kesamaan antar objek. Ada 2 metode yang digunakan, yaitu mengukur korelasi antar sepasang objek variabel dan mengukur jarak antara dua objek.
2. Melakukan proses standardisasi data jika diperlukan. Setelah cara mengukur jarak ditetapkan, yang juga perlu diperhatikan adalah apakah satuan data mempunyai perbedaan yang besar. Sebagai contoh, jika variabel penghasilan mempunyai satuan juta, sedangkan usia seseorang hanya mempunyai satuan puluhan, maka perbedaan yang mencolok ini akan membuat perhitungan jarak menjadi tidak valid.

- Melakukan proses *clustering*. Setelah data yang dianggap mempunyai satuan yang sangat berbeda sudah diseragamkan, langkah selanjutnya adalah membuat *cluster*. Proses inti dari *clustering* adalah pengelompokan data yang bisa dilakukan dengan 2 metode, yaitu metode hirarki dan metode non hirarki.

Lalu terdapat pula asumsi pada analisis kelompok diantaranya adalah sampel yang diambil benar-benar bisa mewakili populasi yang ada dan kemungkinan adanya korelasi antar objek. Jika terdapat korelasi, maka dianjurkan untuk melakukan analisis komponen utama yang akan dijelaskan pada subbab berikutnya.

2.2 Analisis Korelasi

Analisis korelasi mencoba mengukur keeratan hubungan antara dua peubah melalui sebuah bilangan yang disebut koefisien korelasi. Ukuran hubungan linear antara dua peubah diduga dengan koefisien korelasi dirumuskan sebagai berikut (Walpole, 1995):

$$r = \frac{n \sum_{i=1}^n X_{1i} X_{2i} - \left(\sum_{i=1}^n X_{1i} \right) \left(\sum_{i=1}^n X_{2i} \right)}{\sqrt{\left[n \sum_{i=1}^n X_{1i}^2 - \left(\sum_{i=1}^n X_{1i} \right)^2 \right] \left[n \sum_{i=1}^n X_{2i}^2 - \left(\sum_{i=1}^n X_{2i} \right)^2 \right]}}. \quad (2.1)$$

Keterangan:

- r = koefisien korelasi
- X_{1i} = amatan ke- i pada kelompok pertama
- X_{2i} = amatan ke- i pada kelompok kedua
- n = jumlah amatan.

Analisis kelompok tidak dapat dilakukan jika terdapat korelasi antar peubah, sehingga dilakukan analisis komponen utama dengan tujuan membentuk

peubah-peubah baru yang tidak saling berkorelasi, yang akan dibahas pada subbab berikut.

2.3 Analisis Komponen Utama

Analisis Komponen Utama (*Principal Component Analysis*) adalah analisis multivariat yang mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru yang tidak saling berkorelasi dengan mereduksi sejumlah variabel tersebut sehingga mempunyai dimensi yang lebih kecil namun dapat menerangkan sebagian besar keragaman variabel aslinya.

Banyaknya komponen utama yang terbentuk sama dengan banyaknya variabel asli. Pereduksian (penyederhanaan) dimensi dilakukan dengan kriteria persentase keragaman data yang diterangkan oleh beberapa komponen utama pertama. Apabila beberapa komponen utama pertama telah menerangkan lebih dari 75 % keragaman data asli, maka analisis cukup dilakukan sampai dengan komponen utama tersebut.

Bila komponen utama diturunkan dari populasi multivariat normal dengan random vektor $X = (\overline{X}_1, \overline{X}_2, \dots, \overline{X}_p)$ dan vektor rata-rata $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ dan matriks kovarians Σ dengan akar ciri (*eigenvalue*) yaitu $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ didapat kombinasi linier komponen utama yaitu sebagai berikut (Johnson & Wichern, 2002).

$$Y_1 = e'_1 X = e'_{11} X_1 + e'_{21} X_2 + \dots + e'_{p1} X_p$$

$$Y_2 = e'_2 X = e'_{12} X_1 + e'_{22} X_2 + \dots + e'_{p2} X_p$$

$$\vdots$$

$$Y_p = e'_p X = e'_{1p} X_1 + e'_{2p} X_2 + \dots + e'_{pp} X_p$$

Maka $Var(Y_i) = ei'\Sigma ei$ dan $Cov(Y_i, Y_k) = ei'\Sigma ek$ dimana $i, k = 1, 2, \dots, p$.

Syarat untuk membentuk komponen utama yang merupakan kombinasi linear dari variabel X agar mempunyai varian maksimum adalah dengan memilih vektor ciri (*eigen vector*) yaitu $e = (e_1, e_2, \dots, e_p)$ sedemikian hingga $Var(Y_i) = ei'\Sigma ei$ maksimum dan $ei'ei = 1$. Pembentukan komponen utama dijelaskan seperti berikut:

1. Komponen utama pertama adalah kombinasi linear e'_1X yang memaksimalkan $Var(e'_1X)$ dengan syarat $e'_1e_1 = 1$.
2. Komponen utama kedua adalah kombinasi linear e'_2X yang memaksimalkan $Var(e'_2X)$ dengan syarat $e'_2e_2 = 1$.
3. Komponen utama ke- i adalah kombinasi linear e'_iX yang memaksimumkan $Var(e'_iX)$ dengan syarat $e'_ie_k = 1$ dan $Cov(e'_ie_k) = 0$ untuk $k < i$.

Antar komponen utama tersebut tidak berkorelasi dan mempunyai variasi yang sama dengan akar ciri dari Σ merupakan varian dari komponen utama \mathbf{Y} , sehingga matriks ragam peragam dari \mathbf{Y} adalah:

$$\Sigma = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}.$$

Total keragaman variabel asal akan sama dengan total keragaman yang diterangkan oleh komponen utama yaitu:

$$\sum_{j=1}^p \text{var}(X_j) = \text{tr}(\Sigma) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{j=1}^p \text{var}(\mathbf{Y}_j).$$

Penyusutan dimensi dari variabel asal dilakukan dengan mengambil sejumlah kecil komponen yang mampu menerangkan bagian terbesar keragaman

data. Apabila komponen utama yang diambil sebanyak q komponen, dimana $q < p$, maka proporsi dari keragaman total yang bisa diterangkan oleh komponen utama ke- i adalah:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p}; i = 1, 2, \dots, p.$$

Penurunan komponen utama dari matriks korelasi dilakukan apabila data sudah terlebih dahulu ditransformasikan ke dalam bentuk baku \mathbf{Z} . Transformasi ini dilakukan terhadap data yang satuan pengamatannya tidak sama. Bila variabel yang diamati ukurannya pada skala dengan perbedaan yang sangat lebar atau satuan ukurannya tidak sama, maka variabel tersebut perlu dibakukan (*standardized*). Variabel baku (\mathbf{Z}) didapat dari transformasi terhadap variabel asal dalam matriks berikut:

$$Z = \left(V^{1/2} \right)^{-1} (X - \mu).$$

$V^{1/2}$ adalah matriks simpangan baku dengan unsur diagonal utama adalah $a_{ii}^{1/2}$ sedangkan unsur lainnya adalah nol. Nilai harapan $E(\mathbf{Z}) = 0$ dan keragamannya adalah:

$$Cov(Z) = \left(V^{1/2} \right)^{-1} \Sigma \left(V^{1/2} \right)^{-1} = \rho.$$

Dengan demikian komponen utama dari \mathbf{Z} dapat ditentukan dari vektor ciri yang didapat melalui matriks korelasi variabel asal ρ . Untuk mencari akar ciri dan menentukan vektor pembobotnya sama seperti pada matriks Σ . Sementara *trace* matriks korelasi ρ akan sama dengan jumlah p variabel yang dipakai.

Salah satu tujuan dari analisis komponen utama adalah mereduksi dimensi data asal yang semula terdapat p variabel bebas menjadi q komponen utama (dimana $q < p$). Kriteria pemilihan q yaitu:

1. Proporsi kumulatif keragaman data asal yang dijelaskan oleh q komponen utama minimal 80%, dan proporsi total variansi populasi bernilai cukup besar (Johnson, 2002).
2. Dengan menggunakan *scree plot* yaitu plot antara i dengan λ_i , berdasarkan *scree plot* ditentukan dengan melihat letak terjadinya belokan dengan menghapus komponen utama yang menghasilkan beberapa nilai eigen kecil membentuk pola garis lurus (Johnson, 2002).

Contoh 2.3.1. Berikut akan diberikan data tentang skor jawaban benar dari 20 butir soal yang diujikan kepada 8 peserta tes sebagai berikut:

Tabel 2.1: Hasil nilai Tes A dan Tes B

Peserta	Tes A	Tes B
1	2	9
2	16	20
3	8	0
4	18	11
5	10	13
6	4	4
7	10	17
8	12	16

Analisis yang akan dilakukan adalah analisis komponen utama untuk mengetahui apakah peserta tes harus menghadapi kedua tes tersebut atau cukup salah satunya saja. Langkah pertama yang dilakukan adalah melihat nilai korelasi yang bertujuan untuk mencari keeratan hubungan antar 2 peubah. Untuk mencari nilai korelasi menggunakan persamaan pada (2.1), maka hasilnya adalah sebagai berikut:

$$r = \frac{8 \cdot 1044 - (80 \cdot 90)}{\sqrt{[(8 \cdot 1008) - 80^2][(8 \cdot 1332) - 90^2]}}$$

$$r = 0,56$$

Dari hasil perhitungan koefisien korelasi di atas, menunjukkan nilai koefisien-nya kecil dan kurang signifikan, sehingga perlu dilakukan analisis komponen utama, dengan langkah sebagai berikut

Tabel 2.2: Hasil nilai Tes A dan Tes B

Peserta	Tes A	Tes B
1	2	9
2	16	20
3	8	0
4	18	11
5	10	13
6	4	4
7	10	17
8	12	16

Didapat matriks kovarians

$$\Sigma = \begin{bmatrix} 29,7142 & 20,5714 \\ 20,5714 & 45,6429 \end{bmatrix}$$

Dari matriks kovarians didapat nilai eigen

$$\lambda_1 = 59,7379$$

$$\lambda_2 = 15,6193$$

Kemudian didapat vektor eigen berdasarkan λ_1 dan λ_2 sebagai berikut

$$e'_1 = \begin{bmatrix} 0,5652 & 0,8249 \end{bmatrix}$$

$$e'_2 = \begin{bmatrix} 0,8249 & -0,5652 \end{bmatrix}$$

Dari vektor eigen di atas, maka didapat persamaan komponen utamanya

$$Y_1 = 0,56X_1 + 0,82X_2$$

$$Y_2 = 0,82X_1 - 0,56X_2$$

Proporsi total variansi dari matriks kovarians yang dijelaskan komponen utama adalah

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = 0,8$$

$$\frac{\lambda_2}{\lambda_1 + \lambda_2} = 0,2$$

Terlihat dari data di atas $\lambda_1 = 59,7379$ persen sedangkan proporsi total variansi pertama yaitu 80 persen, sehingga dapat disimpulkan cukup diperlukan satu tes saja karena kontribusinya sudah cukup besar dalam menggambarkan keragaman data.

