

BAB III

LANDASAN TEORI

3.1. Pertanian

Pertanian adalah kegiatan pemanfaatan sumberdaya hayati yang dilakukan manusia untuk menghasilkan bahan pangan, bahan baku industri, atau sumber energi, serta untuk mengelola lingkungan hidupnya. Kegiatan pemanfaatan sumberdaya hayati yang termasuk dalam pertanian biasa dipahami orang sebagai budidaya tanaman atau bercocok tanam (bahasa Inggris: *crop cultivation*) serta pembesaran hewan ternak (*raising*), meskipun cakupannya dapat pula berupa pemanfaatan mikroorganisme dan bioenzim dalam pengolahan produk lanjutan, seperti pembuatan keju dan tempe, atau sekedar ekstraksi semata, seperti penangkapan ikan atau eksploitasi hutan.

Bagian terbesar penduduk dunia bermata pencaharian dalam bidang-bidang di lingkup pertanian, namun pertanian hanya menyumbang 4% dari PDB dunia. Sejarah Indonesia sejak masa kolonial sampai sekarang tidak dapat dipisahkan dari sektor pertanian dan perkebunan, karena sektor-sektor ini memiliki arti yang sangat penting dalam menentukan pembentukan berbagai realitas ekonomi dan sosial masyarakat di berbagai wilayah Indonesia. Berdasarkan data BPS tahun 2002, bidang pertanian di Indonesia menyediakan lapangan kerja bagi sekitar 44,3% penduduk meskipun hanya menyumbang sekitar 17,3% dari total pendapatan domestik bruto.

Kelompok ilmu-ilmu pertanian mengkaji pertanian dengan dukungan ilmu-ilmu pendukungnya. Karena pertanian selalu terikat dengan ruang dan waktu, ilmu-ilmu pendukung, seperti ilmu tanah, meteorologi, teknik pertanian, biokimia, dan statistika juga dipelajari dalam pertanian. Usaha tani (*farming*) adalah bagian inti dari pertanian karena menyangkut sekumpulan kegiatan yang dilakukan dalam budidaya. Petani adalah sebutan bagi mereka yang menyelenggarakan usaha tani, sebagai contoh "petani tembakau" atau "petani ikan". Pelaku budidaya hewan ternak (*livestock*) secara khusus disebut sebagai peternak.

3.2. Cangkupan Pertanian

Pertanian dalam pengertian yang luas mencakup semua kegiatan yang melibatkan pemanfaatan makhluk hidup (termasuk tanaman, hewan, dan mikrobia) untuk kepentingan manusia. Dalam arti sempit, pertanian diartikan sebagai kegiatan pembudidayaan tanaman. Usaha pertanian diberi nama khusus untuk subjek usaha tani tertentu. Kehutanan adalah usaha tani dengan subjek tumbuhan (biasanya pohon) dan diusahakan pada lahan yang setengah liar atau liar (hutan). Peternakan menggunakan subjek hewan darat kering (khususnya semua vertebrata kecuali ikan dan amfibia) atau serangga (misalnya lebah). Perikanan memiliki subjek hewan perairan (termasuk amfibia dan semua non-vertebrata air). Suatu usaha pertanian dapat melibatkan berbagai subjek ini bersama-sama dengan alasan efisiensi dan peningkatan keuntungan. Pertimbangan akan kelestarian lingkungan mengakibatkan aspek-aspek konservasi sumber daya alam juga menjadi bagi andalam usaha pertanian.

Semua usaha pertanian pada dasarnya adalah kegiatan ekonomi sehingga memerlukan dasar-dasar pengetahuan yang sama akan pengelolaan tempat usaha, pemilihan benih/ bibit, metode budidaya, pengumpulan hasil, distribusi produk, pengolahan dan pengemasan produk, dan pemasaran. Apabila seorang petani memandang semua aspek ini dengan pertimbangan efisiensi untuk mencapai keuntungan maksimal maka ia melakukan pertanian intensif (*intensive farming*). Usaha pertanian yang dipandang dengan cara ini dikenal sebagai agribisnis. Program dan kebijakan yang mengarahkan usaha pertanian ke cara pandang demikian dikenal sebagai *intensifikasi*. Karena pertanian industri selalu menerapkan pertanian intensif, keduanya sering kali disamakan.

