PENGELOMPOKAN BERBASIS HIRARKI PADA MANGGIS

Retno Nugroho Whidhiasih[1], Malikus Sumadyo[2]
Email : retno.nw@gmail.com [1], malikussumadyo@gmail [2]
[1,2]Instansi (Teknik Komputer, Universitas Islam 45)

Abstract

Jumlah variable yang besar sering menjadi kendala dalam proses pengelompokan data, sehingga perlu dilakukan reduksi jumlah variable. Analisis komponen utama digunakan dapat digunakan untuk mereduksi sejumlah variable. Metode tersebut akan merepresentasikan informasi dalam bentuk variabel variabel baru yang merupakan kombinasi linier dari variabel yang lama. Clustering merupakan metode pengelompokan objek berdasarkan ukuran kemiripan. Hierarchical clustering digunakan untuk pengelompokan manggis berbasis hirarki dengan metode perhitungan jarak euclidean. Dari hasil analisis didapat hasil eigenvalue yang mampu memberikan informasi sebesar 81.4%. selanjutnya dilakukan clustering menggunakan metode berhirarki.
Keywords : Analisis Komponen Utama, Cluster berhierarki, Jarak Euclidean

A. Pendahuluan

Jumlah variable yang cukup besar sering menjadi kendala dalam proses pengelompokan suatu data, sehingga reduksi jumlah variable perlu untuk dilakukan. Reduksi jumlah variabel dilakukan untuk mendapatkan variabel baru dalam jumlah yang lebih kecil. Pengelompokan suatu data atau clustering ke dalam beberapa kelompok yang jumlahnya lebih sedikit. Clustering merupakan teknik mengelompokkan data (objek) ke dalam beberapa cluster (kelompok) yang belum diketahui jumlah kelas baru pengelompokan tersebut.

Pada kegiatan pasca panen untuk mengelompokkan buah dapat juga dilakukan menggunakan clustering. Hal ini dapat dilakukan untuk membantu pengelompokan dengan meminimalisasi kesalahan-kesalahan secara manual jika dilakukan oleh manusia yang akan berakibat terjadinya kerugian baik kerugian di pihak petani maupun pihak konsumen.

Analisis komponen utama digunakan di perusahaan kulit kikil sapi yang menggunakan 15 variabel untuk menentukan kualitas, sehingga digunakan analisis komponen utama dan mendapatkan 2 komponen utama yang sangat berpengaruh terhadap kualitas (Saepurohman dan Putro, 2019). Analisis komponen utama juga lebih akurat dalam pengenalan wajah dibandingkan dengan hidden markov model dengan akurasi 86,6% (Syakala *et. Al*, 2015). Selain itu analisis komponen utama juga salah satu alternative untuk tingkat kecepatan dan keakuratan dari pelatihan jaringan syaraf tiruan (Puspiraningrum *et. al*., 2014).

Aglomeratif hierarchical clustering dapat digunakan untuk pengelompokan pelanggan dan dikaitkan dengan strategi pemasaran (Widyawati *et al*., 2020). Pada perbandingan nilai rasio simpangan baku metode complete linkage lebih baik dibandingkan dengan single linkage dan average linkage pada pengelompokan kecamatan berdasar variable ternak di

Kabupaten Sidoarjo (Mu'afa dan Ulinnuha, 2019).

Dalam penelitian ini dilakukan clustering terhadap 103 manggis yang belum diketahui kelompoknya berdasarkan 15 variabel penduga. Variabel penduga tersebut adalah r, g, b, H, S, V, L, a*, b*, u*, v*, entropi, kontras, energi dan homogenitas. Dengan jumlah variabel yang cukup banyak akan dilakukan reduksi dimensi menggunakan analisis komponen utama. Pengelompokan dilakukan menggunakan algorithma agglomerative hierarchical, yaitu metode single linkage, average linkage dan complete linkage. Penelitian ini diharapkan dapat mengetahui perbedaan pengelompokan dengan menggunakan metode yang sama dan cluster berbeda.