Setelah penjelasan teori tentang analisis klaster, analisis korelasi, dan analisis komponen utama, berikutnya kita akan mulai memasuki metode yang akan digunakan pada skripsi ini yaitu metode *Hybrid Mutual Cluster* yang akan dijelaskan pada subbab selanjutnya

2.4 Mutual Cluster

Mutual cluster adalah suatu pengelompokan yang menggunakan jarak terbesar antara pasangan dalam kelompok yang lebih kecil dari jarak terpendek ke setiap titik di luar kelompok. Hal ini berarti bahwa jarak maksimal antar obyek dalam sebuah mutual cluster lebih kecil dibandingkan jarak minimal beberapa obyek di luar mutual cluster. Data yang terkandung dalam sebuah mutual cluster tidak pernah dipisahkan (Chipman dan Tibshirani, 2006). Metode tersebut memiliki beberapa implikasi dalam sebuah *mutual cluster*. Implikasi yang paling jelas adalah untuk mendukung gagasan bahwa dalam sebuah *mutual cluster* berisi informasi pengelompokan yang kuat, tidak peduli pendekatan *linkage* mana yang digunakan. Hal ini dapat membantu dalam interpretasi metode *bottom-up*. Informasi tambahan tersebut dapat membantu dalam interpretasi dari *mutual cluster*, atau dalam menentukan keputusan untuk pembagian kelompok. *Hybrid* ini juga mempertahankan metode *top-down*

yang akurat membagi data menjadi pengelompokan yang baik. Tahapan awal akan dilakukan pengelompokan secara *bottom-up*. Jarak obyek satu dengan obyek yang lain dihitung yang selanjutnya akan dicari jarak terdekat (minimal). Kelompok inilah yang menjadi *mutual cluster* pertama. Setelah itu, jarak antara kelompok yang terbentuk dengan obyek yang lain dihitung kembali. Lalu jarak terdekat (minimal) juga dicari kembali. Langkah tersebut terus dilakukan sampai semua obyek bergabung menjadi satu kelompok besar. Penentuan *mutual cluster* harus memiliki jarak minimal antar obyek di luar *mutual cluster*. Pada tahapan ini *mutual cluster* yang telah terbentuk harus dipertahankan. Oleh karena itu, obyek-obyek tersebut akan dibagi 2 kelompok. Selanjutnya, koordinat dari pusat kelompok (*means*) masing-masing kelompok ditentukan. Kemudian jarak masing-masing obyek dari koordinat pusat dihitung dan kembali menentukan obyek ke kelompok terdekat. Jika obyek dipindahkan dari posisi awal, pusat kelompok harus diperbaharui sebelum diproses lebih lanjut.

2.5 Metode Pengelompokkan Objek

2.5.1 Metode Pengelompokan Hirarki

1. Single Linkage

Input pada algoritma *single linkage* dapat berupa jarak atau kesamaan antara pasangan-pasangan objek. Grup dibentuk dari kesatuan individu dengan menggabungkan tetangga terdekatnya, dimana kata "tetangga terdekat" mengandung arti jarak terkecil atau kesamaan terbesar (terbanyak).

Sebagai langkah awal kita harus menemukan jarak terkecil pada $D =$

$\{d_{ik}\}$ dan menggabungkan objek-objek yang saling berkorespondensi, katakanlah U dan V , untuk mendapatkan kelompok (UV) . Jarak antara (UV) dan kelompok yang lainnya, katakanlah W , dihitung dengan cara

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\}$$

Di sini, nilai d_{UW} dan d_{VW} adalah jarak antara tetangga terdekat dari kelompok U dan W serta kelompok V dan W , begitupun sebaliknya.

2. Complete Linkage

Prosedur pengelompokan *complete linkage* hampir sama dengan *single linkage*, dengan satu pengecualian. Pada setiap tingkat, jarak (kesamaan) antar kelompok ditentukan dengan jarak (kesamaan) antara dua elemen, satu dari setiap kelompok, yakni yang paling jauh. Dengan demikian *complete linkage* menjamin bahwa dalam seluruh item pada kelompok terdapat jarak maksimum (atau kesamaan minimum).

Algoritma aglomeratif umum dimulai dengan menemukan entri (elemen) dalam $D = \{d_{ik}\}$ dan menggabungkan objek yang berkorespondensi, misalkan U dan V , untuk membentuk kelompok (UV) . Pada langkah ketiga, dalam algoritma umum, jarak (UV) dan kelompok lainnya, misalkan W ditentukan sebagai berikut:

$$d_{(uv)w} = \max\{d_{uw}, d_{vw}\}$$

di mana d_{uw} dan d_{vw} merupakan jarak terjauh antara anggota kelompok U dan W serta kelompok V dan W , begitupun sebaliknya.

3. Average Linkage

Perhitungan *Average Linkage* didasarkan pada rata-rata jarak dari seluruh objek pada suatu cluster dengan seluruh objek pada kelompok lain.

Algoritma yang digunakan dalam *Average Linkage* hampir sama dengan algoritma *agglomerative hierarchical clustering*. Dimulai dengan mencari jarak dari matriks, yaitu $D = \{d_{ik}\}$.

Untuk mencari objek terdekat, sebagai contoh U dan V , objek ini digabung ke dalam bentuk kluster UV . Untuk tahap ketiga, jarak antara UV dan kluster W adalah:

$$d_{(uv)w} = \frac{\sum_i \sum_j d_{(ik)j}}{N_{(uv)}N_W}$$

di mana d_{ik} adalah jarak antara objek I pada kluster (UV) dan objek k pada kluster W , $N_{(UV)}$ dan N_W adalah jumlah item-item pada kluster (UV) dan W .

2.5.2 Metode Pengelompokan non-Hirarki

Metode Pengelompokan hirarki digunakan apabila belum ada informasi jumlah kelompok. Sedangkan metode pengelompokan nonhirarki bertujuan pengelompokan n objek ke dalam k kelompok ($k < n$). Salah satu pengelompokan pada non hirarki adalah dengan menggunakan metode K-Means.

Metode ini merupakan metode pengelompokan yang bertujuan pengelompokan objek sedemikian sehingga jarak tiap-tiap objek ke pusat kelompok di dalam satu kelompok adalah kelompok minimum. Algoritma K-Means adalah:

1. Tentukan jumlah *cluster*.
2. Cari data yang lebih dekat dengan pusat *cluster*. Hitung jarak dari masing-masing objek dari pusat *cluster*. Tentukan kembali pusat *cluster*.
3. Ulangi langkah 2 sampai tidak ada yang berpindah posisi.

2.6 Jarak *Square Euclidean*

Jarak *Square Euclidean* merupakan jarak yang dikembangkan dari jarak *Euclidean*. Pada jarak *Euclidean*, jarak tersebut mempunyai tiga asumsi yang diantaranya adalah antar peubah tidak saling berkorelasi, memiliki satuan pengukuran yang sama, dan pengukuran pembakuan mempunyai rata-rata nol dan standar deviasi satu. Jarak *Euclidean* merupakan jarak antar objek, misalkan dua objek ke- i dan ke- j yang berada pada p dimensi dimana formulanya sebagai berikut (Johnson & Wichern, 2002):

$$d_{(X_i, X_j)} = \sqrt{\sum_{q=1}^p (X_{iq} - X_{jq})^2},$$

Keterangan:

$d_{(X_i, X_j)}$ = jarak antar objek pada X_i dengan objek pada X_j ; $X_i \neq X_j$

q = banyak peubah; $1, 2, \dots, p$

X_{iq} = nilai dari obyek X_i pada variabel ke- q

X_{jq} = nilai dari obyek X_j pada variabel ke- q

sedangkan pada jarak *Square Euclidean* yang merupakan pengembangan dari jarak *Euclidean* dapat diartikan sebagai suatu ukuran kesamaan jumlah kuadrat perbedaan tanpa akar kuadrat. Jarak *Square Euclidean* antara dua unit/observasi yang berdimensi p dengan koordinat $X_{iq} = (X_{i1}, X_{i2}, \dots, X_{ip})$ dan $X_{jq} = (X_{j1}, X_{j2}, \dots, X_{jp})$, dengan $X_i \neq X_j$. Formula jarak *Square Euclidean* adalah sebagai berikut (Hair dkk., 2010):

$$d_{(X_i, X_j)} = \sum_{q=1}^p (X_{iq} - X_{jq})^2.$$

Keterangan:

$d_{(X_i, X_j)}$ = kuadrat jarak antar objek pada X_i dengan objek pada X_j ; $i \neq j$

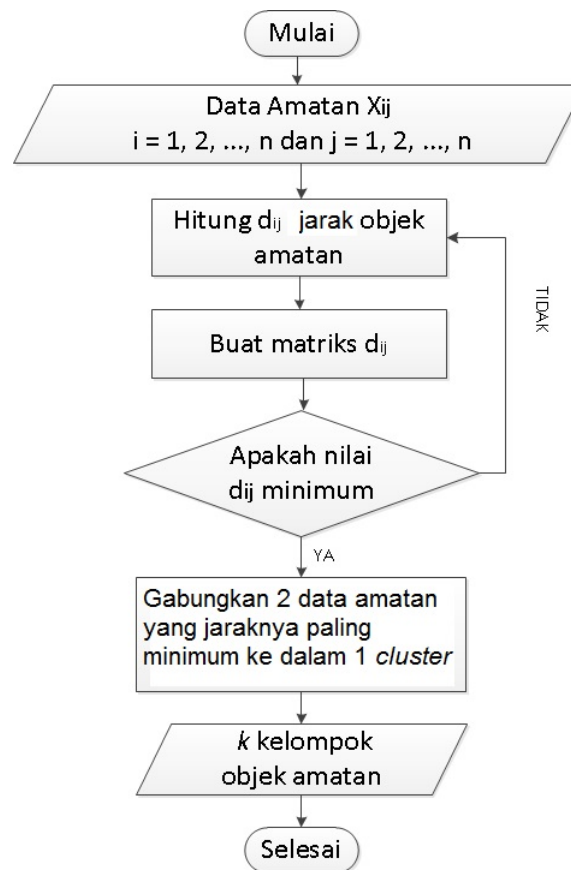
- q = banyak peubah; $1, 2, \dots, p$
 X_{iq} = nilai dari obyek X_i pada variabel ke- q
 X_{jq} = nilai dari obyek X_j pada variabel ke- q

Jarak *Square Euclidean* digunakan dalam pengelompokan metode *bottom-up* dan *top-down*. Berikut akan dijelaskan tentang metode *bottom-up* dan *top-down*.