Sisi pertanian industrial yang memperhatikan lingkungannya adalah pertanian berkelanjutan (*sustainable agriculture*). Pertanian berkelanjutan, dikenal juga dengan variasinya seperti pertanian organik atau permakultur, memasukkan aspek kelestarian daya dukung lahan maupun lingkungan dan pengetahuan lokal sebagai faktor penting dalam perhitungan efisiensinya. Akibatnya, pertanian

berkelanjutan biasanya memberikan hasil yang lebih rendah dari pada pertanian industrial.

Pertanian modern masa kini biasanya menerapkan sebagian komponen dari kedua kutub ideologi pertanian yang disebutkan di atas. Selain keduanya, dikenal pula bentuk pertanian ekstensif (pertanian masukan rendah) yang dalam bentuk paling ekstrem dan tradisional akan berbentuk pertanian subsisten, yaitu hanya dilakukan tanpa motif bisnis dan semata hanya untuk memenuhi kebutuhan sendiri atau komunitasnya.

Sebagai suatu usaha, pertanian memiliki dua ciri penting: selalu melibatkan barang dalam volume besar dan proses produksi memiliki risiko yang relatif tinggi. Dua ciri khas ini muncul karena pertanian melibatkan makhluk hidup dalam satu atau beberapa tahapnya dan memerlukan ruang untuk kegiatan itu serta jangka waktu tertentu dalam proses produksi. Beberapa bentuk pertanian modern (misalnya budidaya alga, hidroponik) telah dapat mengurangi ciri-ciri ini tetapi sebagian besar usaha pertanian dunia masih tetap demikian.

3.3. Subsektor Tanaman Pangan

Pangan diartikan sebagai segala sesuatu yang bersumber dari sumber hayati dan air, baik yang diolah maupun yang tidak diolah. Pangan diperuntukkan bagi konsumsi manusia sebagai makanan atau minuman, termasuk bahan tambahan pangan, bahan baku pangan, dan bahan-bahan lain yang digunakan dalam proses penyiapan, pengolahan, dan atau pembuatan makanan atau minuman.

Komoditas pangan harus mengandung zat gizi yang terdiri atas karbohidrat, protein, lemak, vitamin, dan mineral yang bermanfaat bagi pertumbuhan dan kesehatan manusia. Kelompok tanaman budidaya yang tergolong komoditas ini meliputi kelompok tanaman pangan, tanaman hortikultura nontanaman hias, dan kelompok tanaman lain penghasil bahan baku produk yang memenuhi batasan pangan.

Batasan untuk tanaman pangan adalah kelompok tanaman sumber karbohidrat dan protein. Namun, secara sempit, tanaman pangan biasanya dibatasi pada kelompok tanaman yang berumur semusim. Batasan ini di masa mendatang

harus diperbaiki karena akan menyebabkan sumber karbohidrat menjadi terbatas. Tanaman pangan sebaiknya memasukkan jenis tanaman lain yang dapat menjadi sumber karbohidrat tanpa dibatasi pada kelompok tanaman semusim. Dengan perbaikan batasan ini, tanaman umbian selain ubi kayu, ubi jalar, dan talas dapat masuk ke dalam kelompok tanaman pangan, misalnya garut, ganyong, dan kimpul. Demikian juga dengan buah yang merupakan sumber karbohidrat dapat masuk ke dalam tanaman pangan, misalnya sukun.

3.4. Subsektor Hortikultural

Hortikultura (horticulture) berasal dari bahasa Latin *hortus* (tanaman kebun) dan *cultura/ colere* (budidaya), dan dapat diartikan sebagai budidaya tanaman kebun. Kemudian hortikultura digunakan secara lebih luas bukan hanya untuk budidaya di kebun. Istilah hortikultura digunakan pada jenis tanaman yang dibudidayakan. Bidang kerja hortikultura meliputi pembenihan, pembibitan, kultur jaringan, produksi tanaman, hama dan penyakit, panen, pengemasan dan distribusi. Hortikultura merupakan salah satu metode budidaya pertanian modern. Hortikultura merupakan cabang dari agronomi. Berbeda dengan agronomi, hortikultura memfokuskan pada budi daya tanaman buah (pomologi/ frutikultur), tanaman bunga (florikultura), tanaman sayuran (olerikultura), tanaman obat-obatan (biofarmaka), dan taman (lansekap). Salah satu ciri khas produk hortikultura adalah perisabel atau mudah rusak karena segar. Orang yang menekuni bidang hortikultura dengan profesional disebut sebagai hortikultoris.

Indonesia yang terletak di tropis membuat Indonesia menjadi surga biodiversitas komoditas hortikultura.