B. Landasan Teori

Analisis Komponen Utama

Analisis Komponen Utama adalah sebuah teknik untuk membangun variabel baru yang merupakan kombinasi linier dari variabel – variabel asli. Jumlah maksimum variabel baru tersebut akan sama dengan jumlah variabel lama. Variabel baru tersebut akan tidak saling berkorelasi satu sama lain. Karena itulah AKU merupakan salah satu metode multivariate untuk mengatasi multikolinearitas antar variable.

Salah satu tujuan dari analisis komponen utama adalah mereduksi dimensi data asal yang semula terdapat p variabel bebas menjadi k komponen utama (dimana $k < p$). Langkah awal yaitu menghitung skor masing-masing komponen utama. Lalu dipilih k komponen ($k < m$) untuk digunakan sebagai peubah bebas dalam MKT. Secara umum kriteria pemilihan k komponen utama yaitu :

1. Dalam pemilihan jumlah komponen teresebut belum ada aturan tertentu yang disepakati oleh semua ahli statistika. Sebagian ahli statistika ada yang mengambil akar ciri yang lebih besar dari 1 atau mengambil komponen utama tertentu, dimana proporsi keragaman y yang dapat diterangkan oleh komponen tersebut dianggap cukup berarti.
2. Proporsi kumulatif keragaman data asal yang dijelaskan oleh k komponen utama minimal 80%, dan proporsi total variansi populasi bernilai cukup besar.
3. Dengan menggunakan scree plot yaitu plot antara i dengan l_i, pemilihan nilai k berdasarkan scree plot ditentukan dengan melihat letak terjadinya belokan dengan menghapus komponen utama yang menghasilkan beberapa nilai eigen kecil membentuk pola garis lurus.

Analisis Clustering

Analisis clustering bertujuan untuk menggerombolkan unit-unit pengamatan ke dalam beberapa cluster dimana setiap unit pengamatan dalam satu gerombol akan mempunyai ciri yang relatif sama sedangkan antar cluster unit pengamatan memiliki sifat yang berbeda. Hal-hal yang penting dalam analisis cluster adalah Ukuran kesamaan atau kemiripan untuk semua pasangan unit, Kriteria dan algoritma clustering, Penafsiran hasil penggerombolan.

Sebelum dilakukan clustering terlebih dulu dilakukan penentuan jarak kedekatan (similarity) antar objek dengan menggunakan jarak euclidean. Jarak

euclidean cukup fleksibel digunakan untuk mengatasi data dengan skala pengukuran yang berbeda dengan menggunakan transformasi baku (Z). Jarak euclidean dapat diukur menggunakan persamaan 1. Dalam memperbaiki matrik jarak menggunakan metode single linkage dihitung dengan persamaan 2, menggunakan average linkage dihitung dengan persamaan 3 dan menggunakan complete linkage dihitung dengan persamaan 4.

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

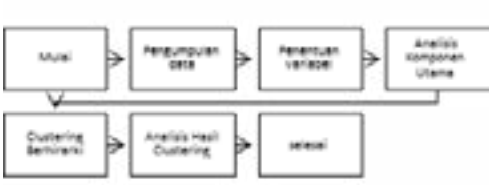
Analisis cluster ini dibagi menjadi dua bagian utama, yaitu metode berhirarki (Hierarchical Clustering Method) dan metode tidak berhirarki (Non Hierarchical Clustering Method). Metode berhirarki sering digunakan apabila jumlah kelompok yang dibentuk belum diketahui, sedang metode tak berhirarki dipakai bila banyaknya kelompok yang akan dibentuk telah ditentukan.

Pada metode analisis cluster berhirarki terdapat beberapa metode untuk memperbaharui matrik jarak antara lain metode pautan lengkap (complete linkage), metode pautan rataan (average linkage) dan metode pautan tunggal (single linkage).

C. Metode Penelitian

Penelitian ini terbagi menjadi lima tahapan, yaitu pengumpulan data, penentuan variable, analisis komponen utama utama, clustering berhirarki dan analisis hasil clustering, disajikan pada Gambar 1.