2.7 Pengelompokan Metode *Bottom-Up*

Pengelompokan dengan menggunakan metode *bottom-up* adalah suatu metode hierarki dimana n buah kelompok digabungkan menjadi satu kelompok tunggal. Metode *bottom-up* ini meletakkan setiap objek data sebagai sebuah kelompok tersendiri (*atomic cluster*) yang selanjutnya kelompok-kelompok tersebut bergabung menjadi kelompok besar sampai akhirnya semua objek menyatu dalam sebuah kelompok tunggal. Jarak antar objek diperlukan pada tahap awal dalam penggabungan 2 kelompok dengan metode *agglomerative* (Hair dkk., 2010).

Algoritma *bottom-up* dimulai dari menginput data amatan X_{ij} yang kemudian dihitung jarak objek amatan (d_{ij}). Selanjutnya buat matriks d_{ij} yang kemudian dilihat apakah nilai d_{ij} minimum, ulangi hitung nilai jarak objek amatan apabila nilai d_{ij} belum minimum. Setelah dapat nilai d_{ij} minimum, gabungkan objek amatan yang sama menjadi satu kelompok. Output berupa k kelompok objek amatan. Berikut adalah diagram alir pengelompokan metode *bottom-up (agglomerative)*:



Gambar 2.1: Diagram alir pengelompokan metode *bottom-up* (agglomerative).

Contoh 2.7.1. Sebagai contoh dapat dilihat pada kasus berikut ini. Akan diberikan pengelompokan data hasil nilai renang mahasiswa sebagai berikut:

Tabel 2.3: Matriks Jarak Hasil Nilai Renang Mahasiswa

	M1	M2	M3	M4
M1	0	7	2	5
M2	7	0	5	2
M3	2	5	0	3
M4	5	2	3	0

Dari tabel di atas, berikut ini merupakan penyelesaian pengelompokan nilai dengan menggunakan metode klasterisasi hirarki *bottom-up* dengan *agglomerative*.

1. Tahap 1

Tabel 2.4: Matriks Jarak Hasil Nilai Renang Mahasiswa

	M1	M2	M3	M4
M1	0	7	2	5
M2	7	0	5	2
M3	2	5	0	3
M4	5	2	3	0

Dilihat dari matriks jarak diatas, jarak terkecil berada di M1 dan M3, yang kemudian akan dilakukan penggabungan M1 dan M3

2. Tahap 2

Tabel 2.5: Matriks Setelah Penggabungan M1 dan M3

	M1/M3	M2	M4
M1/M3	0	5	3
M2	5	0	2
M4	3	2	0

Setelah M1 dan M3 digabungkan menjadi satu, jarak terkecil sekarang berada di M2 dan M4. Maka selanjutnya akan digabungkan M2 dan M4.

3. Tahap 3

Tabel 2.6: Matriks Setelah Penggabungan M1/M3 dan M2/M4

	M1/M3	M2/M4
M1/M3	0	3
M2/M4	3	0

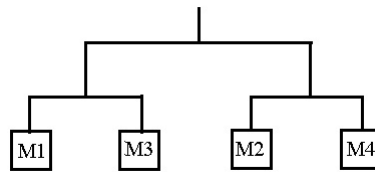
Setelah itu, jarak terkecil sekarang berada di M1/M3 dan M2/M4, gabungkan keduanya.

4. Tahap 4

Tabel 2.7: Matriks Setelah Penggabungan M1/M3/M2/M4

	M1/M2/M3/M4
M1/M2/M3/M4	0

Dari penyelesaian di atas maka dapat digambarkan pengelompokan tersebut menggunakan dendogram sebagai berikut:



Gambar 2.2: Dendogram Data Hasil Nilai Mahasiswa.

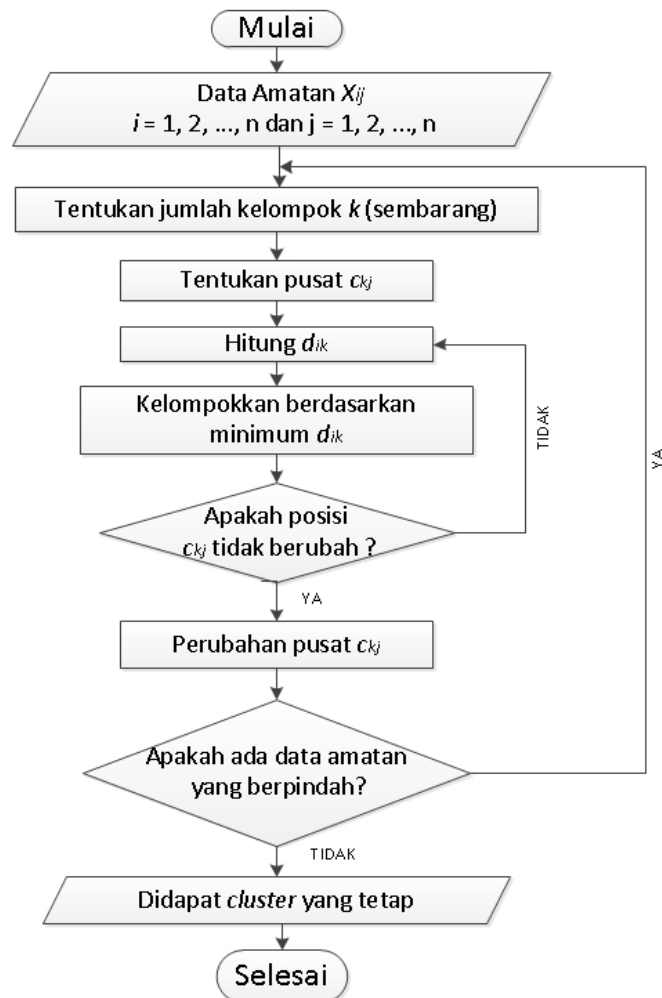
Dapat dilihat dari dendogram di atas disimpulkan mahasiswa 1 dan mahasiswa 3 berada dalam 1 *cluster*, dan mahasiswa 2 dan mahasiswa 4 berada dalam 1 *cluster*. Dalam pengelompokan metode *Hybrid Mutual Clustering*, selain metode *bottom-up*, dilakukan juga pengelompokan dengan metode *top-down*, sehingga pada subbab selanjutnya akan dijelaskan secara rinci mengenai pengelompokan metode *top-down*.

2.8 Pengelompokan Metode *Top-Down*

Pengelompokan dengan metode *top-down* adalah membagi n objek ke dalam k kelompok yang bertujuan untuk mengelompokkan objek sehingga jarak antar objek ke pusat kelompok di dalam satu kelompok minimum.

Proses pertama dalam mengelompokkan dengan menggunakan metode *top-down* yang bersifat non-hierarki (*k-means*) adalah terdapat data amatan X_i dan X_j . Kemudian, partisikan obyek ke dalam k kelompok. Langkah selanjutnya, hitung pusat kelompok dimana pusat kelompok itu sendiri merupakan

rata-rata dari keseluruhan obyek yang berada dalam kelompok tersebut. Setelah itu, hitung jarak setiap obyek ke pusat kelompok dengan menggunakan jarak *Square Euclidean*. Jika terdapat obyek yang berpindah dari posisi awal, maka pusat kelompok dihitung kembali dan periksa kembali posisi obyek. Ulangi langkah-langkah tersebut sampai tidak ada obyek yang berpindah posisi. Perhitungan berhenti ketika obyek sudah tidak berpindah posisi dan membentuk kelompok *Hybrid Mutual Clustering*. Berikut adalah contoh pengelompokan dengan menggunakan metode *top-down (k-means)*:



Gambar 2.3: Diagram alir pengelompokan metode *top-down (k-means)*.

Contoh 2.8.1. Berikut ini merupakan data padi pada tahun 2013 di provinsi Jawa Timur, data yang digunakan hanya data produksi dan luas lahan (sumber: <https://syafrudinmtop.blogspot.co.id/>):

Tabel 2.8: Data Produksi dan Luas Lahan di Provinsi Jawa Timur Tahun 2013

No.	Kota /Kab	Luas Lahan	Produksi
1	Ponorogo	66,693	402,047
2	Trenggalek	31,136	182,848
3	Tulungagung	49,23	259,581
4	Blitar	50,577	289,494
5	Kediri	51,083	281,392
6	Malang	65,597	464,498
7	Lumajang	72,552	387,168
8	Jember	162,619	964,001
9	Banyuwangi	113,609	706,419
10	Bondowoso	61,33	329,557
11	Situbondo	48,902	290,954
12	Probolinggo	59,130	311,258

Selanjutnya akan dilakukan pengelompokan menggunakan *k-means cluster*, dengan pengambilan *cluster* awal 3 dan iterasi sebanyak 4 kali. Hasil tabel adalah sebagai berikut:

Tabel 2.9: Data Hasil Akhir pengelompokan Menggunakan *K-means Cluster*

NO	Kota /Kab	C1	C2	C3	Jarak Terpendek
1	Ponorogo	570,083	212,882	307,967	212,882
2	Trenggalek	792,141	25,362	530,027	25,362
3	Tulungagung	713,488	70,419	451,452	70,419
4	Blitar	683,749	100,168	421,663	100,168
5	Kediri	691,661	92,051	429,602	92,051
6	Malang	508,838	275,226	246,639	246,639
7	Lumajang	583,822	198,439	321,880	198,439
8	Jember	0	781,903	262,203	0
9	Banyuwangi	262,203	520,208	0	0
10	Bondowoso	642,479	140,220	380,471	140,220
11	Situbondo	682,586	101,723	420,474	101,723
12	Probolinggo	660,896	121,856	398,899	121,856

