Pengelompokan Aksesi Jeruk Persilangan Berdasarkan Karakter Kuantitatif dan Kualitatif Menggunakan *Fuzzy C-Means* dan *K-Modes*

[(1)]Candra Widhi Saputra [(2)]Sutikno [(3)]Chaireni Martasari
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: [(2)] sutikno@statistika.its.ac.id, [(3)] c_martasari03@yahoo.com,
[(1)] candrawsaputra@mhs.statistika.its.ac.id

Abstrak — ***Balitjestro telah memulai program pemuliaan jeruk sejak tahun 2006 dengan cara persilangan antar 2 jenis jeruk. Satu proses persilangan dapat menghasilkan sekitar 150 varietas tanaman baru. Banyaknya hasil varietas baru dari hasil persilangan di dapat dari biji buah hasil persilangan antara jeruk jenis Siam Pontianak dan jeruk jenis Soe. Proses seleksi sangat penting untuk memilah antara varietas biasa dan varietas unggul. Seleksi salah satunya dapat menggunakan karakter tanaman jeruk. Penelitian ini dilakukan untuk mengetahui pengelompokan aksesi jeruk persilangan dengan menggunakan metode*** **Fuzzy C-Means, K-Modes,** ***dan*** **Ensemble Cluster.** ***Hasil penelitian menunjukkan bahwa Fuzzy C-Means menghasilkan 3 kelompok optimum, K-Modes menghasilkan 4 kelompok dengan akurasi 100%, dan ensemble cluster akan dibuat 4 kelompok dengan akurasi 97%. Metode fuzzy c-means cluster yang digunakan pada karakter kuantitatif cukup untuk mengelompokkan kedua tipe data karena memiliki nilai icdrate 0,27 dan akurasi 97%***

Kata kunci ***: Persilangan. Fuzzy C-Means, K-Modes, Ensemble Cluster, Balitjestro***

I. PENDAHULUAN

Sejak tahun 2006, Balai Penelitian Tanaman Jeruk dan Buah Subtropika (BALITJESTRO) telah memulai program pemuliaan jeruk dengan cara persilangan antar 2 jenis jeruk. Dalam 1 proses persilangan dapat menghasilkan kira-kira 150 varietas tanaman baru. Banyaknya hasil varietas baru dari hasil persilangan di dapat dari biji buah hasil persilangan antara jeruk jenis Siam Pontianak dan jeruk jenis Soe. Untuk dapat membedakan varietas baru tersebut, maka varietas yang baru muncul tersebut nantinya akan diberi nama. Pemberian nama tersebut dalam bidang pertanian disebut dengan aksesi. Persilangan tanaman dilakukan untuk menciptakan suatu varietas baru yang memiliki kualitas baik dari kedua induk yang disilangkan. Jeruk Siam telah menjadi target utama dari pihak Balitjestro untuk diperkaya varietasnya. Hal itu dilakukan karena jeruk jenis Siam memiliki cita rasa yang manis. Jeruk jenis Siam memiliki suatu kelemahan yaitu memiliki warna kulit yang kurang menarik, menarik tidaknya kulit tersebut diindikasikan dengan warna kuning atau orange. Tujuan utama dari Balitjestro menyilangkan jeruk jenis Siam dengan jeruk jenis yang lain adalah mendapatkan varietas jeruk yang memiliki cita rasa manis, warna kulit buah kuning atau orange, dan mudah dikupas. [1].

Setelah didapatkan suatu varietas-varietas baru dilakukan proses seleksi. Proses seleksi sangat penting untuk memilah antara varietas biasa dan varietas unggul. Seleksi salah satunya dapat menggunakan karakterisasi tanaman jeruk tersebut. Terdapat 2 jenis data yang akan dihasilkan dari karakterisasi tanaman jeruk ini, yaitu data kuantitatif dan data kualitatif.

Karakterisasi tanaman jeruk tersebut menghasilkan 2 jenis data, sehingga menghasilkan masalah yang ada di Balai Penelitian Jeruk dan Buah Subtropika (BALITJESTRO). Sehingga dalam penelitian ini diuji bagaimana metode *Fuzzy C-Means Cluster* dapat bekerja pada data kuantitatif, K-Modes pada data kualitatif, dan *Ensemble Cluster* untuk penggabungan 2 jenis data.

II. TINJAUAN PUSTAKA

A. Chernoff Face

Analisis ini pertama kali diperkenalkan oleh Herman Chernoff yaitu teknik visualisasi berupa metode grafik untuk merepresentasikan data dengan banyak variabel dalam bentuk wajah kartun (Chernoff faces) yang ditentukan lebih dari 20 parameter yaitu terdiri dari panjang hidung, kelengkungan mulut, panjang alis, besar sudut alis dan lain-lain.