- a. Pomologi/ Frutikultur : Manggis, Mangga, Apel, Durian, Salak, dan lain sebagainya.
- b. Florikultura : Melati, Mawar, Krisan, Anyelir, Begonia, Bugenvil, dan lain sebagainya.
- c. Olerikultura : Tomat, Selada, Bayam, Wortel, Kentang, (Melon & Semangka: termasuk kelompok tanaman sayuran yang di panen buahnya) dan lain sebagainya.
- d. Biofarmaka : Purwoceng, Rosela, Kunyit, dan lain sebagainya.

- e. Lansekap : Taman Bali, Taman Jawa, dan lain sebagainya.
- f. Puslitbang Hortikultura melakukan penelitian pada komoditas utama, yaitu :
- g. Sayuran : cabai merah, bawang merah, kentang, tomat, kubis, kacang panjang, bawang putih, dan sayuran asli Indonesia (*indigenous*).
- h. Buah : jeruk, pisang, mangga, manggis, durian, pepaya, salak, nenas, semangka, melon, apel, anggur, markisa, jambu, kesemek, rambutan, dan alpokat.
- i. Tanaman Hias : anggrek, lili, mawar, anyelir, krisan, sedap malam, dracaena, dan tanaman hias tropika.

3.5. Subsektor Perkebunan

Perkebunan adalah segala kegiatan yang mengusahakan tanaman tertentu pada tanah dan/ atau media tumbuh lainnya dalam ekosistem yang sesuai mengolah, dan memasarkan barang dan jasa hasil tanaman tersebut, dengan bantuan ilmu pengetahuan dan teknologi, permodalan serta manajemen untuk mewujudkan kesejahteraan bagi pelaku usaha perkebunan dan masyarakat. Tanaman yang ditanam bukanlah tanaman yang menjadi makanan pokok maupun sayuran untuk membedakannya dengan usaha ladang dan hortikultura sayur-mayur dan bunga, meski usaha penanaman pohon buah masih disebut usaha perkebunan. Tanaman yang ditanam umumnya berukuran besar dengan waktu penanaman yang relatif lama, antara kurang dari setahun hingga tahunan.

Perkebunan dibedakan dari agroforestri dan silvikultur (budidaya hutan) karena sifat intensifnya. Dalam perkebunan pemeliharaan memegang peranan penting, sementara dalam agroforestri dan silvikultur, tanaman cenderung dibiarkan untuk tumbuh sesuai kondisi alam. Karena sifatnya intensif, perkebunan hampir selalu menerapkan cara budidaya monokultur, kecuali untuk komoditas tertentu, seperti lada dan vanili. Penciri sekunder yang tidak selalu berlaku adalah adanya instalasi pengolahan atau pengemasan terhadap hasil panen dari lahan perkebunan itu, sebelum produknya dipasarkan. Perkebunan dibedakan dari usaha tani pekarangan terutama karena skala usaha dan pasar produknya.

3.6. Subsektor Peternakan

Peternakan adalah kegiatan mengembangbiakkan dan membudidayakan hewan ternak untuk mendapatkan manfaat dan hasil dari kegiatan tersebut. Pengertian peternakan tidak terbatas pada pemeliharaan saja, memelihara dan peternakan perbedaannya terletak pada tujuan yang ditetapkan. Tujuan peternakan adalah mencari keuntungan dengan penerapan prinsip-prinsip manajemen pada faktor-faktor produksi yang telah dikombinasikan secara optimal.

Kegiatan di bidang peternakan dapat dibagi atas dua golongan, yaitu peternakan hewan besar seperti sapi, kerbau dan kuda, sedang kelompok kedua yaitu peternakan hewan kecil seperti ayam, kelinci dan lain sebagainya.

3.7. Subsektor Perikanan

Perikanan adalah kegiatan manusia yang berhubungan dengan pengelolaan dan pemanfaatan sumberdaya hayati perairan. Sumberdaya hayati perairan tidak dibatasi secara tegas dan pada umumnya mencakup ikan, amfibi, dan berbagai avertebrata penghuni perairan dan wilayah yang berdekatan, serta lingkungannya. Di Indonesia, menurut UU RI no. 9/1985 dan UU RI no. 31/2004, kegiatan yang termasuk dalam perikanan dimulai dari praproduksi, produksi, pengolahan sampai dengan pemasaran, yang dilaksanakan dalam suatu sistem bisnis perikanan. Dengan demikian, perikanan dapat dianggap merupakan usaha agribisnis. Umumnya, perikanan dimaksudkan untuk kepentingan penyediaan pangan bagi manusia. Selain itu, tujuan lain dari perikanan meliputi olahraga, rekreasi (pemancingan ikan), dan mungkin juga untuk tujuan membuat perhiasan atau mengambil minyak ikan.