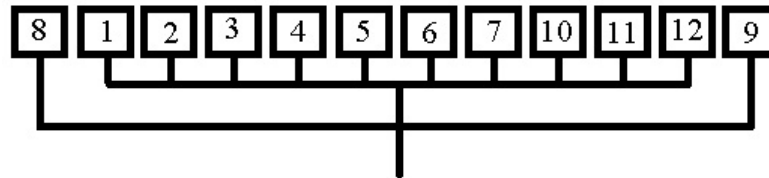
Maka diperoleh hasil akhir adalah pusat *cluster* 1 dengan menggunakan data ke-8, pusat *cluster* 2 dengan menggunakan data ke-7, pusat *cluster* 3 dengan menggunakan data ke-2. Dari hasil tabel di atas, pengelompokannya menjadi sebagai berikut:

Tabel 2.10: Data Hasil Pengelompokan Kota dengan Menggunakan *K-means Cluster*

No.	C1	C2	C3
1		1	
2		1	
3		1	
4		1	
5		1	
6		1	
7		1	
8	1		
9			1
10		1	
11		1	
12		1	

Pada tabel di atas, jika kota tersebut bernilai 1 pada setiap *cluster*, maka kota tersebut menjadi satu kelompok untuk tiap *cluster* tersebut. Dari penyelesaian di atas, maka dapat disimpulkan bahwa dalam *cluster* 1 terdapat satu kota, yaitu Kota Jember. Pada *cluster* 2 terdapat sepuluh kota untuk dijadikan satu kelompok, yaitu Kota Ponorogo, Kota Trenggalek, Kota Tulungagung, Kota Blitar, Kota Kediri, Kota Malang, Kota Lumajang, Kota Bondowoso, Kota Situbondo, dan Kota Probolinggo. Sedangkan pada *cluster* 3 terdapat satu kota, yaitu Kota Banyuwangi. Kota-kota tersebut dijadikan kelompok berdasarkan jarak terpendek terhadap pusat *cluster* awal.

Selanjutnya, penyelesaian di atas menghasilkan dendogram sebagai berikut:



Gambar 2.4: Dendrogram Data Hasil Pengelompokkan Kota.

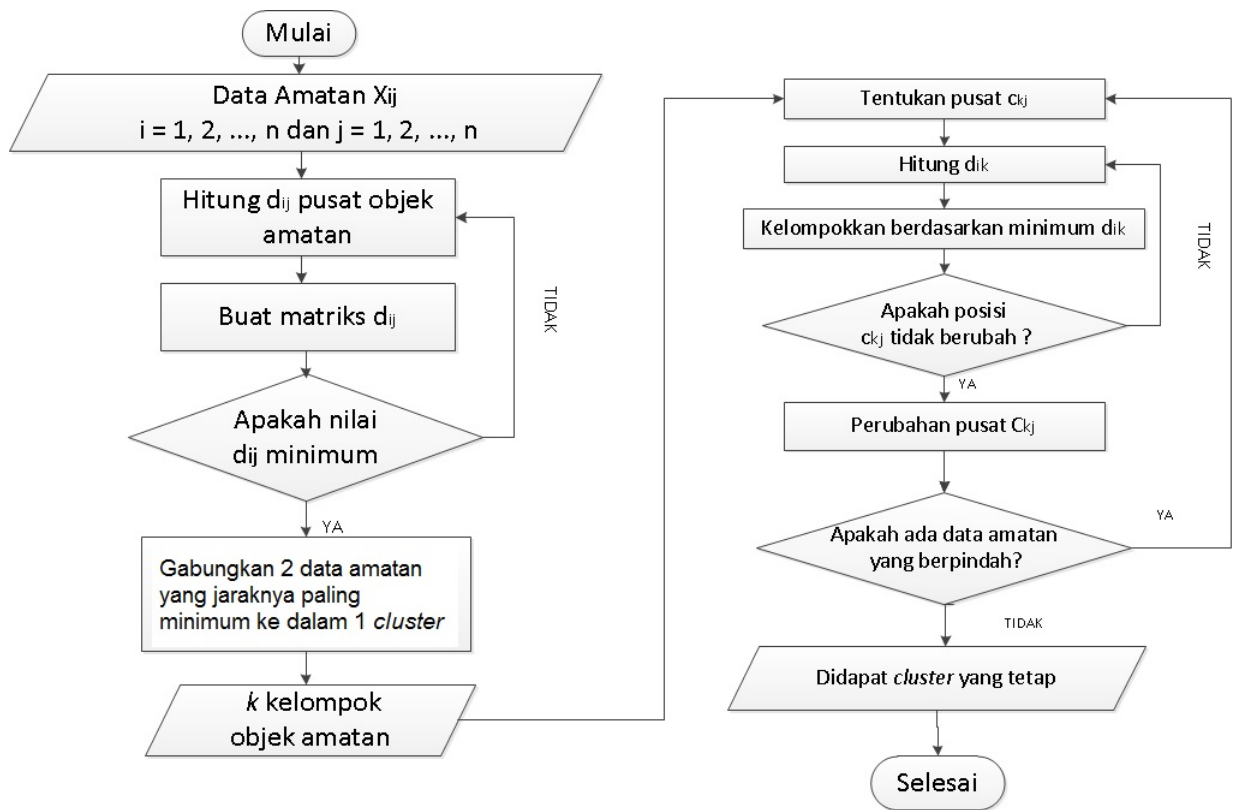
BAB III

PEMBAHASAN

Pada bab ini akan dibahas secara lebih terperinci mengenai pengelompokan menggunakan metode *Hybrid Mutual Clustering* dengan jarak *Square Euclidean* dan juga penerapan metode *Hybrid Mutual Clustering* ke dalam contoh kasus. Sebelumnya, akan dibahas terlebih dahulu mengenai pengelompokan metode *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean*.

3.1 Pengelompokan Menggunakan Metode *Hybrid Mutual Clustering* dengan Jarak *Square Euclidean*

Metode *Hybrid Mutual Clustering* merupakan metode penggabungan antara metode *bottom-up* dan *top-down*. Pengelompokan *Hybrid Mutual Clustering* ini menggunakan jarak *Square Euclidean*. *Hybrid mutual* disini mengkombinasikan kelebihan metode *bottom-up clustering (agglomerative)* dan *top-down clustering (k-means)*. Metode *Hybrid* terdiri dari dua metode yaitu *bottom-up* yaitu metode pengelompokan dimulai dari kelompok kecil menjadi kelompok yang lebih besar (*agglomerative*) dan *top-down* yaitu metode pengelompokan dengan memecah kelompok besar menjadi kelompok lebih kecil seperti metode *k-means* membagi sebanyak k kelompok. Berikut adalah diagram alir pengelompokan *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean*:



Gambar 3.1: Diagram alir pengelompokan menggunakan metode *Hybrid Mutual Clustering* dengan jarak *Square Euclidean*.

Berikut akan dijelaskan algoritma pengelompokan *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean*

1. Masukkan data amatan X_{ij} dengan $i = 1, \dots, n$ dan $j = 1, \dots, n$.
2. Hitung d_{ij} untuk mencari jarak antar objek amatan, dengan persamaan berikut:

$$d_{ij} = \sum_{q=1}^p (x_{iq} - x_{jq})^2.$$

3. Bentuk matriks d_{ij} .
4. Tentukan nilai minimum d_{ij} . Jika d_{ij} minimum maka proses berlanjut ke tahap 4 dari diagram alir. Namun, jika d_{ij} tidak minimum, maka proses

mengulang dari tahap 2 diagram alir.

5. Gabungkan 2 data amatan yang jaraknya paling minimum ke dalam satu *cluster* sampai diperoleh k kelompok, dimana $d_{ij}=0$.
6. Bentuk sebanyak k kelompok objek amatan.
7. Tentukan pusat $c_{kj} = \frac{\sum_{j=1}^p x_{ij}}{p}$ dengan peubah sebanyak p .
8. Hitung $d_{ik} = \sum_{j=1}^p (x_{ij} - c_{kj})^2$.
9. Kelompokkan berdasarkan minimum d_{ik} .
10. Tentukan nilai c_{kj} . Jika c_{kj} berubah maka proses berlanjut ke tahap 11. Jika c_{kj} tidak berubah, maka proses mengulang dari tahap 8.
11. Pusat c_{kj} berubah.
12. Identifikasi data amatan. Jika data amatan berubah posisi maka proses mengulang dari tahap 7. Jika data amatan tidak berubah, maka proses berlanjut ke tahap 13.
13. Didapat *Cluster* tetap yang memiliki proporsi terbaik.

Dari pengelompokkan *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean* didapatkan kelompok sebanyak k dan memiliki jarak berdekatan yg menunjukkan bahwa karakteristik setiap objek amatan pada kelompok adalah sama.

Pada subbab selanjutnya akan dibahas mengenai penerapan metode *Hybrid Mutual Clustering* dengan jarak *Square Euclidean* ke dalam contoh kasus menggunakan program aplikasi SPSS

3.2 Penerapan Metode *Hybrid Mutual Clustering* dengan Jarak *Square Euclidean* ke dalam contoh kasus

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) dengan judul Persentase Penduduk Buta Aksara Menurut Provinsi di Indonesia. Pada situs tersebut data yang tertera ada dalam periode 2003-2015, tetapi yang diambil oleh penulis hanya tahun 2015 saja. Unit pengamatan yang dipakai ada 34 provinsi di Indonesia, meliputi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua. Data yang dipakai hanya mencakup 3 variabel, yaitu penduduk dengan umur di bawah 14 tahun, penduduk dengan umur 15 sampai 44 tahun, dan penduduk dengan umur di atas 45 tahun. Variabel tersebut digunakan karna tersediaan data.

Metode *Hybrid Mutual Clustering* ini akan diaplikasikan ke dalam pengelompokan Penduduk Buta Aksara Menurut Provinsi di Indonesia tahun 2015 menggunakan *software SPSS*. Data yang berhasil dikumpulkan ditampilkan di Lampiran 4.1.

