PENERAPAN METODE *K-MEDOIDS* PADA PENGELOMPOKAN DAERAH PENGHASIL KELAPA SAWIT DENGAN VALIDASI INDEKS *SILHOUETTE*

Ewaldus Okta, Neva Satyahadewi, Naomi Nesyana Debataraaja

INTISARI

K-medoids termasuk metode *partitioning clustering* untuk mengelompokkan sekumpulan n objek menjadi sekumpulan k cluster. Penelitian ini dilakukan untuk mengelompokkan kabupaten/kota yang memiliki luas lahan dan rata-rata produksi perkebunan kelapa sawit tertinggi sampai terendah. Tahap ini diawali dengan menghitung kemiripan antar objek menggunakan Jarak Euclidean, kemudian memilih k objek untuk menjadi medoids awal. Langkah selanjutnya adalah mengelompokkan objek berdasarkan nilai Jarak yang paling mirip dengan medoids, lalu menghitung nilai absolute error dari cluster yang terbentuk. Tahap selanjutnya adalah melakukan hal yang sama terhadap medoids random, kemudian membandingkan nilai absolute error medoids awal dengan medoids random. Langkah selanjutnya adalah mengukur validitas dengan menggunakan validasi indeks *silhouette*. Cluster kesatu yaitu Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Mempawah, Kabupaten Kapuas Hulu, Kabupaten Kayong Utara, Kabupaten Kubu Raya dan Kota Singkawang untuk variabel luas lahan dan rata-rata produksi termasuk kedalam kategori rendah. Cluster kedua yaitu Kabupaten Landak, Kabupaten Sanggau, Kabupaten Ketapang dan Kabupaten Sintang untuk variabel luas lahan dan rata-rata produksi termasuk kedalam kategori sedang. Cluster ketiga yaitu Kabupaten Sekadau dan Kabupaten Melawi untuk variabel luas lahan termasuk kedalam kategori rendah dan rata-rata produksi termasuk kedalam kategori tinggi.

Kata kunci: Analisis Cluster, *K-medoids*, Validasi Indeks *Silhouette*.

PENDAHULUAN

Peraturan Daerah Kalimantan Barat No.1 Tahun 1995 tentang Rencana Tata Ruang Wilayah Provinsi (RTRWP) Kalimantan Barat dengan jelas merencanakan kelapa sawit sebagai salah satu sumber pendapatan daerah pengganti kayu yang saat ini hampir habis. Hasil perkebunan kelapa sawit ini memberikan kontribusi terhadap pembangunan di daerah Kalimantan Barat dan salah satu mata pencaharian di Kalimantan Barat. Menurut Dinas Perkebunan Provinsi Kalimantan Barat pada tahun 2015, lahan terluas yang digunakan untuk perkebunan kelapa sawit terletak di Kabupaten Ketapang dengan luas lahan 350.201 hektar. Perkebunan kelapa sawit terluas kedua terletak di Kabupaten Sanggau dengan luas lahan 257.528 hektar, sedangkan untuk terluas ketiga terletak di Kabupaten Sintang dengan luas lahan 151.587 hektar

Berdasarkan fakta-fakta tersebut dilakukan pengelompokan kabupaten/kota di Kalimantan Barat berdasarkan luas lahan dan rata-rata produksi perkebunan kelapa sawit pada tahun 2015 dengan menggunakan metode *K-medoids*. Pengelompokan ini dilakukan untuk mengetahui kabupaten/kota yang memiliki luas lahan dan rata-rata produksi perkebunan kelapa sawit tertinggi sampai terendah. Analisis *cluster* mengelompokkan objek yang paling mirip kedalam satu kelompok yang sama, sehingga karakteristik objek dalam satu *cluster* bersifat homogen sedangkan karakteristik objek antar *cluster* yang lain bersifat heterogen [1]. Analisis *cluster* pada dasarnya ada dua yaitu metode hierarki dan metode non-hierarki. Analisis *cluster* yang menggunakan prosedur partisi atau metode non-hierarki salah satunya adalah *K-medoids*.

K-medoids atau *Partitioning Around Medoid* (PAM) ditemukan oleh Kaufman dan Rousseeuw pada tahun 1990. Dalam prosedur *partitioning cluster* objek data yang dikelompokkan, dipisah atau disekat menjadi k -cluster dimana jumlah k tersebut kurang dari atau sama dengan banyaknya data yang di kelompokkan. *Medoids* adalah objek yang dianggap mewakili *cluster* sekaligus sebagai pusat

cluster. Analisis ini mencoba meminimumkan ketidaksamaan setiap objek dalam satu *cluster* dengan meminimumkan nilai *absolute error* [2].