Usaha perikanan adalah semua usaha perorangan atau badan hukum untuk menangkap atau membudidayakan (usaha penetasan, pembibitan, pembesaran) ikan, termasuk kegiatan menyimpan, mendinginkan, pengeringan, atau mengawetkan ikan dengan tujuan untuk menciptakan nilai tambah ekonomi bagi pelaku usaha (komersial/ bisnis).

3.8. Subsektor Kehutanan

Kehutanan adalah suatu praktik untuk membuat, mengelola, menggunakan dan melestarikan hutan untuk kepentingan manusia. Undang-Undang Republik

Indonesia No 41 tahun 1999 tentang kehutanan, definisi kehutanan adalah sistem pengurusan yang bersangkutan paut dengan hutan, kawasan hutan, dan hasil hutan yang diselenggarakan secara terpadu.

Menurut Simon (1998), perkembangan teori pengelolaan hutan dapat dikelompokkan ke dalam dua kategori, yaitu kategori kehutanan konvensional dan kategori kehutanan modern (kehutanan sosial). Teori pengelolaan hutan yang termasuk ke dalam kehutanan konvensional adalah penambangan kayu atau timber extraction (TE) dan perkebunan kayu atau timber management (TM). Kehutanan sosial adalah pengelolaan hutan sebagai sumberdaya atau forest resource management (FRM) dan pengelolaan hutan sebagai ekosistem atau forest ecosystem management (FEM). Keduanya disebut juga dengan istilah lain Sustainable Forestry Management (SFM). Ketiga teori pengelolaan hutan tersebut, secara evolutif berkembang, sejak dari mulai penambangan kayu (TE) hingga sampai pada pengelolaan ekosistem hutan (FEM).

3.9. Usaha Pertanian

Usaha pertanian adalah kegiatan yang menghasilkan produk pertanian dengan tujuan sebagian atau seluruh hasil produksi dijual/ ditukar atas risiko usaha (bukan buruh tani atau pekerja keluarga). Usaha pertanian meliputi usaha tanaman pangan, hortikultura, perkebunan, peternakan, perikanan, dan kehutanan, termasuk jasa pertanian. Khusus tanaman pangan (padi dan palawija) meskipun tidak untuk dijual (dikonsumsi sendiri) tetap dicakup sebagai usaha (BPS, 2013).

3.10. Rumah Tangga Usaha Pertanian

Rumah tangga usaha pertanian adalah rumah tangga yang salah satu atau lebih anggota rumah tangganya mengelola usaha pertanian dengan tujuan sebagian atau seluruh hasilnya untuk dijual, baik usaha pertanian milik sendiri, secara bagi hasil, atau milik orang lain dengan menerima upah, dalam hal ini termasuk jasa pertanian (BPS, 2013).

3.11. Perusahaan Pertanian Berbadan Hukum

Perusahaan pertanian berbadan hukum adalah setiap bentuk usaha yang menjalankan jenis usaha di sektor pertanian yang bersifat tetap, terus menerus yang didirikan dengan tujuan memperoleh laba yang pendirian perusahaan dilindungi

hukum atau izin dari instansi yang berwenang minimal pada tingkat kabupaten/kota, untuk setiap kegiatan budidaya pertanian seperti penanaman, pemupukan, pemeliharaan, dan pemanenan. Contoh bentuk badan hukum: PT, CV, Koperasi, Yayasan, SIP Pemuda (BPS, 2013).

3.12. Perusahaan Tidak Berbadan Hukum

Perusahaan tidak berbadan hukum adalah usaha pertanian yang dikelola oleh bukan badan hukum dan bukan oleh rumah tangga seperti, pesantren, seminari, kelompok usaha bersama, tanksi militer, lembaga masyarakat, lembaga pendidikan, dan lain-lain yang mengusahakan pertanian (BPS, 2013).