Chernoff faces menjadi alat peraga yang sangat efektif karena menghubungkan data dengan raut wajah yang mana terkadang dapat menunjukkan keadaan seseorang atau kelompok bahkan suatu wilayah. Dimensi data yang berbeda dipetakan untuk raut wajah yang berbeda, sebagai contoh lebar muka, lebar telinga, tinggi telinga, lebar dari mulut, panjang hidung dan lain-lain.

B. Fuzzy C-Means Cluster

Fuzzy Cluster merupakan penerapan dari konsep fuzzy terhadap cluster. Konsep fuzzy diharapkan mampu untuk meminimalkan kejadian konvergen yang biasa dialami oleh metode cluster biasa. Metode FCM merupakan pengembangan dari metode tak berhirarki c-means cluster, karena pada awalnya menentukan jumlah kelompok yang akan dibentuk. Berikut ini adalah algoritma FCM [4]:

1. *Input* data yang akan di *cluster* X, berupa matriks berukuran $n \times m$ (n = banyaknya data, m = banyaknya variabel setiap data). X_{ij} = data sampel ke-*i* (*i*=1,2,...,n), variabel ke-*j* (*j*=1,2,...,m).
2. Menentukan jumlah *cluster* (*c*) , *weighting exponent* (*w*=2), maksimum iterasi, error terkecil ($\varepsilon = 10^{-6}$), fungsi objektif awal (P_0=0), dan iterasi awal (t=1).
3. Membangkitkan bilangan *random* u_{ik}, *i* =1,2,...,n; *k*= 1,2,...,c sebagai elemen matriks partisi awal partisi *U*

4. Menghitung *centriod* dari masing-masing kelompok sesuai persamaan berikut.

$$v_{kj} = \frac{\sum_{i=1}^{n} (u_{ik})^w x_{ij}}{\sum_{i=1}^{n} (u_{ik})^w} \quad (1)$$

dimana :

n : Banyaknya data

i : Indeks objek ke-i

k : Indeks *cluster* ke-k

u_{ik} : Keanggotaan data objek ke-i dan *cluster* ke-k

v_{kj} : *Centroid*/rata rata *cluster* ke-k untuk variabel ke-j

w : *Weighting exponent*

x_{ij} : Nilai objek ke-i yang ada didalam *cluster* tersebut untuk variabel ke-j

5. Menghitung derajat keanggotaan setiap data pada setiap *cluster*. Dimana untuk nilai derajat keanggotaan mempunyai jangkauan nilai $0 \le u_{ik} \le 1$

$$u_{ik} = \sum_{j=1}^{c} \left[\left(\frac{d_{ki}}{d_{ji}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (2)$$

untuk nilai d_{ki} menggunakan persamaan (2)

u_{ik}: Keanggotaan data objek ke-i dan *cluster* ke-k

d_{ki} : Jarak *Euclidean cluster* ke-k objek ke-i

d_{ji} : Jarak *Euclidean* variabel ke-j objek ke-i

m : *Weighting exponent*

c : Banyaknya *cluster*

6. Menentukan kriteria penghentian iterasi, yaitu perubahan matriks partisi pada iterasi sekarang dan iterasi sebelumnya. Apabila $|U^{l} - U^{(l-1)}| < \varepsilon$ maka proses berhenti.

$$U^{l} = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} \left(x_{ij} - v_{kj} \right)^2 \right] (u_{ik})^m \right) \quad (3)$$

dimana :

u_{ik}: Keanggotaan data objek ke-i dan *cluster* ke-k

v_{kj}: *Centroid* rata rata *cluster* ke-k untuk variabel ke-j

x_{ij} : Nilai objek ke-i yang ada didalam *cluster* tersebut untuk variabel ke-j

m : *Weighting exponent*

c : Banyaknya *cluster*

n : Banyaknya data

i : Indeks objek ke-i

k : Indeks *cluster* ke-k

Namun apabila perubahan nilai *membership function* masih diatas nilai *threshold* (ε), maka kembali ke langkah 4, dimana l : iterasi ke-t ; U : derajat keanggotaan [4]

C. Algoritma K-Modes

K-Modes hampir sama dengan metode K-Means. Kedua metode sama-sama menggunakan ukuran rata-rata untuk menentukan pusat clusternya. Bedanya jika K-Means menggunakan rata-rata, K-Modes menggunakan nilai Modus untuk menjadi pusat cluster nya. Adapun langkah untuk memperoleh cluster pada algoritma K-Modes adalah sebagai berikut [5].