Gambar 1. Tahapan Penelitian



Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa 103 buah citra buah manggis Padang yang berukuran 640 x 480 piksel. Data tersebut didapatkan dari laboratorium sistem dan manajeman keteknikan pertanian Universitas Padjajaran Bandung. Citra tersebut merupakan hasil dari pengambilan buah manggis didalam kotak instrument tertutup yang diberi pelapis kain warna hitam, menggunakan kamera Change Couple Device (CCD) Telview tipe ST205 color, dua buah lampu PL Philips warna putih 11 watt dan bidang dasar pemotretan berwarna putih, dengan jarak rekam 30 cm dan posisi sudut pencahayaan 45.

Selanjutnya dilakukan konversi citra menggunakan software matlab R2008a ke bentuk vektor berukuran 2400 x 1. Citra 103 buah manggis yang telah berubah menjadi vektor tersebut kemudian digabungkan menjadi sebuah matrik berukuran 2400 x 103.

Penentuan Variabel

Penentuan variable dilakukan untuk menentukan variable penduga yang akan digunakan untuk keperluan clustering. Diambil ciri-ciri khusus yang dapat membedakan objek satu dengan lainnya yang jumlahnya akan dibatasi sesuai

dengan tingkat kepentingannya, tanpa mengesampingkan tujuan utama yaitu mendapatkan hasil yang optimal. Variabel yang digunakan sejumlah 15 variabel, yaitu r, g, b, H, S, V, L, a*, b*, u*, v*, entropi, kontras, energi dan homogenitas.

Analisis Komponen Utama

Analisis komponen utama digunakan untuk mereduksi dimensi dari variable-variabel penduga yang berjumlah besar tanpa menghilangkan sifat aslinya. Analisis komponen utama dilakukan dalam beberapa jumlah komponen utama kemudian dilakukan penentuan komponen utama berdasarkan hasil yang muncul. Analisis komponen utama yang menggunakan beberapa komponen utama diamati hasilnya sehingga dapat ditentukan jumlah komponen utama yang digunakan.

Terdapat empat langkah dalam analisis komponen utama. Langkah tersebut adalah (1) menginput data (mxn), dengan m adalah jumlah observasi sedangkan n adalah jumlah sampel. (2) Preprocessing (pre-PCA) menggunakan standarisasi data dan kovarian atau corelastion. (3) Proses PCA menggunakan eigen value decomposition (EVD) atau singular value decomposition (SVD). (4) Output berupa data hasil transformasi (mxk), m adalah jumlah observasi sedangkan k adalah jumlah principal component.

Clustering Berhirarki

Clustering berhirarki digunakan dikarenakan jumlah kelompok yang akan dibentuk belum diketahui. Dalam clustering berhirarki dilakukan clustering berdasarkan jarak kemiripannya untuk memperbaharui matrik jarak, yaitu menggunakan metode pautan lengkap (*complete linkage*), metode pautan rataan (*average linkage*) dan metode pautan tunggal (*single linkage*).

Sebelum dilakukan clustering terlebih dulu dilakukan penentuan jarak kedekatan (similarity) antar objek dengan menggunakan jarak euclidean. Jarak euclidean cukup fleksibel digunakan untuk mengatasi data dengan skala pengukuran yang berbeda dengan menggunakan transformasi baku (Z). Jarak euclidean dapat diukur menggunakan persamaan 1. Dalam memperbaiki matrik jarak menggunakan metode single linkage dihitung dengan persamaan 2, menggunakan average linkage dihitung dengan persamaan 3 dan menggunakan complete linkage dihitung dengan persamaan 4.

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

d_{ij} : jarak antara objek i dan j

x_{ij} : nilai objek I pada variabel ke j

x_{jk} : nilai objek j pada variabel ke k

p : banyaknya variabel yang diamati

$$d_{(ab)c} = \min\{d_{a,c}; d_{b,c}\} \quad \ldots\ldots\ldots\ldots\ldots(2)$$

$$d_{(ab)c} = \text{average}\{d_{a,c}; d_{b,c}\} \quad \ldots\ldots\ldots\ldots\ldots(3)$$

$$d_{(ab)c} = \max\{d_{a,c}; d_{b,c}\} \quad \ldots\ldots\ldots\ldots\ldots(4)$$

Analisis Clustering

Hasil clustering berhirarki dengan menggunakan single linkage, average linkage dan complete linkage dianalisis perbedaan hasil clusteringnya.