Sebelum melakukan *clustering*, akan dilakukan terlebih dahulu standarisasi data. *Output* yang dihasilkan adalah sebagai berikut:

Zkurangdari14	Zumur15sampai44	Zlebihdari45
-0.39798	-0.30073	-0.30378
-0.59457	-0.25090	-0.85261
-0.57210	-0.29035	-0.80067
-0.63014	-0.28827	-0.81248
-0.43730	-0.25505	-0.50089
-0.50844	-0.25713	-0.65786
-0.39798	-0.25713	-0.41709
-0.21823	-0.28620	-0.09251
-0.39798	-0.17616	-0.52449
-0.61516	-0.29658	-0.69445
-0.76495	-0.34433	-1.06741
-0.46912	-0.29658	-0.57288
0.44644	-0.25298	0.68410
0.18806	-0.31734	0.29462
0.75537	-0.09935	1.05471
-0.34930	-0.28827	-0.19047
0.51197	-0.23014	0.94494
1.59791	0.33040	2.77082
0.75911	0.28681	1.08186
0.59622	0.05844	1.23647
-0.63202	-0.29450	-0.82428
-0.50657	-0.31734	-0.57878
-0.59644	-0.32979	-0.70389
0.09631	-0.07443	0.54129
-0.77244	-0.32149	-1.13233
-0.34368	-0.16786	-0.45840
0.78907	0.10411	1.30257
0.26295	-0.07236	0.79859
-0.51219	-0.23014	-0.70271
0.53631	0.33456	0.83400
-0.62640	-0.19070	-0.98480
-0.55900	-0.25921	-0.71097
-0.25755	0.07712	-0.47020
4.61982	5.55385	2.50998

Gambar 3.2: Tabel *Descriptive Z-Score*.

Selain tabel di atas, terdapat *output* lain yang menginformasikan data deskriptif dari variabel-variabel di atas, yang memperlihatkan nilai minimum, nilai maksimum, mean, dan deviasi standar. Tabel tersebut akan digunakan saat penafsiran hasil *cluster*. Berikut adalah tabel *Descriptive Statistics*:

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
kurangdari4	34	.37	29.17	4.4956	5.34099
umur15sampai44	34	.06	28.47	1.7185	4.81674
lebihdari45	34	.71	33.78	10.3038	8.47265
Valid N (listwise)	34				

Gambar 3.3: Tabel *Descriptive Statistics*.

Setelah data distandardisasi, proses dilanjutkan dengan melakukan metode pengelompokan *bottom-up (agglomerative)* terlebih dahulu dan dihasilkan output dengan penjelasan sebagai berikut:

- Tabel *output* pertama menyatakan bahwa semua data sejumlah 34 obyek telah diproses tanpa ada data yang hilang. Berikut adalah tabel *Case Processing Summary*:

Case Processing Summary ^a					
Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
34	100.0%	0	.0%	34	100.0%

a. Squared Euclidean Distance used

Gambar 3.4: Tabel *Case Processing Summary*

- Tabel *output Proximity Matrix* yang terdapat pada lampiran. Pada tabel tersebut menunjukkan bahwa semua data sejumlah 34 obyek telah diproses tanpa ada data yang hilang. Tabel tersebut menunjukkan matriks jarak antara variabel satu dengan variabel yang lain. Semakin kecil jarak *Euclidean*, maka semakin mirip kedua variabel tersebut sehingga akan membentuk kelompok (*cluster*).
- Selain tabel *Proximity Matrix*, terdapat pula tabel *output* yang berjudul *Agglomeration Schedule*. Tabel tersebut merupakan hasil proses clustering dengan metode Between Group Linkage. Setelah jarak antar variabel diukur dengan jarak *Square Euclidean*, maka dilakukan pengelom-

pokan, yang dilakukan secara bertingkat. Berikut adalah tabel *Agglomeration Schedule*:

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	21	,000	0	0	5
2	10	23	,002	0	0	11
3	12	22	,002	0	0	15
4	6	29	,003	0	0	6
5	3	4	,004	0	1	7
6	6	32	,004	4	0	11
7	2	3	,004	0	5	14
8	11	25	,005	0	0	18
9	9	26	,007	0	0	12
10	5	7	,009	0	0	12
11	6	10	,012	6	2	14
12	5	9	,014	10	9	15
13	1	16	,015	0	0	17
14	2	6	,024	7	11	22
15	5	12	,033	12	3	22
16	20	27	,044	0	0	20
17	1	8	,052	13	0	25
18	11	31	,055	8	0	27
19	13	17	,073	0	0	21
20	19	20	,093	0	16	23
21	13	28	,094	19	0	28
22	2	5	,100	14	15	27
23	15	19	,112	0	20	26
24	14	24	,128	0	0	28
25	1	33	,229	17	0	29
26	15	30	,244	23	0	30
27	2	11	,248	22	18	29
28	13	14	,284	21	24	30
29	1	2	,380	25	27	31
30	13	15	,603	28	26	31
31	1	13	3,636	29	30	32
32	1	18	12,940	31	0	33
33	1	34	63,260	32	0	0

Gambar 3.5: Tabel *Agglomeration Schedule*.

Penjelasan tabel *Agglomeration Schedule*:

1. Stage 1: terbentuk 1 *cluster* yang beranggotakan Kota 4 (Riau) dan Kota 21 (Kalimantan Tengah) dengan jarak hampir mendekati nol (perhatikan pada kolom *Coefficients*). Karena proses aglomerasi dimulai dari 2 obyek yang terdekat, maka jarak tersebut adalah yang terdekat dari sekian kombinasi jarak 34 obyek yang ada. Selanjutnya lihat kolom terakhir (*Next Stage*), terlihat angka 5. Hal ini berarti *clustering* selanjutnya dilakukan dengan melihat *Stage*

5, dengan penjelasan berikut.

2. Baris ke-5 (*stage 5*) terlihat obyek ke-3 (Kota Sumatera Barat) membentuk *cluster* dengan kota 4 (Riau). Dengan demikian, sekarang *cluster* terdiri dari 3 obyek yaitu Kota Sumatera Barat, Riau, dan Kalimantan Tengah. Sedangkan jarak sebesar 0,004 merupakan jarak rata-rata obyek terakhir yang bergabung dengan 2 obyek sebelumnya, seperti tampak dalam *Proximity matrix* dan dapat dihitung sebagai berikut:
 - (a) Jarak Kota Sumatera Barat dan Riau = 0,004
 - (b) Jarak Kota Sumatera Barat dan Kalimantan Tengah = 0,004
 - (c) Jarak rata-rata = $(0,004 + 0,004) / 2 = 0,004$
3. Stage 2 : terjadi pembentukan *cluster* Kota 10 (Kepulauan Riau) dan Kota 23 (Kalimantan Timur) yang berjarak 0,002, yang kemudian berlanjut ke *Stage 11*.
4. Demikian seterusnya dari *stage 3* sampai ke *stage* terakhir.

Proses aglomerasi ini bersifat kompleks, khususnya perhitungan koefisien yang melibatkan sekian banyak obyek dan terus bertambah. Proses aglomerasi pada akhirnya akan menyatukan semua obyek menjadi satu *cluster*. Hanya saja dalam prosesnya dihasilkan beberapa *cluster* dengan masing-masing anggotanya, tergantung jumlah *cluster* yang dibentuk.

- Hasil *output cluster membership* dapat dilihat pada tabel berikut ini:

Perincian jumlah cluster dengan anggota yang terbentuk dijabarkan sebagai berikut:

Cluster Membership			
Case	4 Clusters	3 Clusters	2 Clusters
1:ACEH	1	1	1
2:SUMATERA UTARA	1	1	1
3:SUMATERA BARAT	1	1	1
4:RIAU	1	1	1
5:JAMBI	1	1	1
6:SUMATERA SELATAN	1	1	1
7:BENGKULU	1	1	1
8:LAMPUNG	1	1	1
9:KEP. BANGKA BELITUNG	1	1	1
10:KEP. RIAU	1	1	1
11:DKI JAKARTA	1	1	1
12:JAWA BARAT	1	1	1
13:JAWA TENGAH	2	1	1
14:DI YOGYAKARTA	2	1	1
15:JAWA TIMUR	2	1	1
16:BANTEN	1	1	1
17:BALI	2	1	1
18:NUSA TENGGARA BARAT	3	2	1
19:NUSA TENGGARA TIMUR	2	1	1
20:KALIMANTAN BARAT	2	1	1
21:KALIMANTAN TENGAH	1	1	1
22:KALIMANTAN SELATAN	1	1	1
23:KALIMANTAN TIMUR	1	1	1
24:KALIMANTAN UTARA	2	1	1
25:SULAWESI UTARA	1	1	1
26:SULAWESI TENGAH	1	1	1
27:SULAWESI SELATAN	2	1	1
28:SULAWESI TENGGARA	2	1	1
29:GORONTALO	1	1	1
30:SULAWESI BARAT	2	1	1
31:MALUKU	1	1	1
32:MALUKU UTARA	1	1	1
33:PAPUA BARAT	1	1	1
34:PAPUA	4	3	2

Gambar 3.6: Tabel *Cluster Membership*.

1. Apabila diinginkan dibentuk 4 cluster, maka:
 - (a) Anggota cluster 1 adalah Kota Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, dan Papua Barat
 - (b) Anggota cluster 2 adalah Kota Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Timur, Kalimantan Barat,

Kalimantan Utara, Sulawesi Selatan, Sulawesi Tengah, dan Sulawesi Barat.

(c) Anggota cluster 3 adalah Kota Nusa Tenggara Barat.

(d) Anggota cluster 4 adalah Kota Papua.

2. Apabila ditentukan dibentuk 3 cluster, maka :

(a) Anggota cluster 1 adalah Kota Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Barat, Gorontalo, Maluku, Maluku Utara, dan Papua Barat.

(b) Anggota cluster 2 adalah Kota Nusa Tenggara Barat.

(c) Anggota cluster 3 adalah Kota Papua.

3. Apabila ditentukan dibentuk 2 cluster, maka :

(a) Anggota cluster 1 adalah Kota Kota Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Timur, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Barat, Gorontalo, Maluku, Maluku Utara, dan Papua Barat .

(b) Anggota cluster 2 adalah Kota Papua.