1) Tentukan jumlah cluster
2) Alokasikan data ke dalam cluster secara random
3) Hitung modes dari data yang ada di masing-masing cluster
4) Alokasikan masing-masing data ke cluster terdekat menggunakan Hamming Distance
5) Kembali ke Step 3, apabila masih ada data yang berpindah cluster

Pada algoritma *K-Modes* nantinya akan dihitung akurasi dan cluster optimum untuk penentuan hasil cluster terbaik.

$$r = \frac{1}{n} \sum_{c=1}^{c} a_c \quad (4)$$

Dimana n adalah banyaknya data, a_c adalah banyaknya kategori yang mendominasi pada kelompok c. Akurasi akan disajikan dalam bentuk presentase, dengan mengalikan hasil r dengan 100%. *Error* atau kesalahan pengelompokkan juga dapat dihitung dari tingkat akurasi, yaitu $e = 1 - r$

D. Pseudo F Statistics

Penentuan jumlah *cluster* optimum yang pembentukan *cluster* ditentukan oleh jarak *Euclidean* sesuai metode yang digunakan, maka menggunakan *pseudo-statistic* [6]. Nilai *pseudo f-statistics* tertinggi menunjukkan bahwa jumlah kelompok telah optimal, dimana keseragaman dalam kelompok sangat homogen sedangkan antar kelompok sangat heterogen. Rumus yang digunakan dalam menghitung nilai *pseudo f-statistics* adalah sebagai berikut

$$Pseudo\ F\text{-}statistics = \frac{\left(\frac{R^2}{I-1}\right)}{\left(\frac{1-R^2}{N-I}\right)} \quad (5)$$

dengan

$$R^2 = \frac{(SST - SSW)}{SST} \quad (6)$$

$$SST = \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ni}^{j} - \bar{x}^{j})^2 \quad (7)$$

$$SSW = \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ni}^{j} - \bar{x}_{i}^{j})^2 \quad (8)$$

dimana

S : total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan

SS : total jumlah dari kuadrat jarak objek terhadap rata-rata kelompoknya

N : banyak objek

I : banyak *cluster*

j : banyak variabel

x_{ni}^{j} : sampel ke-n kelompok ke-i variabel ke$-j$

$\bar{x}^{j}$: rata-rata seluruh sampel pada variabel ke-j

$\bar{x}_{i}^{j}$: rata-rata sampel pada kelompok ke-i variabel ke-j

E. Ensemble Cluster

Pengelompokan Ensemble merupakan metode untuk menggabungkan beberapa algoritma yang berbeda untuk medapatkan partisi umum dari hasil pengelompokan individu

[7]. Tujuan dari pengelompokan ensemble adalah untuk menggabungkan hasil pengelompokan dari beberapa algoritma pengelompokan untuk mendapatkan hasil pengelompokan yang lebih baik [8]

Langkah-langkah dalam metode pengelompokan ensembel adalah sebagai berikut

1. Kumpulan data yang terdiri atas variabel kualitatif dan kontinu, dibagi menjadi dua subdata, yaitu murni kualitatif dan murni kontinu.
2. Lakukan pengelompokan objek dengan variabel berskala kuantitatif dengan pendekatan fuzzy C-Means
3. Lakukan pengelompokan objek dengan variabel berskala kualitatif dengan pendekatan K-Modes
4. Menggabungkan hasil pengelompokan dari langkah (2) dan langkah (3) yang disebut proses ensembel.
5. Lakukan pengelompokan pada langkah (4) menggunakan algoritma yang telah ditentukan untuk mendapatkan kelompok akhir.

F. Internal Dispersion Rate (Icdrate)

[9] Membandingkan metode *cluster* yang terbaik dengan mengevaluasi performansi algoritma dengan menggunakan prosentase rata-rata dari klasifikasi yang benar *(recovery rate)* dan nilai persebaran data-data dalam *cluster (internal cluster dispersion rate)* dari hasil akhir pengelompokkan.

$$Icdrate = 1-\frac{S}{S} = 1-\frac{S\ \ -SS}{S} = 1- R^2 \qquad (9)$$

dimana :

SST : Total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan.