D. Hasil dan Pembahasan

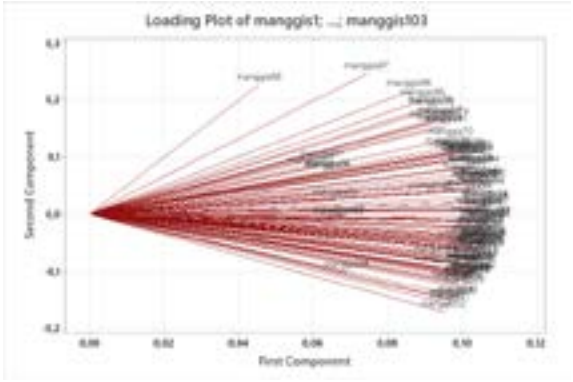
Data yang digunakan adalah citra buah manggis Padang yang merupakan data sekunder dan terdiri dari 103 data dari. Data tersebut didapatkan dari laboratorium sistem dan manajeman keteknikan pertanian Universitas Padjajaran Bandung.

Analisis Principal Component

Data yang dianalisis adalah data dari citra buah manggis yang terdiri dari 103 data. Data telah dikonversi dan tersusun dalam matrik berukuran 2400x103, yang terdiri dari 15 variabel yaitu r, g, b, H, S, V, L*, a*, b*, u*, v*, entropi, kontras, energi dan homogenitas. Berdasarkan pengelolaan data, reduksi dimensi variabel tahap kematangan buah manggis menggunakan analisis principal component.

Nilai eigenvalue (λ) untuk principal component pertama (PC1) adalah 86,596, nilai eigenvalue untuk principal component kedua (PC2) adalah 6,860. Eigenvalue kedua komponen mewakili 84,1% dan 6,7% dari seluruh variabilitas. Untuk mendapatkan informasi dari variabel baru yang dapat mewakili keseluruhan sekitar 80% sampai 90% dapat digunakan PC1. Hal ini berarti jika 15 variabel yang terdiri dari r, g, b, H, S, V, L*, a*, b*, u*, v*, entropi, kontras, energi dan homogenitas direduksi menjadi 1 variabel, maka 1 variabel baru tersebut dapat menjelaskan atau mewakili 84,1% dari total variabilitas dari 15 variabel.

Scree plot untuk menunjukkan nilai eigenvalue dari nilai terkecil ke nilai terbesar dan jumlah komponen tersaji pada Gambar 4.2. Pengambilan nilai eigenvalue didasarkan pada belokan tajam yang menunjukkan perubahan yang signifikan dibanding titik-titik yang relative stabil. Dari Scree Plot diatas Nilai Eigenvalue ke-2 dapat disimpulkan sebagai titik belok yang tajam. Sedangkan nilai eigenvalue ke-1 diambil berdasarkan tingkat kepercayaan mendekati 84,1 %.



Gambar 4.3 Posisi jarak antar buah manggis

Posisi jarak antar buah manggis disajikan pada Gambar 4.3. Terlihat himpunan manggis1-manggis28 mempunyai posisi yang berdekatan, manggis88, manggis87 dan manggis96 mempunyai posisi yang jauh dengan manggis-manggis lainnya, dan dapat dilihat posisi manggis-manggis lainnya yang saling berdekatan. Dapat dilihat dengan perbedaan nilai-nilai variabel yang kecil memberikan perbedaan posisi yang berdekatan.

Koefisien masing-masing variabel terhadap variabel baru (PC1) terlihat jelas pada hasil kombinasi linier yang dihasilkan oleh masing-masing PC.