Dari hasil di atas dapat dilihat bahwa peralihan dari kolom 4 *Clusters* ke kolom 3 *Clusters*, yang terjadi adalah penggabungan variabel-variabel yang sudah ada, bukan mengacak variabel dari awal. Pada kolom 3 *Clusters*, terlihat anggota pada *cluster* 1 adalah nama-nama kota yang sebelumnya menjadi anggota *cluster* 1 dan *cluster* 2 dari kolom 4 *Clusters*. Demikian pula dari anggota *cluster* 1 dari kolom 2 *Clusters* adalah penggabungan dari anggota *cluster* 1 dan *cluster* 2 dari kolom 3 *Clusters*. Dari proses tersebut terlihat bahwa kota Nusa Tenggara Barat dan Papua memang jauh berbeda (tidak similar) dengan kota lainnya, dan kota Papua adalah yang paling berbeda dengan kota-kota lainnya.

- Terdapat satu lagi hasil *output SPSS* yaitu dendogram. Dendogram berguna untuk menunjukkan anggota *cluster* yang ada jika akan ditentukan berapa *cluster* yang seharusnya dibentuk. Sebagai contoh yang terlihat dalam dendogram, apabila akan dibentuk 2 *cluster*, maka *cluster* 1 beranggotakan Kota DKI Jakarta sampai dengan Kota Jawa Tengah (sesuai urutan dalam dendogram); dan *cluster* 2 beranggotakan sisanya. Demikian seterusnya dapat dengan mudah dilihat anggota tiap cluster sesuai jumlah *cluster* yang diinginkan. Dapat dilihat proses dendogram sama persis dengan hasil proses tabel output *Cluster Membership*, dengan perbedaan Dendogram adalah visualisasi proses *clustering* yang terjadi, sedangkan *Cluster Membership* langsung menunjukkan anggota *cluster* yang ada. Berikut adalah grafik dendogram:

```

***** H I E R A R C H I C A L   C L U S T E R   A N A L Y S I S *****
Dendrogram using Average Linkage (Between Groups)

Rescaled Distance Cluster Combine

  C A S E      0      5      10      15      20      25
Label      Num  +-----+-----+-----+-----+
RIAU              4  --+
KALIMANTAN TENGAH 21  --+
SUMATERA BARAT    3  --+
SUMATERA UTARA    2  --+
KEP. RIAU         10  --+
KALIMANTAN TIMUR 23  --+
SUMATERA SELATAN  6  --+
GORONTALO        29  --+
MALUKU UTARA     32  --+
JAWA BARAT       12  --+
KALIMANTAN SELATAN 22  --+
KEP. BANGKA BELITUNG 9  --+
SULAWESI TENGAH  26  --+
JAMBI            5  --+
BENGKULU         7  --+
DKI JAKARTA      11  --++
SULAWESI UTARA   25  --+ |
MALUKU           31  --+ |
ACEH             1  --+ |
BANTEN           16  --+ |
LAMPUNG          8  --+ |
PAPUA BARAT      33  --+ +-----+
KALIMANTAN BARAT 20  --+ | |
SULAWESI SELATAN 27  --+ | |
NUSA TENGGARA TIMUR 19  --+ | |
JAWA TIMUR       15  --+ | |
SULAWESI BARAT   30  --+ | +-----+
JAWA TENGAH      13  --++ | |
BALI             17  --+ | |
SULAWESI TENGGARA 28  --+ | |
DI YOGYAKARTA    14  --+ | |
KALIMANTAN UTARA 24  --+ | |
NUSA TENGGARA BARAT 18  -----+ |
PAPUA            34  -----+

```

Gambar 3.7: Dendrogram Analisis Klaster Hierarki

Dalam metode *Hybrid Mutual Clustering*, selain metode *bottom-up (agglomerative)*, terdapat juga metode *top-down (k-means)*. Dari proses yang dilakukan melalui program aplikasi *SPSS*, didapatkan beberapa output seperti berikut:

- *Output* pertama berupa tambahan 2 variabel, variabel QCL1 berisi nomor *cluster* untuk setiap kasus, dan varian QCL2 berisi jarak antara kasus tertentu dengan pusat kelompok. Tabel QCL terletak pada lampiran

- *Output* selanjutnya adalah tampilan pertama (*initial*) proses *clustering* data sebelum dilakukan iterasi

Initial Cluster Centers				
	Cluster			
	1	2	3	4
Zscore(umur14keawah)	.09631	4.61982	1.59791	-.77244
Zscore (umur15sampai44)	-.07443	5.55385	.33040	-.32149
Zscore(umur45keatas)	.54129	2.50998	2.77082	-1.13233

Gambar 3.8: Tabel *Initial Cluster Centers*

- Kemudian terdapat *output* proses iterasi. Tabel berikut adalah proses iterasi yang mencoba mengubah-ubah *cluster* yang ada sebelumnya sehingga menjadi lebih tepat dalam mengelompokkan.

Iteration History ^a				
Iteration	Change in Cluster Centers			
	1	2	3	4
1	.316	.000	.000	.511
2	.218	.000	.000	.055
3	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 3. The minimum distance between initial centers is 1,902.

Gambar 3.9: Tabel *Iteration History*

- Setelah terjadi tiga tahapan iterasi (proses pengulangan dengan ketepatan lebih tinggi dari sebelumnya)

Final Cluster Centers				
	Cluster			
	1	2	3	4
Zscore(umur14keawah)	.49418	4.61982	1.59791	-.50725
Zscore (umur15sampai44)	-.02627	5.55385	.33040	-.25553
Zscore(umur45keatas)	.87731	2.50998	2.77082	-.63882

Gambar 3.10: Tabel *Final Cluster Centers*

Angka pada tabel di atas berkaitan dengan proses standardisasi data sebelumnya, yang mengacu pada angka z , dengan ketentuan:

1. Angka negatif berarti data di bawah rata-rata total
2. Angka positif berarti data di atas rata-rata total

Pada tabel di atas terlihat dari ke 4 cluster yang ada, penduduk dengan umur di bawah 14 tahun memiliki nilai yang lebih besar dibandingkan dengan penduduk dengan umur 15 sampai 44 dan penduduk dengan umur di atas 45 tahun.

BAB IV

PENUTUP

4.1 Kesimpulan

Pada bab sebelumnya telah dibahas secara detil setiap rumusan masalah yang diteliti dalam skripsi ini. Berdasarkan pembahasan tersebut maka dapat ditarik kesimpulan yaitu langkah-langkah dalam pengelompokan *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean* antara lain:

- (a) Masukkan data amatan X_{ij} dengan $i = 1, \dots, n$ dan $j = 1, \dots, n$.
- (b) Hitung d_{ij} untuk mencari pusat antar objek amatan, dengan persamaan berikut:

$$d_{ij} = \sum_{q=1}^p (x_{iq} - x_{jq})^2.$$

- (c) Bentuk matriks d_{ij} .
- (d) Tentukan nilai minimum d_{ij} . Jika d_{ij} minimum maka proses berlanjut ke tahap 4. Jika d_{ij} tidak minimum, maka proses mengulang dari tahap 2.
- (e) Gabungkan objek amatan yang nilai $d_{ij} = 0$ menjadi satu kelompok.
- (f) Bentuk sebanyak k kelompok objek amatan.
- (g) Tentukan pusat $c_{kj} = \frac{\sum_{j=1}^p x_{ij}}{p}$ dengan peubah sebanyak p .

- (h) Hitung $d_{ik} = \sum_{j=1}^p (x_{ij} - c_{kj})^2$.
- (i) Kelompokkan berdasarkan minimum d_{ik} .
- (j) Tentukan nilai c_{kj} . Jika c_{kj} berubah maka proses berlanjut ke tahap 11. Jika c_{kj} tidak berubah, maka proses mengulang dari tahap 8.
- (k) Pusat c_{kj} berubah.
- (l) Identifikasi data amatan. Jika data amatan berubah posisi maka proses mengulang dari tahap 7. Jika data amatan tidak berubah, maka proses berlanjut ke tahap 13.
- (m) Didapat *Cluster* tetap yang memiliki proporsi terbaik.

Penerapan metode *Hybrid Mutual Clustering* ini pada kasus Penduduk Buta Aksara di Indonesia tahun 2015 didapatkan hasil bahwa proporsi untuk komponen utama terbesar terdapat pada variabel umur di bawah 14 tahun. Proporsi tujuannya melihat variabel mana yang memiliki pengaruh lebih besar dari yang lainnya, proporsi juga memengaruhi jarak yang diperoleh dari setiap provinsi yang memiliki karakteristik yang sama terhadap variabel umur. Anggota sebuah *cluster* tentu mempunyai kemiripan satu dengan yang lain, dan mereka tentu juga berbeda dengan anggota *cluster* yang lain. Pada kasus ini, terlihat kota Nusa Tenggara Barat dan Papua mempunyai karakteristik yang berbeda dengan kota lain.

4.2 Saran

Penelitian yang dilakukan dalam skripsi ini dapat dikembangkan lebih lanjut sehingga akan lebih bermanfaat dalam dunia statistika. Berikut

beberapa saran yang dapat digunakan untuk penelitian selanjutnya yaitu:

- Penerapan dari metode *Hybrid Hierarchical Clustering* dapat menggunakan jarak selain *Square Euclidean*, antara lain: jarak *Pearson*, jarak *Mahalanobis*
- Dalam metode pengelompokan, data yang digunakan harus sesuai dengan karakteristik dalam metode tersebut, untuk mendapatkan hasil yang sesuai.

DAFTAR PUSTAKA

- Agustina, Mitakda, dan Solimun. 2013. "Pemilihan Metode Pengelompokan Terbaik Kabupaten/Kota Berdasarkan Indikator Pendidikan Menggunakan Hybrid Melalui Mutual Cluster, Bottom-up, dan Top-down", *Jurnal Mahasiswa Statistik Universitas Brawijaya-Malang Vol. 1 No. 3*. Hal. 205-208.
- Bikriyah. 2014. "Analisis Hybrid Hierarchical Clustering Melalui Mutual Cluster, Bottom-up, dan Top-down Menggunakan Jarak Euclidean dan Mahalanobis", *Jurnal Mahasiswa Statistik Universitas Brawijaya-Malang Vol. 2 No. 5*. Hal. 397-400.
- Chipman dan Tibshirani. 2006. Hybrid Hierarchical Clustering With Applications To Microarray Data, *Biostatistics Journal-Oxford England*. Hal. 286-301.
- Hair, dkk. 2010. *Multivariate Data Analysis, Seventh Ed*. New Jersey: Prentice Hall International, Inc.
- Johnson dan Wichern. 2002. *Applied Multivariate Analysis, Fifth Edition*. New Jersey: Prentice Hall, Inc.
- Manly. 1988. *Multivariate Statistical Methods*. New York: Chapman Hall.
- Mariyani, dkk. 2011. "Penerapan *Hybrid Hierarchical Clustering* melalui *Mutual Cluster* dalam Pengelompokan Kabupaten di Jawa Timur berdasarkan Variabel Sektor Pertanian". *Jurnal Mahasiswa Statistik FMIPA Institut Sepuluh November Surabaya*.
- Santoso. 2010. *Statistik Multivariat*. Jakarta: PT Elex Media Komputindo.
- Walpole. 1995. *Pengantar Statistika*. Edisi ke-5. Jakarta: Terjemahan

Bambang Sumantri, Gramedia.