SSW : Total jumlah dari kuadrat jarak sampel terhadap rata-rata

SSB : (*Sum Square Between*) SST-SSW

R^2 : (*Recovery Rate*) SSB/SST

G. One-Way *ANOVA*

Menentukan respon mana yang dipengaruhi oleh perlakuan yang dalam hal ini adalah hasil *Cluster* dapat diperoleh melalui Pengujian *One-way* ANOVA (*Analysis of Variance)*. Berikut adalah hipotesis yang digunakan dalam pengujian *One-way* ANOVA

H_0 : $\tau_1 = \tau_2 = ... = \tau_n$ (dimana n = banyak *cluster* yang terbentuk)

H_1 : minimal ada satu yang tidak sama

Statistik uji :

$$F_{hitung} = \frac{\frac{\sum_{t=1}^{T} n_t(\overline{X_t} - \overline{X})^2}{(t-1)}}{\frac{\sum_{j=1}^{J}\sum_{t=1}^{T}(\overline{X_t} - \overline{X_j})^2}{\sum_{t=1}^{T} n_t - t}} \qquad (10)$$

Tolak H_0 jika F_{hitung} lebih besar dari F_{tabel}. [10]

H. Persilangan Tanaman

Ilmu pemuliaan tanaman sebelumnya dikenal dengan nama ilmu seleksi karena dalam pelaksanaannya dilakukan pemilihan terhadap tanaman yang diinginkan, baik secara individu maupun kelompok. Ilmu pemuliaan digunakan untuk menemukan varietas – varietas baru dari proses pemuliaan. [11]

III. METODOLOGI PENELITIAN

A. Sumber data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari pengamatan di Balai Penelitian Jeruk dan Buah Subtropika (BALITJESTRO). Pengamatan ini dilakukan pada buah jeruk hasil persilangan dari kedua induk dan akan dibedakan antara data kuantitatif dan data kualitatif, data yang digunakan sebanyak 34 aksesi P5 (persilangan antara jeruk Siam Pontianak dengan jeruk Soe).

B. Variabel Penelitian

Variabel-variabel penelitian yang digunakan dalam penelitian ini antara lain sebagai berikut.

Variabel kuantitatif:

Tabel 1. Variabel Penelitian

X	Kuantitatif	X	Kualitatif
X1	Diameter Buah (cm)	X9	Bentuk Buah
X2	Tebal Kulit (mm)	X10	bentuk Ujung
X3	Jumlah Juring (Buah)	X11	Bentuk Pangkal
X4	Biji Normal (Buah)	X12	Warna Kulit
X5	Biji Abnormal (Buah)	X13	Permukaan Kulit
X6	Volume Jus (ml/50g)	X14	Keeratan Epicarp-Mesocarp
X7	Brix	X15	Tekstur Pulp
X8	Berat Buah (g)	X16	Rasa

C. Langkah Analisis

Langkah analisis yang dilakukan dalam penelitian ini sebagai berikut.

1. Mendiskripsikan data karakteristik tanaman jeruk dengan menggunakan ukuran means, varians, nilai minimum dan maksimum.
2. Melakukan pemisahan data kualitatif dan kuantitatif
3. Menentukan derajat keanggotaan dengan cara membangkitkan data yang memiliki syarat setiap baris harus berjumlah 1
4. Menentukan jumlah *cluster* optimum dengan metode *Fuzzy C-Means* melalui nilai *Pseudo-f statistics* terbesar. Sedangkan untuk K-*Modes* akan digunakan proporsi.
5. Mengelompokkan aksesi hasil persilangan tanaman jeruk berdasarkan karakteristik kuantitatif dan kualitatif sesuai dengan jumlah *cluster* optimum.
6. Melakukan cluster ensemble dari kedua algoritma yang digunakan, hasil cluster dari kedua algoritma akan menjadi data kualitatif berukuran nx2.
7. Melakukan *Final Cluster* terhadap aksesi hasil persilangan tanaman jeruk sesuai dengan jumlah *cluster* optimum menggunakan K-Modes dan menentukan karakteristik masing – masing cluster melalui nilai mean dan modus.
8. Melakukan uji *one-way* ANOVA pada ketiga metode.
9. Membandingkan ketiga metode menggunakan nilai *icdrate* dan nilai akurasi

IV. ANALISIS DAN PEMBAHASAN

A. Karakterisasi Data Persilangan

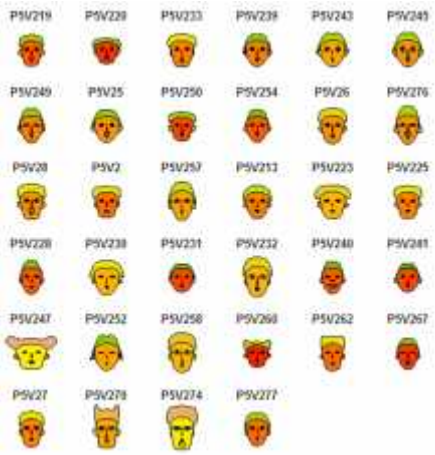
Data persilangan akan di karakteristik menggunakan statistika deskriptif untuk mengetahui secara visual bagaimana keadaan data persilangan. Berikut adalah hasil uraian statistika deskriptif.