3.13. Analisis Cluster

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

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

1. *Agromeration Schedule*, merupakan daftar yang memberikan informasi tentang obyek atau kasus yang akan dikelompokkan di setiap tahap pada proses analisis *cluster* dengan metode hirarki.
2. Rata-rata *cluster* (*Cluster Centroid*), adalah nilai rata-rata variabel dari semua obyek atau observasi dalam *cluster* tertentu.
3. Pemusatan *cluster* (*Cluster Centrsers*), adalah titik awal dimulai pengelompokan di dalam *cluster* non hirarki.
4. Keanggotaan *cluster* adalah keanggotaan yang menunjukan *cluster* untuk tiap obyek yang menjadi anggotanya.
5. *Dendogram* yaitu suatu alat grafis untuk menyajikan hasil dari analisis *cluster* yang dilakukan oleh peneliti. *Dendogram* berguna untuk menunjukan anggota *cluster* yang ada jika akan ditentukan berapa *cluster* yang seharusnya dibentuk.
6. Jarak antara pusat *cluster* (*Distance Between Cluster Center*) merupakan jarak yang menunjukan bagaimana terpisahnya pasangan individu.

3.14. Prosedur Analisis Cluster

Dalam menganalisis suatu data menggunakan analisis *cluster* diperlukan beberapa proses yang harus dilakukan yaitu:

1. Standarisasi Data

Proses standarisasi dilakukan apabila diantara variabel-variabel yang diteliti terdapat perbedaan ukuran satuan yang besar. Perbedaan satuan yang mencolok dapat mengakibatkan perhitungan pada analisis *cluster* menjadi tidak valid. Untuk itu, perlu dilakukan proses standarisasi dengan melakukan transformasi (standarisasi) pada data asli sebelum dianalisis lebih lanjut. Transformasi dilakukan terhadap variabel yang relevan ke dalam bentuk *z-scor*, sebagai berikut:

$$z = \frac{x - \bar{x}}{s} \quad (3.1)$$

dengan, x : nilai data
 \bar{x} : nilai rata-rata
 s : standar deviasi

2. Menentukan Jumlah Cluster

Menurut Subhash (1996), untuk menentukan jumlah *cluster* yang paling tepat, saat menggunakan metode *k-means* dapat dilakukan dengan beberapa cara. Salah satunya adalah dengan cara manual dengan belandaskan teori terkait data yang digunakan. Selain itu ada beberapa cara yang lain juga yang dapat digunakan yaitu seperti *Elbow Criterion (RMSSDT dan RS)*. *Elbow criterion* adalah salah satu cara untuk menentukan jumlah cluster yang paling tepat untuk pemodelan *k-means*. *Elbow criterion* untuk *k-means* ini mengkombinasikan antara nilai *RMSSTD* dan *RS statistics*, dimana cluster yang paling tepat untuk suatu dataset ditentukan apabila perbedaan nilai antara *RMSSTD* dan *RS* menjadi berbanding terbalik dengan keadaan sebelumnya.

RMSSTD (Root Means Square Standard Deviation) merupakan alat ukur tingkat kemiripan (*homogeneity*) data yang terdapat di dalam *cluster* yang ditemukan (*within clusters*). Makin rendah nilai *RMSSTD* makin mirip data di dalam cluster yang ditemukan. *RMSSDT* dihitung menggunakan rumus sebagai berikut:

$$RMSSTD = SQRT \left(\frac{SUM(i=1 \text{ to } k)SUM(j=1 \text{ to } d) \left(SUM(k=1 \text{ to } N_i) \left((x_{kj} - \mu_j)^2 \right) \right)}{SUM(i=1 \text{ to } k)SUM(j=1 \text{ to } d)(N_i - 1)} \right) \quad (3.2)$$

RS (R Squared) digunakan untuk mengukur tingkat kesamaan atau ketidaksamaan antara *cluster* (*between clusters*). RS mempunyai nilai antara 0 dan 1. Nilai 0 untuk *cluster* yang sama dan 1 untuk *cluster* yang benar-benar berbeda. RS dihitung dengan rumus:

$$RS = (SS_t - SS_w) / SS_t \quad (3.3)$$

$$SS_t = SUM(j = 1 \text{ to } d) (SUM(k = 1 \text{ to } N) \left((x_{kj} - \mu_j)^2 \right))$$

$$SS_w = SUM(i = 1 \text{ to } k) SUM(j = 1 \text{ to } d) \left(SUM(k = 1 \text{ to } N_i) \left((x_{kj} - \mu_j)^2 \right) \right)$$

Notasi:

x_{kj} : data ke- k yang ada di dalam cluster untuk dimensi ke- j

μ_j : means/rata-rata nilai dari variabel dimensi ke- j

N_i : jumlah data di dalam cluster ke- i

N : jumlah data keseluruhan

d : jumlah dimensi dari data

k : jumlah cluster

Elbow criterion adalah suatu *modelling criterion* yang bisa digunakan untuk menentukan jumlah *cluster* dengan melihat perubahan perbandingan antara nilai $RMSSTD$ dan RS . Hal ini dilihat dengan membandingkan persentase tingkat perubahan kedua nilai ($RMSSTD$ dan RS). Apabila muncul suatu keadaan yang berbanding terbalik dengan keadaan sebelumnya, maka titik sebelum terjadinya perubahan tersebut dianggap sebagai jumlah cluster yang paling tepat.