PC1 = 0,097 manggis1 + 0,102 manggis2 + 0,099 manggis3 + 0,099 manggis4 + 0,102 manggis5 + 0,103 manggis6 + 0,101 manggis7 + 1,104 manggis8 + 0,101 manggis9 + 0,098 manggis10 + 0,102 manggis11 + 1,104 manggis12 + 0,105 manggis13 + 0,102 manggis14 + 0,102 manggis15 + 0,104 manggis16 + 0,097 manggis17 + 0,104 manggis18 + 0,102 manggis19 + 0,104 manggis20 + 0,101 manggis21 + 0,103 manggis22 + 0,093 manggis23 + 0,099 manggis24 + 0,102 manggis25 + 0,105 manggis26 + 0,106 manggis27 + 0,105 manggis28 + 0,069 manggis29 + 0,105 manggis30 + 0,106 manggis31 + 0,095 manggis32 + 0,101 manggis33 + 0,102 manggis34 + 0,105 manggis35 + 0,106 manggis36 + 0,105 manggis37 + 0,101 manggis38 + 0,105 manggis39 + 0,102 manggis40 + 0,106 manggis41 + 0,105 manggis42 + 0,105 manggis43 + 0,101 manggis44 + 0,093 manggis45 + 0,106 manggis46 + 0,106 manggis47 + 0,102 manggis48 + 0,097 manggis49 + 0,105 manggis50 + 0,106 manggis51 + 0,107 manggis52 + 0,102 manggis53 + 0,094 manggis54 + 0,105 manggis55 + 0,102 manggis56 + 0,107 manggis57 + 0,106 manggis58 + 0,103 manggis59 + 0,106 manggis60 + 0,104 manggis61 + 0,068 manggis62 + 0,098 manggis63 + 0,105 manggis64 + 0,104 manggis65 + 0,092 manggis66 + 0,103 manggis67 + 0,064 manggis68 + 0,102 manggis69 + 0,097 manggis70 + 0,106 manggis71 + 0,101 manggis72 + 0,106 manggis73 + 0,096 manggis74 + 0,091 manggis75 + 0,064 manggis76 + 0,100 manggis77 + 0,104 manggis78 + 0,103 manggis79 + 0,107 manggis80 + 0,101 manggis81 + 0,100 manggis82 + 0,066 manggis83 + 0,103 manggis84 + 0,092 manggis85 + 0,100 manggis86 + 0,074 manggis87 + 0,045 manggis88 + 0,094 manggis89 + 0,096 manggis90 + 0,096 manggis91 + 0,106 manggis92 + 0,059 manggis93 + 0,104 manggis94 + 0,090 manggis95 + 0,086 manggis96 + 0,062 manggis97 + 0,092 manggis98 + 0,105 manggis99 + 0,101 manggis100 + 0,100 manggis101 + 0,102 manggis102 + 0,067 manggis103

Clustering

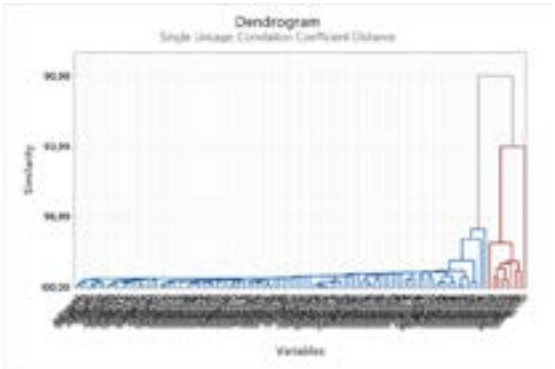
Dalam analisis cluster ini dilakukan analisis cluster berhirarki dengan menggunakan cluster sebanyak 2, 3 dan 4 dengan metode hirarki single linkage, average linkage dan complete linkage. Himpunan objek-objek dari masing-masing cluster pada tiap metode yang digunakan dapat diamati pada Tabel 4.1. Pada tabel tersebut terlihat sedikit perbedaan hasil cluster untuk setiap clustering dengan metode yang sama.

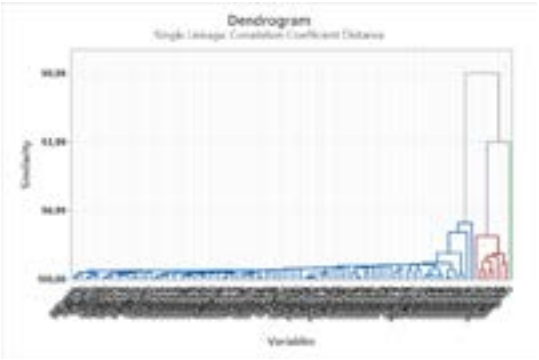
Untuk dapat melihat penggambaran langkah-langkah penggabungan masing-