Tabel 2. Deskripsi Data Kuantitatif Aksesi P5

Variabel	Mean	StDev	Outlier
Diameter Buah (cm)	5,77	0,52	3 data
Tebal Kulit (mm)	3,38	0,88	1 data
Jumlah Juring (buah)	11,00	0,57	0
Biji Normal (biji)	17,00	4,86	0
Biji Abnormal (biji)	4,00	2,13	1 data
Volume Jus (ml/50g)	28,84	3,05	2 data
Brix	12,19	1,73	1 data
Berat Buah (g)	93,56	21,47	2 data

Tabel 2 menjelaskan rata-rata bentuk buah dari aksesi P5 memiliki ukurn dengan diameter 5,77 cm, berat 93,56 g, rasa yang cukup manis, memiliki biji sebanyak 17. Kolom kanan dari niai mean adalah nilai standar deviasi, hanya berat buah yang memiliki standar deviasi terbesar yaitu 21,47 g. Selain itu dari 8 variabel, 6 variabel memiliki nilai outlier. Metode *fuzzy c-means cluster* sangat cocok untuk data yang memiliki varians kecil dan adanya *outlier*.

Deskripsi data selain menggunakan univariate, terdapat deskripsi data secara multivariate dengan menggunakan visualisasi wajah yang disebut *Chernoff Face*. Berikut hasilnya yang dipaparkan di Gambar 1.



Gambar 1. *Chernoff Face* Data Kuantitatif

Pengekspresian wajah mulai dari rambut, mata, hidung, telinga yang ada pada Gambar 1 menunjukkan variabel. Pada Tabel 3 akan dijelaskan dari semua variabel yang ada di data kuantitatif diwakili oleh ekspresi wajah seperti apa.

Tabel 3. Keterangan Bentuk *Chernoff Face*

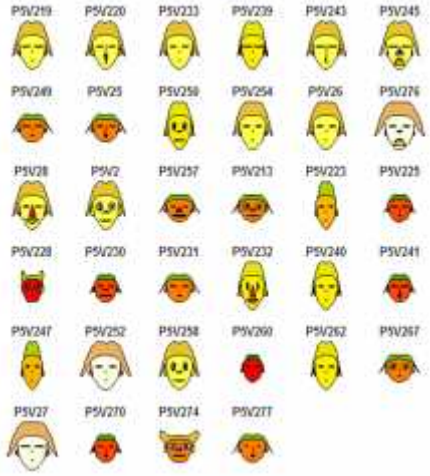
Deskripsi Wajah	Variabel
Tinggi Wajah	Diameter Buah
Lebar Wajah	Tebal Kulit Buah
Gaya Rambut	Jumlah Juring
Tinggi Mulut	Biji Normal
Lebar Mulut	Biji Abnormal
Lebar Telinga	Volume Jus
Tinggi Mata	Brix
Lebar Mata	Berat Buah

Variabel kuantitatif sebagian besar menjelaskan tentang bagian dalam dari sebuah jeruk. Selain variabel kuantitatif, terdapat variabel kualitatif yang dapat menjelaskan bentuk fisik dari sebuah jeruk. Terdapat 8 variabel kualitatif yang menjelaskan tentang bentuk fisik dari sebuah jeruk yaitu bentuk buah Obloid, pangkal buah Truncate, bentuk ujung buah Truncate, warna kulit Kuning kehijauan, permukaan kulit halus, keeratan epicarp mesocarp sedang, tekstur pulp lembut, dan rasa buah sedang.

Tabel 4. Deskripsi Data Kualitatif

Variabel	Deskripsi
Bentuk Buah	Obloid
Pangkal Buah	Truncate
Bentuk Ujung Buah	Truncate
Warna Kulit	Kuning Kehijauan
Permukaan Kulit	Halus
Kerekatan Epicarp-Mesocarp	Sedang
Tekstur Pulp	Lembut
Rasa	Sedang

Penerapan *Chernoff Face* juga akan diterapkan pada data kualitatif untuk melihat bagaimana visualisasi wajah terbentuk. Berikut hasilnya yang disajikan di Gambar 2.



Gambar 2. *Chernoff Face* Data Kualitatif

Gambar 2 menjelaskan setiap ekspresi wajah mulai dari rambut, mata, hidung dapat mewakili deskripsi dari setiap variabel yang ada. Pada Tabel 5 akan dijelaskan dari semua variabel yang ada di data kualitatif diwakili oleh ekspresi wajah seperti apa.