———. Data Penduduk Buta Aksara menurut Provinsi tahun 2015. [ON LINE] <http://data.go.id/dataset/persentase-penduduk-buta-aksara/resource/3e89671a-3199-4716-ba05-b39e00162b54> (diunduh pada tanggal 8 Oktober 2016, pukul 19:04)

LAMPIRAN-LAMPIRAN

NO.	KOTA	umur < 14	umur 15-44	umur > 45
1	ACEH	2.37	0.27	7.73
2	SUMATERA UTARA	1.32	0.51	3.08
3	SUMATERA BARAT	1.44	0.32	3.52
4	RIAU	1.13	0.33	3.42
5	JAMBI	2.16	0.49	6.06
6	SUMATERA SELATAN	1.78	0.48	4.73
7	BENGKULU	2.37	0.48	6.77
8	LAMPUNG	3.33	0.34	9.52
9	KEP. BANGKA BELITUNG	2.37	0.87	5.86
10	KEP. RIAU	1.21	0.29	4.42
11	DKI JAKARTA	0.41	0.06	1.26
12	JAWA BARAT	1.99	0.29	5.45
13	JAWA TENGAH	6.88	0.50	16.10
14	DI YOGYAKARTA	5.50	0.19	12.80
15	JAWA TIMUR	8.53	1.24	19.24
16	BANTEN	2.63	0.33	8.69
17	BALI	7.23	0.61	18.31
18	NUSA TENGGARA BARAT	13.03	3.31	33.78
19	NUSA TENGGARA TIMUR	8.55	3.10	19.47
20	KALIMANTAN BARAT	7.68	2.00	20.78
21	KALIMANTAN TENGAH	1.12	0.30	3.32
22	KALIMANTAN SELATAN	1.79	0.19	5.40
23	KALIMANTAN TIMUR	1.31	0.13	4.34
24	KALIMANTAN UTARA	5.01	1.36	14.89
25	SULAWESI UTARA	0.37	0.17	0.71
26	SULAWESI TENGAH	2.66	0.91	6.42
27	SULAWESI SELATAN	8.71	2.22	21.34
28	SULAWESI TENGGARA	5.90	1.37	17.07
29	GORONTALO	1.76	0.61	4.35
30	SULAWESI BARAT	7.36	3.33	17.37
31	MALUKU	1.15	0.80	1.96
32	MALUKU UTARA	1.51	0.47	4.28
33	PAPUA BARAT	3.12	2.09	6.32
34	PAPUA	29.17	28.47	31.57

Gambar 4.1: Tabel Data Persentase Penduduk Buta Aksara menurut Provinsi di Indonesia Tahun 2015 (sumber: Badan Pusat Statistik Tahun 2015).

Case	Squared Euclidean Distance																
	1:ACEH	2:SUMATERA UTARA	3:SUMATERA BARAT	4:RIAU	5:JAMBI	6:SUMATERA SELATAN	7:BENGKULU	8:LAMPUNG	9:KEP. BANGKA BELITUNG	10:KEP. RIAU	11:DKI JAKARTA	12:JAWA BARAT	13:JAWA TENGAH	14:DI YOGYAKARTA	15:JAWA TIMUR	16:BANTEN	17:BALI
1:ACEH	,000	,342	,277	,313	,042	,139	,015	,077	,084	,200	,720	,077	,1891	,702	,3216	,015	,2392
2:SUMATERA UTARA	,342	,000	,005	,004	,148	,045	,228	,721	,152	,028	,084	,096	,3445	,1933	,5483	,500	,4456
3:SUMATERA BARAT	,277	,005	,000	,004	,109	,026	,179	,627	,120	,013	,111	,063	,3243	,1778	,5241	,422	,4226
4:RIAU	,313	,004	,004	,000	,135	,040	,211	,688	,149	,014	,086	,083	,3400	,1896	,5442	,466	,4396
5:JAMBI	,042	,148	,109	,135	,000	,030	,009	,216	,008	,071	,436	,008	,2185	,1028	,3867	,105	,2992
6:SUMATERA SELATAN	,139	,045	,026	,040	,030	,000	,070	,405	,037	,014	,241	,010	,2713	,1396	,4555	,245	,3611
7:BENGKULU	,015	,228	,179	,211	,009	,070	,000	,139	,018	,126	,565	,031	,1926	,854	,3521	,055	,2684
8:LAMPUNG	,077	,721	,627	,688	,216	,405	,139	,000	,231	,520	,1253	,294	,1046	,316	,2299	,027	,1613
9:KEP. BANGKA BELITUNG	,064	,152	,120	,149	,008	,037	,018	,231	,000	,091	,458	,022	,2180	,1034	,3830	,127	,2990
10:KEP. RIAU	,200	,028	,013	,014	,071	,126	,126	,520	,091	,000	,164	,036	,3029	,1624	,4977	,325	,3962
11:DKI JAKARTA	,720	,084	,111	,086	,436	,241	,565	,1253	,458	,164	,000	,334	,4544	,2764	,6875	,945	,5893
12:JAWA BARAT	,077	,096	,063	,083	,008	,010	,031	,294	,022	,036	,334	,000	,2420	,1185	,4187	,161	,3271
13:JAWA TENGAH	1,691	3,445	3,243	3,400	2,185	2,713	1,926	1,046	2,180	3,029	4,544	2,420	,000	,223	,256	1,399	,073
14:DI YOGYAKARTA	7,02	1,933	1,778	1,896	1,028	1,396	,316	,316	1,034	1,624	2,764	1,185	,223	,000	,947	,525	,535
15:JAWA TIMUR	3,216	5,483	5,241	5,442	3,867	4,555	3,521	2,299	3,830	4,977	6,875	4,187	,256	,947	,000	2,806	,088
16:BANTEN	,015	,500	,422	,466	,105	,245	,055	,027	,127	,325	,945	,161	,181	,525	2,806	,000	2,034
17:BALI	2,392	4,456	4,226	4,396	2,992	3,611	2,684	1,613	2,990	3,962	5,693	3,271	,073	,535	,088	2,034	,000
18:NUSA TENGGARA BARAT	13,835	18,274	17,850	18,187	15,189	16,538	14,492	11,877	15,099	17,299	20,770	15,846	6,021	8,539	3,840	12,944	4,827
19:NUSA TENGGARA TIMUR	3,604	5,864	5,649	5,849	4,230	4,929	3,882	2,663	4,134	5,384	7,340	4,587	,547	1,311	,150	3,178	,347
20:KALIMANTAN BARAT	3,490	5,878	5,637	5,822	4,185	4,908	3,822	2,548	4,144	5,322	7,323	4,535	,425	1,195	,083	3,050	,175
21:KALIMANTAN SELATAN	,326	,004	,004	,000	,144	,044	,222	,707	,159	,017	,079	,090	,3440	,1925	5,494	,482	4,443
22:KALIMANTAN TIMUR	,088	,087	,054	,071	,015	,010	,042	,321	,035	,026	,306	,002	2,507	1,245	4,308	,176	3,367
23:KALIMANTAN UTARA	,200	,028	,012	,015	,072	,015	,127	,519	,085	,002	,161	,034	3,020	1,613	4,973	,326	3,957
24:KALIMANTAN BARAT	1,010	2,451	2,294	2,406	1,403	1,837	1,196	,545	1,391	2,063	3,403	1,610	,175	,128	,699	,780	,360
25:SULAWESI UTARA	,827	,115	,151	,124	,515	,299	,666	1,390	,531	,217	,005	,406	4,790	2,959	7,167	1,067	5,973
26:SULAWESI TENGAH	,045	,225	,184	,222	,018	,075	,013	,164	,007	,146	,560	,045	1,937	,872	3,502	,086	2,705
27:SULAWESI SELATAN	4,153	6,685	6,432	6,842	4,885	5,657	4,487	3,113	4,826	6,120	8,233	5,261	,927	1,555	,104	3,679	,316
28:SULAWESI TENGGARA	1,704	3,494	3,302	3,440	2,212	2,750	1,949	1,071	2,188	3,051	4,613	2,467	,079	,320	,309	1,400	,108
29:GORONTALO	,177	,030	,017	,029	,047	,003	,085	,462	,048	,015	,210	,023	2,843	1,493	4,712	,292	3,764
30:SULAWESI BARAT	2,571	4,466	4,291	4,459	3,077	3,667	2,788	1,813	2,979	4,060	5,770	3,369	,376	,837	,285	2,222	,332
31:MALUKU	,528	,022	,047	,039	,274	,125	,379	,972	,264	,086	,050	,206	3,840	2,316	6,077	,717	5,021
32:MALUKU UTARA	1,93	,021	,009	,016	,059	,112	,499	,068	,005	,177	,029	,029	2,957	1,573	4,871	,316	3,890
33:PAPUA BARAT	,190	,367	,343	,389	,144	,210	,134	,276	,087	,318	,792	,195	1,937	,939	3,322	,220	2,689
34:PAPUA	67,372	72,192	72,071	72,721	68,383	70,102	67,514	64,286	67,219	71,901	76,562	69,629	54,470	59,019	49,010	66,115	52,778

Gambar 4.2: Tabel Matriks Prioritas Hasil Output SPSS.