3. Mengukur kemiripan atau ketakmiripan antar obyek

Sesuai dengan tujuan analisis *cluster* yaitu untuk mengelompokkan obyek yang mirip dalam *cluster* yang sama, maka beberapa ukuran diperlukan untuk

mengetahui beberapa mirip atau berbeda obyek-obyek tersebut. Terdapat tiga metode yang dapat diterapkan dalam mengukur kesamaan antar obyek yaitu ukuran asosiasi, ukuran korelasi, ukuran jarak.

a. Ukuran asosiasi

Ukuran asosiasi dipakai untuk mengukur data berskala non-metrik (nominal atau ordinal) dengan cara mengambil bentuk-bentuk dari koefisien korelasi pada tiap obyeknya, dengan memutlakkan korelasi-korelasi yang bernilai negatif.

b. Ukuran korelasi

Ukuran korelasi dapat diterapkan pada data dengan skala metrik, namun ukuran korelasi jarang digunakan karena titik beratnya pada nilai suatu pola tertentu, padahal titik berat analisis *cluster* adalah besarnya obyek. Kesamaan antar variabel dapat dilihat dari koefisien korelasi antar pasangan obyek yang diukur dengan beberapa variabel.

c. Ukuran kedekatan

Metode ukuran jarak diterapkan pada data berskala metrik. Ukuran ini sebenarnya ukuran ketidakmiripan, dimana jarak yang besar menunjukkan sedikit kesamaan sebaliknya jarak yang kecil/pendek menunjukkan bahwa suatu obyek semakin mirip dengan obyek lainnya. Perbedaan dengan ukuran korelasi adalah bahwa ukuran korelasi bisa saja tidak memiliki kesamaan nilai tetapi memiliki kesamaan pola, sedangkan ukuran jarak lebih memiliki kesamaan nilai meskipun memiliki pola yang berbeda. Jarak *Eulidean* merupakan besarnya jarak suatu garis lurus yang menghubungkan antar obyek yang diteliti. Jarak *Eulidean* biasanya digunakan pada data mentah dan bukan data yang telah dilakukan standarisasi. Misalkan terdapat dua obyek yaitu A dengan koordinat (x_1, y_1) dan B dengan koordinat (x_2, y_2) maka jarak antara kedua obyek tersebut dapat diukur dengan rumus

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3.4)$$

Ukuran jarak antara obyek ke-i dengan obyek ke-j disimbolkan dengan d_{ij} dan variabel ke-k dengan $k=1, \dots, p$. Menurut Simamora (2005), nilai d_{ij}

diperoleh melalui perhitungan jarak kuadrat *Euclidean* yang dirumuskan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3.5)$$

dengan: d_{ij} : jarak kuadrat *Euclidean* antara obyek ke-i dengan obyek ke-j

p : jarak variabel *cluster*

x_{ik} : nilai atau data dari obyek ke-i pada variabel ke-k

x_{jk} : nilai atau data dari obyek ke-j pada variabel ke-k

Terdapat jarak yang merupakan variansi dari jarak *Euclidean*, yaitu jarak *Squared Euclidean*. Menurut Bilson (2005), jarak ini merupakan variansi dari jarak *Euclidean*. Hal yang membedakan pada jarak ini akhirnya dihilangkan, seperti pada rumus berikut:

$$d_{ij} = (v_{ik} - v_{jk}) \quad (3.6)$$

Untuk data yang harus dilakukan standarisasi, maka perhitungan dilakukan dengan menggunakan beberapa tipe ukuran jarak *Euclidean* (Greenacre dan Primicerio, 2013) berikut:

1. Jarak *Standardize Euclidean*

Jarak *Standardize Euclidean* digunakan ketika variabel memiliki skala yang berbeda.