Penelitian ini bertujuan untuk mengaplikasikan metode *K-medoids* untuk mengelompokkan kabupaten atau kota berdasarkan luas lahan dan rata-rata produksi perkebunan kelapa sawit di provinsi Kalimantan Barat pada tahun 2015. Ukuran Jarak yang digunakan untuk menghitung kemiripan antar objek adalah Jarak *Euclidean*. Jumlah *cluster* yang dibentuk dalam penelitian ini sebanyak tiga *cluster*. Ukuran validasi yang digunakan adalah dengan menggunakan validasi indeks *Silhouette*.

Langkah pertama yang dilakukan adalah standarisasi data, setelah dilakukan standarisasi maka langkah kedua adalah pengecekan multikolinearitas pada variabel-variabel yang digunakan dalam penelitian. Selanjutnya langkah ketiga yang harus dilakukan adalah melakukan analisis *cluster*. Metode *cluster* yang digunakan dalam penelitian ini adalah dengan metode *K-medoids*. Adapun langkah-langkah pengerjaannya adalah sebagai berikut:

- Menghitung kemiripan antara objek dengan menggunakan Jarak *Euclidean*.
- Memilih k objek untuk menjadi O_c , dengan O_c adalah objek yang menjadi *medoids* awal di *cluster* ke- c dan $c = 1, 2, \dots, k$
- Mengelompokkan objek *non-medoids* ke dalam kelompok yang paling dekat dengan *medoids*.
- Menghitung nilai *absolute error* O_c .
- Secara acak memilih O_{random} , dengan O_{random} adalah sebuah objek *non-medoids* untuk menggantikan O_c .
- Mengulangi langkah b sampai d untuk objek O_{random} .
- Membandingkan nilai *absolute error* sebelum dan sesudah pertukaran O_c dengan O_{random} , jika $E_{random} < E_c$ maka tukar O_c dengan O_{random} , tetapi jika $E_{random} > E_c$ maka O_c tetap
- Mengulangi langkah e sampai g hingga semua objek *non-medoids* terpilih menjadi O_{random} .

Langkah keempat adalah dengan mengukur validitas dengan menggunakan validasi indeks *Silhouette*. langkah kelima adalah menentukan interval masing-masing kategori. Langkah keenam adalah membuat profil *cluster*, sehingga diperoleh anggota-anggota kelompok dalam satu *cluster*.

ANALISIS CLUSTER

Analisis *cluster* adalah teknik multivariat yang mempunyai tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengelompokkan objek yang paling mirip kedalam satu kelompok yang sama, sehingga karakteristik objek dalam satu *cluster* bersifat homogen sedangkan karakteristik objek antar *cluster* yang lain bersifat heterogen. Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. Tujuan utama analisis *cluster* adalah untuk menempatkan sekumpulan objek kedalam dua atau lebih grup berdasarkan objek atas dasar beberapa karakteristik.

Suatu *cluster* dikatakan baik jika mempunyai homogenitas (kesamaan) yang tinggi antar anggota *cluster* (*within cluster*) dan heterogenitas (perbedaan) yang tinggi antar *cluster* yang satu dengan yang lainnya (*between cluster*). Analisis *cluster* ini terdiri dari beberapa proses dasar yaitu dengan mengukur kesamaan antar objek (*similarity*), dilakukan proses standarisasi jika data memiliki satuan yang berbeda atau bervariasi, kemudian dilakukan dengan membuat *cluster*. Analisis *cluster* terbagi atas dua metode yaitu metode hierarki dan metode non-hierarki.

K-MEDOIDS

K-medoids termasuk metode *partitioning clustering* untuk mengelompokkan sekumpulan n objek menjadi sekumpulan k *cluster*. Representasi *cluster* pada *K-medoids* adalah objek dari sekumpulan objek yang mewakili *cluster* yang disebut *medoids* [3]. Dalam metode *K-medoids* dipilih satu objek representatif yang digunakan untuk mewakili *cluster*. *Cluster* dibentuk dengan menghitung kedekatan antar objek dengan menggunakan ukuran kedekatan. Setiap objek yang tidak dipilih menjadi *medoids*

dikelompokkan dengan objek representatif yang paling mirip. Setiap objek yang tidak dipilih menjadi *medoids* dikelompokkan dengan objek representatif yang paling mirip. Metode ini dilakukan berdasarkan prinsip meminimalkan jumlah ketidaksamaan antara masing-masing objek dalam satu *cluster*, dengan menghitung nilai *absolute error* yang dirumuskan sebagai berikut:

$$E = \sum_{r=1}^m \sum_{p \in C_r} |p - o_r| \quad (1)$$

dengan:

E : jumlah *absolute error* untuk semua objek dalam suatu kumpulan data

p : objek *non-medoids* dalam suatu *cluster*

o_r : objek yang menjadi *medoids* dalam suatu *cluster*

C_r : *cluster* ke- r , dengan $r = 1, 2, \dots, m$

m : banyaknya *cluster* yang terbentuk.