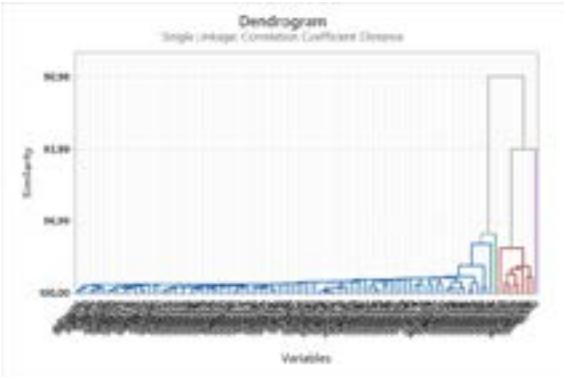
masing variabel hingga menjadi satu bagian dapat menggunakan dendogram. Dalam dendogram berikut dapat diamati perbedaan langkah-langkah penggabungan masing-masing variabel hingga menjadi satu bagian dalam cluster yang sama dengan metode yang berbeda.

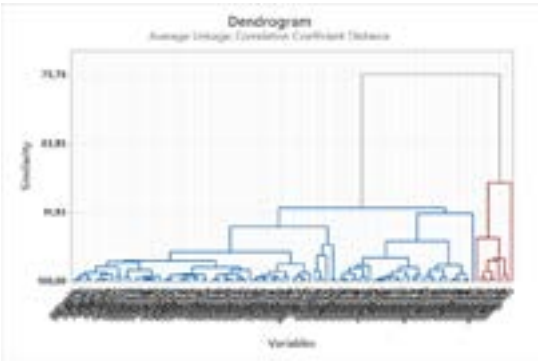
Langkah penggabungan menggunakan metode single linkage dapat diamati pada Gambar 4.4, langkah penggabungan menggunakan metode average linkage dapat diamati pada Gambar 4.5 dan langkah penggabungan menggunakan metode complete linkage dapat diamati pada Gambara 4.6. Pada dendogram tersebut terlihat hampir tidak ada perbedaan hasil clustering untuk setiap clustering menggunakan metode yang sama.

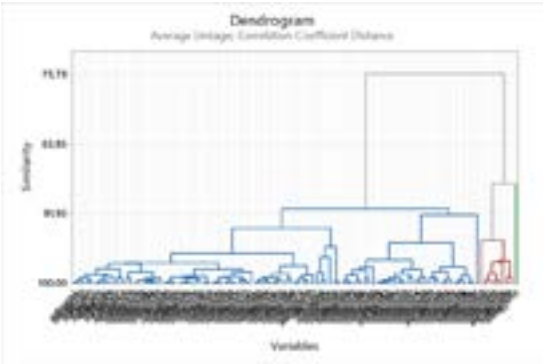
Output pada window session menunjukkan 102 cluster. Pada tahap pertama manggis34 digabungkan dengan manggis40. Tahap selanjutnya manggis33 digabungkan dengan manggis38 dan seterusnya. Penggabungan objek ditunjukkan dengan jelas pada dendogram. Penggabungan tersebut memperlihatkan objek-objek yang mempunyai hubungan erat.