Proximity Matrix

Squared Euclidean Distance

Case	18:NUSA TENGGARA BARAT	19:NUSA TENGGARA TIMUR	20:KALIMANTAN BARAT	21:KALIMANTAN TENGAH	22:KALIMANTAN SELATAN	23:KALIMANTAN TIMUR	24:KALIMANTAN UTARA	25:SULAWESI UTARA	26:SULAWESI TENGAH	27:SULAWESI SELATAN	28:SULAWESI ITENGGARA	29:GORON TALO	30:SULAWESI BARAT	31:MALUKU	32:MALUKU UTARA	33:PAPUA BARAT	34:PAPUA
1:ACEH	13,835	3,604	3,490	3,266	3,088	2,000	1,010	827	045	4,153	1,704	1,177	2,571	528	193	190	67,372
2:SUMATERA UTARA	18,274	5,864	5,878	004	087	028	2,451	115	225	6,885	3,494	030	4,486	022	021	367	72,192
3:SUMATERA BARAT	17,850	5,649	5,637	004	054	012	2,294	151	184	6,432	3,302	017	4,291	047	009	343	72,071
4:RIAU	18,187	5,849	5,822	000	071	015	2,406	124	222	6,642	3,440	029	4,459	039	016	389	72,731
5:JAMBI	15,189	4,230	4,185	015	015	072	1,403	515	018	4,885	2,212	047	3,077	274	059	144	68,383
6:SUMATERA SELATAN	16,538	4,929	4,908	044	010	015	1,837	299	075	5,657	2,750	003	3,667	125	005	210	70,102
7:BENGKULU	14,492	3,862	3,822	042	042	127	1,196	656	013	4,497	1,949	095	2,788	379	112	134	67,514
8:LAMPUNG	11,877	2,663	2,548	070	031	519	1,390	545	164	3,113	1,071	462	1,813	972	499	276	64,288
9:KEP. BANGKA BELITUNG	15,099	4,134	4,144	159	035	095	1,391	531	007	4,826	2,198	048	2,979	264	068	087	67,219
10:KEP. RIAU	17,299	5,384	5,322	017	026	026	2,083	217	146	6,120	3,051	015	4,060	096	005	318	71,901
11:DKI JAKARTA	20,770	7,340	7,323	079	306	181	3,403	005	580	8,233	4,613	210	5,770	050	177	792	76,562
12:JAWA BARAT	15,846	4,587	4,535	090	002	034	1,610	406	045	5,261	2,467	023	3,369	206	029	195	69,629
13:JAWA TENGAH	6,021	547	425	3,440	2,507	3,020	1,175	4,790	1,937	627	079	2,843	376	3,940	2,957	1,937	54,470
14:DIYOGYAKARTA	8,539	1,311	1,195	1,925	1,245	1,613	1,128	2,959	872	1,555	320	1,493	837	2,316	1,573	939	59,019
15:JAWA TIMUR	3,840	1,500	083	5,494	4,308	4,973	699	7,167	3,502	1,104	309	4,712	285	6,077	4,871	3,382	49,010
16:BANTEN	12,944	3,178	3,050	4,82	176	326	780	1,067	086	3,679	1,400	292	2,222	717	316	220	66,115
17:BALI	4,827	347	175	4,443	3,367	3,957	380	5,973	2,705	316	108	3,764	332	5,021	3,890	2,689	52,778
18:NUSA TENGGARA BARAT	000	3,558	3,432	16,288	16,068	17,325	7,389	21,278	14,446	2,861	5,834	16,832	4,878	19,324	17,123	14,011	36,484
19:NUSA TENGGARA TIMUR	3,558	000	1,103	5,907	4,725	5,407	862	7,618	3,795	083	455	5,068	113	6,419	5,250	3,486	44,686
20:KALIMANTAN BARAT	3,432	1,103	000	5,880	4,653	5,338	751	7,629	3,807	044	320	5,072	242	6,491	5,228	3,642	48,011
21:KALIMANTAN SELATAN	18,288	5,907	5,880	000	077	017	2,444	115	233	6,702	3,484	033	4,511	037	019	404	72,902
22:KALIMANTAN TIMUR	16,068	4,725	4,653	000	077	024	1,677	377	063	5,396	2,549	023	3,509	195	024	229	70,291
23:KALIMANTAN TIMUR	17,325	5,407	5,338	000	024	000	2,096	215	150	6,134	3,062	017	4,090	099	006	335	72,156
24:KALIMANTAN UTARA	7,389	862	751	2,444	1,677	2,096	000	3,617	1,202	1,091	094	1,942	447	2,865	2,032	1,171	56,015
25:SULAWESI UTARA	21,278	7,618	7,629	1,115	377	215	3,617	000	662	8,548	4,863	261	6,010	060	227	862	76,863
26:SULAWESI TENGAH	14,446	3,795	3,807	233	063	150	1,202	662	000	4,458	1,957	092	2,697	358	119	068	66,168
27:SULAWESI SELATAN	2,861	083	044	6,702	5,396	6,134	1,091	8,548	4,458	000	562	5,826	337	7,322	6,004	4,239	45,832
28:SULAWESI TENGGARA	5,834	455	320	3,484	2,549	3,062	084	4,863	1,957	562	000	2,880	242	3,985	2,989	1,903	53,565
29:GORONTALO	16,832	5,068	5,072	033	023	017	1,942	261	092	5,826	2,880	000	3,780	084	003	213	70,113
30:SULAWESI BARAT	4,878	4,878	242	4,511	3,509	4,090	447	6,010	2,697	337	242	3,780	000	4,936	3,939	2,397	46,725
31:MALUKU	19,324	6,419	6,491	037	368	195	3,649	060	358	7,322	3,985	094	4,936	000	084	473	72,736
32:MALUKU UTARA	17,123	5,250	5,228	019	024	006	2,032	227	119	6,004	2,989	003	3,939	084	000	262	70,966
33:PAPUA BARAT	14,011	3,486	3,642	4,04	229	335	1,171	862	068	4,239	1,903	213	2,397	084	262	000	62,665
34:PAPUA	36,484	44,686	48,011	72,902	70,291	72,156	56,015	76,863	66,168	45,832	53,565	70,113	46,725	72,736	70,966	62,665	000

This is a dissimilarity matrix

Gambar 4.3: Tabel Matriks Prioritas Hasil Output SPSS (lanjutan).

KOTA	QCL 1	QCL 2
ACEH	2	0,35529
SUMATERA UTARA	2	0,23098
SUMATERA BARAT	2	0,17781
RIAU	2	0,21525
JAMBI	2	0,15466
SUMATERA SELATAN	2	0,01915
BENGKULU	2	0,2472
LAMPUNG	2	0,6188
KEP. BANGKA BELITUNG	2	0,17695
KEP. RIAU	2	0,12816
DKI JAKARTA	2	0,50793
JAWA BARAT	2	0,08652
JAWA TENGAH	3	0,47176
DI YOGYAKARTA	3	0,91632
JAWA TIMUR	3	0,19243
BANTEN	2	0,47648
BALI	3	0,27124
NUSA TENGGARA BARAT	3	2,01867
NUSA TENGGARA TIMUR	3	0,32696
KALIMANTAN BARAT	3	0,1941
KALIMANTAN TENGAH	2	0,2269
KALIMANTAN SELATAN	2	0,08517
KALIMANTAN TIMUR	2	0,13306
KALIMANTAN UTARA	3	0,71619
SULAWESI UTARA	2	0,56412
SULAWESI TENGAH	2	0,25883
SULAWESI SELATAN	3	0,33393
SULAWESI TENGGARA	3	0,42312
GORONTALO	2	0,06893
SULAWESI BARAT	3	0,39706
MALUKU	2	0,37162
MALUKU UTARA	2	0,08887
PAPUA BARAT	2	0,44882
PAPUA	1	0

Gambar 4.4: Tabel QCL Hasil *Output SPSS*.

SURAT PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya yang bertanda tangan di bawah ini, mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta:

Nama : Astrid Alfira

No. Registrasi : 3125121985

Program Studi : Matematika

Menyatakan bahwa skripsi ini yang saya buat dengan judul "**Analisis *Hybrid Hierarchical Clustering* menggunakan Jarak *Square Euclidean***" adalah :

1. Dibuat dan diselesaikan oleh saya sendiri.
2. Bukan merupakan duplikat skripsi yang pernah dibuat oleh orang lain atau jiplakan karya tulis orang lain.

Pernyataan ini dibuat dengan sesungguhnya dan saya bersedia menanggung segala akibat yang timbul jika pernyataan saya tidak benar.

Jakarta, Februari 2017

Yang membuat pernyataan

Astrid Alfira

DAFTAR RIWAYAT HIDUP

DAFTAR RIWAYAT HIDUP

Astrid Alfira. Lahir di Jakarta, 2 Agustus 1995. Anak kedua dari pasangan Bapak Eddie Sutriyono dan Ibu Iswanti. Saat ini bertempat tinggal di Komplek Pondok Benda Residence blok A1/3, Pamulang, Tangerang Selatan, Banten, 15416.

No. Ponsel : +6281290223143

Email : astridalfiraaa@gmail.com

Riwayat Pendidikan: Penulis mengawali pendidikan di TK Marita, Tangerang pada tahun 2000-2001, kemudian melanjutkan di SDN Joglo 5 Pagi, Jakarta pada tahun 2001-2007. Setelah itu, penulis melanjutkan di SMPN 19 Jakarta pada tahun 2007-2009. Kemudian penulis kembali melanjutkan di SMAN 47 Jakarta pada tahun 2009-2012. Setelah lulus SMA, penulis melanjutkan ke Universitas Negeri Jakarta (UNJ), program studi Matematika, melalui jalur SNMPTN Tulis. Pada pertengahan tahun 2017 penulis memperoleh gelar Sarjana Sains untuk Program Studi Matematika, FMIPA, UNJ.

Riwayat Organisasi: Selama kuliah, penulis terlibat kegiatan di BEM jurusan, seperti menjadi panitia keamanan dalam kegiatan Calculus Cup serta aktif dalam berbagai kegiatan BEM lainnya. Di luar itu, penulis pernah menjadi wakil ketua dari Organisasi Tari Saman di SMAN 47 Jakarta dan pernah menjadi Juara 1 tingkat Jakarta di tahun 2011 serta menjuarai beberapa lomba saman yang diadakan oleh SMA di Jakarta.

Riwayat Pekerjaan: Penulis menjadi pengajar privat untuk pelajaran matematika dan IPA sejak semester 5. Pada tahun 2015 penulis pernah menjalani praktik kerja lapangan di Bank Indonesia pada Divisi Statistik Data Sekunder, Grup Statistik Domestik, Departemen Statistik.