Tabel 5. Keterangan Bentuk *Chernoff Face*

Deskripsi Wajah	Variabel	Bentuk	Keterangan
Tinggi Wajah	Bentuk Buah	Tinggi	Obloid
		Pendek	Spheroid
Lebar Wajah	Pangkal Buah	Lebar	Concave
		Agak Lebar	Truncate
		Sedang	Convex
		Sempit	Necked
Gaya Rambut	Bentuk Ujung Buah	Rapi	Truncate
		Bertanduk	Depressed
Lebar Mulut	Permukaan Kulit	Sempit	Halus
		Lebar	Berpori
		Sedang	Halus Dengan Pori Samar
Ekspresi Senyum	Keeratan Epicarp-Mesocarp	Tak Berekspresi	Kuat
		Tersenyum Dan Datar	Sedang
		Kaget	Lepas
Tinggi Mata	Tekstur Pulp	Pendek	Lembut
		Tinggi	Sedang

B. Fuzzy C-Means

Penentuan cluster optimum dilakukan dengan menghitung nilai dari pseudo f-statistics yang tertinggi. Penentuan cluster optimum dilakukan dengan membentuk cluster yang berjumlah 2-4 cluster, dan selanjutnya setiap cluster akan dicari nilai pseudo f-statistics yang tertinggi, terpilih sebanyak 3 cluster karena memiliki nilai pseudo f-statistics tertinggi dengan 41,65. Berikut adalah anggota dari masing-masing kelompok.

Tabel 6. Anggota Cluster Fuzzy C-Means

Cluster 1	Cluster 2	Cluster 3	
P5V2 76	P5V2 20	P5V2 19	P5V2 23
P5V2 57	P5V2 33	P5V2 39	P5V2 25
P5V2 32	P5V2 5	P5V2 43	P5V2 28
P5V2 58	P5V2 50	P5V2 45	P5V2 30
P5V2 7	P5V2 31	P5V2 49	P5V2 40
P5V2 70	P5V2 41	P5V2 54	P5V2 52
P5V2 74	P5V2 47	P5V2 6	P5V2 60
		P5V2 8	P5V2 62
		P5V2	P5V2 67
		P5V2 13	P5V2 77

Anggota di cluster 1 adalah 7 aksesi, pada kelompok 2 sebanyak 7 aksesi, kelompok 3 sebanyak 15 aksesi, dan kelompok 4 sebanyak 3 aksesi. Berikut adalah karakteristik dari masing-masing cluster.

Tabel 7. Karakteristik Anggota Cluster Fuzzy

Cluster	X1	X2	X3	X4	X5	X6	X7	X8
1	6,48	3,18	11	20	4	28,50	12,16	124,82
2	5,23	3,83	11	18	4	28,57	13,17	67,36
3	5,68	3,25	10	16	4	28,99	11,85	90,48

C. Algoritma K-Modes

Perhitungan cluster optimum untuk data kualitatif adalah hanya dengan menghitung proporsi tiap kelompoknya. Pembentukan kelompok dilakukan antara 2-4, hasilnya kelompok 4 memiliki akurasi 100%. Oleh karena itu kelompok optimum pada data kualitatif akan ditentukan yaitu sebanyak 4.

Tabel 8. Anggota Cluster K-Modes

Cluster 1	Cluster 2	Cluster 3		Cluster 4
P5V2 50	P5V2 19	P5V2 20	P5V2 28	P5V2 57
P5V2	P5V2 39	P5V2 33	P5V2 30	P5V2 70
P5V2 13	P5V2 43	P5V2 5	P5V2 31	P5V2 77
P5V2 32	P5V2 45	P5V2 54	P5V2 40	
P5V2 58	P5V2 49	P5V2 6	P5V2 47	
P5V2 67	P5V2 41	P5V2 76	P5V2 52	
P5V2 74	P5V2 7	P5V2 8	P5V2 60	
		P5V2 23	P5V2 62	
		P5V2 25		

Kelompok 1 yaitu sebanyak 7 aksesi, pada kelompok 2 sebanyak 7 aksesi, kelompok 3 sebanyak 15 aksesi, dan kelompok 4 sebanyak 3 aksesi. Berikut adalah karakteristk dari masing – masing kelompok.