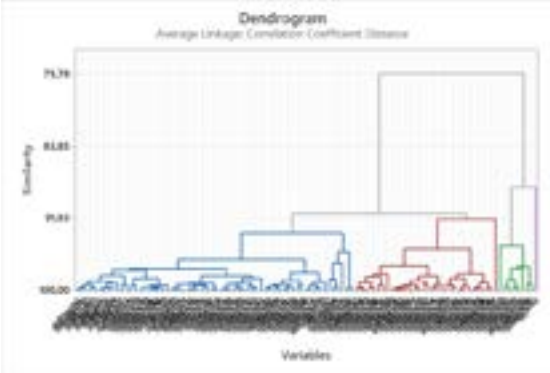


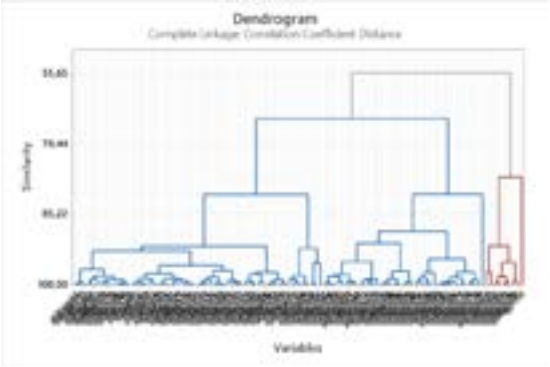


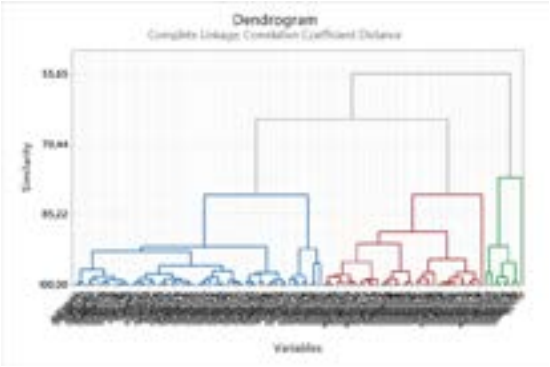


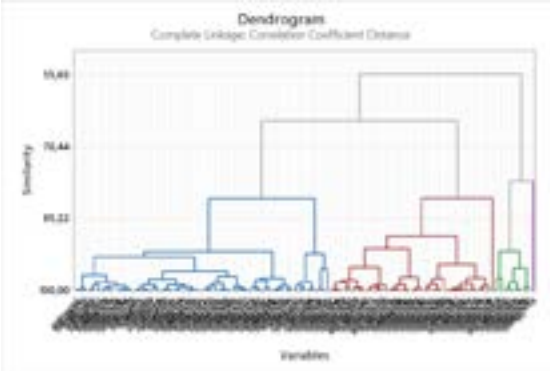












E. Penutup

Kesimpulan dari penelitian ini (1) Hampir tidak ada perbedaan hasil dari clustering manggis untuk clustering menggunakan metode berhirarki yang sama, (2) Clustering manggis dengan metode berhirarki sama dengan jumlah cluster 2, 3 dan 4 memberikan hasil yang sama, (3) Tidak ada ketetapan untuk memindahkan objek yang mungkin tidak tepat pengelompokannya pada tiap tahap pada analisis cluster metode berhirarki. Perlu dicoba menggunakan jumlah objek yang lebih kecil sehingga bisa dicermati lebih detil kemungkinan satu dan lainnya

DAFTAR PUSTAKA

Mu'afa Sulthan Fikri, Ulinnuha Nurissaidah. 2019. Perbandingan Metode Single Linkage, complete Linkage Dan Average Linkage dalam Pengelompokan Kecamatan Berdasarkan Variabel Jenis Ternak Kabupaten Sidoarjo. *Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi* (4):2.

Richard A. Johnson & Dean W. Wicherin, *Applied Multivariate Statistical Analysis*, Fifth Edition, Prentice Hall, New Jersey.

Saepurohman Tatan dan Putro Bramantyo Eko. 2019. Analisis Principal Component Analysis (PCA) Untuk Mereduksi Faktor-Faktor yang Mempengaruhi Kualitas Kulit Kikil Sapi. *Seminar dan Konferensi Nasional IDEC.*

Syakala, Abdu Rakhman, Puspitaningrum Dyah, Purwandari Erdina Putri. 2015. Perbandingan Metode Principal Component Analisys Dengan Hidden

Markov Model Dalam Pengenalan Identitas Seseorang Melalui Wajah. *Jurnal Rekursif* (3) : 2.

Puspitaningrum Dyah, Sari Dyan Kemala, Susilo Boko. 2014. Dampak Reduksi Sampel Menggunakan Principal Component Analisys Pada Pelatihan Jaringan Syaraf Tiruan Terawasi. *Jurnal Pseudocode* (2):1.

Widyawati, Saptomo wawan Laksito yuli, Utami Yustina Retno Wahyu. 2020. Penerapan Aglomeratif Hierarchical Clustering Untuk Segmentasi Pelanggan. *JUrnal Ilmiah Sinus (JIS*) (18): 1.