2. Jarak *Weighted Euclidean*

Standarisasi antara dua dimensi vektor J dapat ditulis sebagai berikut:

$$d_{x,y} = \sqrt{\sum_{j=1}^J \left(\frac{x_j}{s_j} - \frac{y_j}{s_j} \right)^2} \quad (3.7)$$

Dengan s_j merupakan standar deviasi dari variabel ke-j. Pada *Weighted Euclidean* tidak perlu dilakukan pengurangan rata-rata dari x_j dan y_j .

$$\begin{aligned} d_{x,y} &= \sqrt{\sum_{j=1}^J \frac{1}{s_j^2} (x_j - y_j)^2} \\ &= \sqrt{\sum_{j=1}^J w_j (x_j - y_j)^2} \end{aligned} \quad (3.8)$$

Dengan $w_j = \frac{1}{s_j^2}$ dan w_j sebagai *weight* (bobot) untuk variabel ke-j.

Jarak pada data yang telah dilakukan standarisasi dianggap sebagai bobot dari variabel. Perhitungan ini disebut jarak *Weighted Euclidean*.

3. Jarak *Chi-Square*

Jarak ini digunakan pada tipe *count* data. Jarak *Chi-Square* terbentuk dari rumus (2.6) pada *Weighted Euclidean*. Perhitungan pada jarak *Chi-Square* dilakukan pada data yang saling berhubungan dan bukan pada data mentah/asli. Standarisasi pada jarak *chi-square* dihitung berdasarkan rata-rata dan bukan berdasarkan simpangan baku (Greenacre dan Primicerio, 2013).

Dinotasikan c_j merupakan elemen ke- j dari rata-rata setiap variabel. Kemudian jarak *chi-square* dinotasikan dengan χ^2 , jika terdapat dua kelompok dengan variabel $x=[x_1, x_2, \dots, x_j]$ dan $y=[y_1, y_2, \dots, y_j]$, maka didefinisikan sebagai berikut:

$$\chi_{x,y} = \sqrt{\sum_{j=1}^J \frac{1}{c_j} (x_j - y_j)^2} \quad (3.9)$$

Ketiga tipe jarak *Euclidean* diatas menggunakan proses standarisasi, sedangkan pada penelitian ini data yang digunakan tidak memiliki perbedaan skala, sehingga tidak perlu dilakukan standarisasi. Perhitungan kesamaan obyek dilakukan dengan menggunakan rumus (2.3). Hasil perhitungan menggunakan jarak *Euclidean distance* tersebut dituangkan dalam *proximity*. *Proximity* menampilkan jarak antara variabel satu dengan variabel lain dalam bentuk matrik $N \times N$, dan biasa disebut dengan *proximity matrix*. Pada *proximity matrix*, semakin kecil jarak *Euclidean*, maka semakin mirip kedua variabel.

3. Memilih Suatu Prosedur Analisis Cluster

Prosedur *cluster* atau pengelompokan data dapat dilakukan dengan dua metode yaitu metode hierarki dan metode non hierarki.

a. Metode Hierarki

Tipe dasar dalam metode hierarki bisa aglomeratif atau devisif. Pada pengclusteran aglomeratif, dimulai dengan menempatkan obyek dalam *cluster-cluster* yang berbeda kemudian mengelompokan obyek secara bertahap ke dalam *cluster-cluster* yang lebih besar, sedangkan pada

pengclusteran devisif dimulai dengan menempatkan semua obyek sebagai satu *cluster*. Kemudian secara bertahap obyek-obyek dipisahkan ke dalam *cluster-cluster* yang berbeda, dua *cluster*, tiga *cluster*, dan seterusnya (Simamora: 2005).

Ada lima metode hierarki aglomeratif dalam pembentukan *cluster* yaitu:

- i. Pautan Tunggal (*Single Linkage*)
- ii. Pautan Lengkap (*Complete Linkage*)
- iii. Pautan Rata-rata (*Average Linkage*)
- iv. Metode Ward (*Ward's Method*)
- v. Metode Centroit (*Pusat*)

Secara umum langkah-langkah dalam metode *cluster* hierarki aglomeratif untuk membentuk kelompok dari N obyek sebagai berikut:

- a) Dimulai dengan N *cluster*, dimana masing-masing memuat satu kesatuan. Jika terdapat matriks N x N dengan jarak $D = \{d_{ik}\}$
- b) Mencari matriks jarak untuk pasangan *cluster* terdekat. Misalkan pasangan *cluster* paling mirip obyek U dan V maka $D = \{d_{uv}\}$, sehingga U dan V dipilih.
- c) Menggabungkan *cluster* U dan V menjadi *cluster* baru (UV). Memperbaharui masukan dalam matriks jarak dengan cara
 - 1) Menghapus baris dan kolom sesuai dengan *cluster* U dan V
 - 2) Menambahkan baris dan kolom dengan menambah nilai jarak antara *cluster* baru (UV) dan semua sisa *cluster*.
- d) Mengulangi langkah (b) dan (c) sebanyak (n-1) kali. (Semua obyek akan berada dalam *cluster* tunggal pada berakhirnya algoritma). Mencatat identitas dari *cluster* yang digabungkan dan tingkat (jarak atau similaritas) dimana penggabungan terjadi.