Tabel 9. Karakteristik Anggota Cluster K-Modes

Cluster	X1	X2	X3	X4	X5	X6	X7	X8
1	Obloid	Truncate	Truncate	Kuning Kehijauan	Halus	Sedang	Lembut	Sedang
2	Obloid	Truncate	Truncate	Kuning Oranye	Halus	Sedang	Lembut	Enak
3	Spheroid	Truncate	Truncate	Kuning	Halus	Sedang	Lembut	Buruk

D. Kombinasi Cluster *dengan Algoritma* Ensemble Cluster

Kombinasi Cluster menggunakan Algoritma Ensemble Cluster adalah tahap terakhir dari proses pengelompokkan data persilangan dengan aksesi P5. Hasil output dari pengelompokkan data kuantitatif dan pengelompokkan kualitatif dijadikan data baru dengan ukuran matriks 34×2. Data gabungan tersebut akan dikelompokkan menggunakan Algoritma K-Modes karena data gabungan tersebut otomatis telah menjadi data bertipe kualitatif. Jumlah kelompok yang akan dibentuk sudah ditentukan yaitu sebanyak 4 kelompok.

Tabel 10. Anggota Cluster Ensemble Cluster

Cluster 1		Cluster 2	Cluster 3	Cluster 4
P5V2 54	P5V2 28	P5V2 19	P5V2 50	P5V2 20
P5V2 6	P5V2 30	P5V2 39	P5V2 57	P5V2 33
P5V2 76	P5V2 40	P5V2 43	P5V2 32	P5V2 5
P5V2 8	P5V2 52	P5V2 45	P5V2 58	P5V2 31
P5V2	P5V2 60	P5V2 49	P5V2 70	P5V2 47
P5V2 13	P5V2 62	P5V2 41	P5V2 74	
P5V2 23	P5V2 67	P5V2 7		
P5V2 25	P5V2 77			

Kelompok 1 memiliki anggota yang paling besar yaitu sebanyak 16 aksesi, kelompok 2 sebanyak 7 aksesi, kelompok 3 sebanyak 6 aksesi, dan kelompok 4 sebanyak 5 aksesi. Berikut adalah karakteristik dari masing – masing kelompok.

Tabel 11. Karakteristik Anggota Cluster Ensemble

Cluster	X1	X2	X3	X4	X5	X6	X7	X8
1	5,81	3,26	10	16	5	28,77	12,48	96,35
2	5,23	3,83	11	18	4	28,57	13,17	67,36
3	6,34	3,06	11	21	3	29,38	11,83	116,81

Cluster	X1	X2	X3	X4	X5	X6	X7	X8
1	Obloid	Truncate	Truncate	Kuning Kehijauan	Halus	Sedang	Lembut	Sedang
2	Obloid	Truncate	Truncate	Kuning Kehijauan	Halus	Sedang	Lembut	Sedang
3	Spheroid	Convex	Truncate	Kuning	berpori	Sedang	Lembut	Buruk

E. Perbedaan Karakteristik Setiap Cluster

Perbandingan rata – rata setiap kelompok akan dilakukan pada hasil kelompok dari *fuzzy c-means, k-modes, dan ensemble cluster* adalah pada data kuantitatifnya saja. Untuk mengetahui variabel mana yang menjadi pembeda setiap kelompok akan menggunakan uji *one-way* ANOVA. Berikut adalah hasil dari uji *one-way* ANOVA untuk ketiga metode tersebut.

Tabel 12. One Way ANOVA

Variabel	Fuzzy	K-Modes	Ensemble
Diameter Buah	0,000*	0,166	0,002*
Tebal Kulit	0,284	0,687	0,179
Jumlah Juring	0,372	0,719	0,559
Biji Normal	0,164	0,513	0,044*
Biji Abnormal	0,874	0,154	0,450
Volume Jus	0,919	0,239	0,663
Brix	0,235	0,278	0,020*
Berat Buah	0,000*	0,178	0,000*

Hasil dari *p-value* pada tabel 4.10 adalah hasil dari uji ANOVA *one-way* pada metode *Fuzzy C-Means, K-Modes*, dan *Ensemble* yang telah dilampirkan pada Lampiran 12, Lampiran 13, dan 14. Nilai dari tabel 4.10 menunjukkan ada tidaknya perbedaan karakteristik antar *cluster*. Pada variabel terkait, metode *Fuzzy C-Means* diameter buahdan berat buah memiliki nilai p-value < α (5%), sedangkan pada *Ensemble* diameter buah, biji normal, brix, dan berat buah memiliki nilai p-value < α (5%). yang artinya tolak H_0 berarti terjadi perbedaan karakteristik dari variabel – variabel tersebut pada masing – masing *cluster*.

F. Perbandingan Ketiga Metode

Langkah terakhir dari penelitian ini adalah membandingkan antar ketiga metode yaitu *fuzzy c-means cluster*, *k-modes*, dan *ensemble cluster*. Untuk bisa membandingkan ketiga metode adalah dengan cara melihat nilai *icdrate* untuk data kuantitatif dan nilai akurasi yang dihitung dari proporsi untuk data kualitatif. Hasil dari nilai *icdrate* dan nilai akurasi dari masing – masing metode telah dijelaskan pada Tabel 13

Tabel 13. Nilai *Icdrate* dan akurasi

Metode	Data Kuantitatif	Data Kualitatif
	ICD Rate	Akurasi
Fuzzy	0,27	97%
K-Modes	0,93	100%
Gabungan	0,56	97%

Tabel diatas menghasilkan bahwa metode *fuzzy c-means* memiliki nilai *icdrate* yang terkecil, sedangkan untuk akurasi pada data kualitatif semuanya sama yaitu 97%. Hal ini berarti bahwa dengan metode *fuzzy c-means cluster* yang digunakan pada data kuantitatif cukup untuk mengelompokkan kedua tipe data. Caranya adalah pertama mengelompokkan data kuantitatif dengan menggunakan *fuzzy c-means cluster*, setelah itu aksesi dengan variabel kualitatif tinggal mengikuti anggota yang terbentuk pada metode *fuzzy c-means*.

V. KESIMPULAN DAN SARAN

Metode *Fuzzy C-Means Cluster* mampu mengelompokkan data persilangan dengan baik karena memiliki nilai *icdrate* yang rendah yaitu 0,27. Kondisi data yang memiliki varians kecil dan adanya *outlier* juga mendukung digunakannya metode *Fuzzy C-Means Cluster*. Pada metode K-Modes, mampu mengelompokkan data kategori dengan memiliki nilai akurasi 97%.

Perbandingan dari ketiga metode, *Fuzzy C-Means* menjadi algoritma terbaik untuk pengelompokkan data persilangan ini karena memiliki nilai *icdrate* terendah dan akurasi yang cukup tinggi. Sehingga dengan memakai data kuantitatif saja dengan algoritma *Fuzzy C-Means* mampu mengelompokkan kedua jenis data.

Saran yang keluar dari penelitian ini adalah, untuk penelitian selanjutnya dengan menambah metode pembanding untuk data kuantitatif dan kualitatif sehingga dapat mengetahui sejauh mana metode utama dapat bekerja dengan baik. Selain itu, lebih teliti lagi untuk menyeleksi variabel penelitian yang digunakan agar tidak ada salah 1 variabel yang mendominasi variabel lain dan ditambahkan data tetua nya.

DAFTAR PUSTAKA

[1] Martasari, C., 2014. *Kajian Genetik dan Percepatan Pembungaan Tanaman Hasil Fusi Protoplasma Jeruk Siam Madu (Citrus nobilis Lour.) dan Satsuma Mandarin (Citrus unshiu).* Malang: Universitas Brawijaya

[2] Naba, Agus. (2009). *Belajar Cepat Fuzzy Logic Menggunakan MATLAB.* Yogyakarta : CV ANDI OFFSET.

[3] Kusumadewi, Sri dan Hari Purnomo. (2004). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan.* Yogyakarta : Graha Ilmu.

[4] Bezdek, J.C., Ehrlich, R., Full, W. 1984. *FCM: Fuzzy C-Means Clustering Algorithm.* USA: Computers & Geosciences Vol. 10, No. 2-3, pp. 191-203

[5] Huang, Z. And Ng.M. (1999). A fuzzy K-Modes algorithm for clustering categorical data. *IEEE Transactions Fuzzy System*, 7(4): 446-452

[6] Orpin, A., & Kostylev, V. (2006). Towards a Statistically Valid Method of Textural Sea Floor Characterization of Benthic Habitats. *Marine Geology*, 209-222.

[7] Dewi, A., 2012. *Metode Cluster Ensemble Untuk Pengelompokan Desa Perdesaan di Provinsi Riau. Thesis,* Jurusan Statistika FMIPA-ITS, Surabaya

[8] Yoon, H. S. 2006. Heterogeneous Clustering Ensemble Method For Combining Different Cluster Results. *BioDM 2006*, pp. 82-91

[9] Mingoti, S., & Lima, J. (2006). Comparing SOM Neural Network with Fuzzy C-Means, C-Means and Traditional Hierarchical Clustering Algorithms. *European Journal of Operational Research, 174*, 1742-1759.

[10] Johnson, R. A. and Wichern, D. W. 2007. .*Applied Multivariat Analysis, Sixth Edition.* Prentice Hall Inc. New Jersey.

[11] Mangoendidjojo, W. 2003. *Dasar – Dasar Pemuliaan Tanaman Edisi ke-6*. Yogyakarta: Penerbit Kanisius (Anggota IKAPI)