Metode hierarki yang digunakan dalam penelitian ini adalah metode *average linkage* (pautan rata-rata).

- 1) Pautan Lengkap (*Complete Linkage*)

Metode pautan lengkap (*complete linkage*) didasarkan pada jarak maksimum. Menurut Simamora (2005), jarak antara satu *cluster* dan

cluster lain diukur berdasarkan obyek yang mempunyai jarak terjauh. Pada awal perhitungan, terlebih dahulu mencari nilai minimum dalam $D=\{d_{ij}\}$ dan menggabungkan obyek-obyek yang bersesuaian, misalnya U dan V, untuk mendapatkan *cluster* (UV). Pada langkah (c) dari algoritma yang dijelaskan sebelumnya, jarak antara (UV) dan *cluster* lain W, dihitung dengan cara:

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (3.10)$$

Disini d_{UW} dan d_{VW} merupakan jarak paling jauh antara anggota *cluster-cluster* U dan W, dan juga *cluster-cluster* V dan W (Jonshon dan Wichern, 1996).

2) Pautan Rata-rata (*Average Linkage*)

Average Linkage menghitung jarak antara dua *cluster* yang disebut sebagai jarak rata-rata dimana jarak tersebut dihitung pada masing-masing *cluster*.

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)}N_W} \quad (3.11)$$

Dengan d_{ik} (Johnson dan Wichern, 1996) merupakan jaak antar obyek i dalam *cluster* (UV) dan obyek k dalam *cluster* W. Sedangkan $N_{(UV)}$ dan N_w berturut-turut meupakan jumlah obyek dalam *cluster* (UV) dan (W).

b. Metode Non-Hierarki

Pada metode non-hierarki, banyaknya *cluster* yang ingin di bentuk harus di tentukan terlebih dahulu. Metode non-hierarki sering disebut *K-Means clustering* . Pusat *cluster* yang dipilih pada metode ini merupakan pusat sementara dengan terus memperbaharui pusat *cluster* sampai kriteria pemberhentian tercapai.

Metode *K-means* merupakan metode non-hierarki yang bersifat tanpa arahan, hal ini dikarenakan data yang dianalisis tidak mempunyai label kelas, yang berarti dalam proses pengelompokannya tidak mempunyai anggota *cluster* yang pasti. Obyek yang sudah masuk dalam *cluster* tertentu masih bisa pindah ke *cluster* lainnya. MacQueen berpendapat (Johnson dan Wichern, 1996) bahwa istilah *K-means* untuk mendeskripsikan bahwa algoritma *K-*

means menandai setiap obyek masuk kedalam *cluster* yang mempunyai pusat *cluster* (rata-rata) terdekat.

Untuk membantu dalam menganalisis data, penulis menggunakan *SPSS 20 for windows* dan *Ms. Excel*.

4. Interpretasi Cluster

Pada tahap interpretasi meliputi pengujian pada masing-masing cluster yang terbentuk untuk memberikan nama atau keterangan secara tepat sebagai gambaran sifat dari cluster tersebut. Ketika memulai proses interpretasi digunakan rata-rata (centroid) setiap cluster pada setiap variabel.

3.15. Pemilihan Metode Terbaik dengan Simpangan Baku

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

$$s_w = \frac{1}{K} \sum_{k=1}^K s_k \quad (3.12)$$

dengan: K : Banyaknya kelompok yang terbentuk
 s_k : Simpangan baku kelompok ke-k

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

$$s_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_k)^2} \quad (3.13)$$

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

$$s_b = \left[\frac{1}{(K-1)} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2 \right]^{1/2} \quad (3.14)$$

dengan: K : banyaknya kelompok yang terbentuk
 \bar{X}_k : rata-rata kelompok ke-k

\bar{X} : rata-rata keseluruhan kelompok

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

$$s = \frac{s_w}{s_b} \times 100\% \quad (3.15)$$