STANDARISASI DATA

Standarisasi data adalah proses mengubah masing-masing observasi menjadi nilai standar dengan rata-rata 0 dan standar deviasi 1 untuk menghilangkan bias yang disebabkan karena perbedaan skala dari beberapa variabel yang digunakan dalam analisis. Untuk menghitung nilai standar tersebut dapat digunakan rumus sebagai berikut:

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (2)$$

Dengan:

Z_i : nilai yang telah standarisasi (*Z-score*)

x_i : observasi ke- i sebelum distandarisasi

$\bar{x}$: rata-rata dari data

s : simpangan baku data.

UKURAN KESAMAAN ANTAR OBJEK

Mengukur kesamaan antar objek merupakan bagian penting yang harus diperhatikan karena memengaruhi perhitungan dalam menentukan anggota suatu *cluster*. Dalam analisis *cluster* terdapat tiga metode yang digunakan dalam mengukur kesamaan antar objek, yaitu ukuran korelasi, ukuran jarak, dan ukuran asosiasi. Mengukur kedekatan antar objek dengan menggunakan ukuran jarak dapat diterapkan pada data kuantitatif, sebenarnya metode ini merupakan ukuran ketidakmiripan. Dalam hal ini jarak yang besar menunjukkan sedikit kesamaan, sebaliknya jarak yang kecil menunjukkan bahwa suatu objek semakin mirip dengan objek lain. Ada beberapa tipe ukuran jarak dan salah satunya adalah jarak *euclidean*. Ukuran jarak *euclidean* adalah pengukuran jarak objek dan pusat *cluster* yang sering digunakan dalam proses pengelompokan. Cara menghitung Jarak antar objek dengan menggunakan Jarak *Euclidean* dapat digunakan persamaan sebagai berikut [4]:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3)$$

dengan:

d_{ij} : nilai Jarak *euclidean* antara objek ke- i dan objek ke- j

x_{ik} : nilai atau data dari objek ke- i pada variabel ke- k , dengan $k = 1, 2, \dots, p$

x_{jk} : nilai atau data dari objek ke- j pada variabel ke- k , dengan $k = 1, 2, \dots, p$

p : banyaknya variabel./

UJI MULTIKOLINEARITAS

Multikolinearitas antar variabel adalah suatu pelanggaran dalam analisis *cluster* karena dapat memengaruhi hasil *cluster*. Untuk mengetahui ada atau tidaknya multikolinearitas antar variabel dapat digunakan nilai *variance inflation factor* (VIF) dengan menggunakan rumus sebagai berikut

$$VIF = \frac{1}{1 - R^2} \quad (4)$$

Dengan kriteria pengujian jika nilai $VIF < 10$, dapat disimpulkan bahwa tidak terjadi multikolinearitas antar variabel, tetapi apabila nilai $VIF > 10$, berarti terdapat multikolinearitas antar variabel.

VALIDASI INDEKS *SILHOUETTE*

Validasi indeks *silhouette* merupakan metode yang digunakan untuk melihat kualitas dari sebuah *cluster*. Validasi indeks *silhouette* pertama kali diperkenalkan sebagai teknik yang digunakan untuk mengevaluasi objek secara visual yang berada baik didalam *cluster* maupun yang berada diluar *cluster*, berdasarkan rentang nilai *silhouettenya*. Nilai *silhouette* objek mendefinisikan kedekatannya dengan *clusternya* sendiri dengan kelompok lain. Rentang nilai *silhouette* dirumuskan sebagai berikut [5]:

$$s_i = \frac{b_i - a_i}{\max[b_i, a_i]} \quad (5)$$

dengan:

s_i : nilai koefisien *silhouette* objek ke- i dengan $i = 1, 2, \dots, n$

a_i : rata-rata Jarak antara objek ke- i dengan objek lainnya dalam satu *cluster*

b_i : minimum rata-rata Jarak antara objek ke- i dengan objek lainnya di masing-masing *cluster*.

Jika nilai koefisien *silhouette* < 0 maka objek berada dalam kelompok yang salah, sedangkan jika nilai koefisien *silhouette* > 0 maka objek berada dalam kelompok yang benar dan jika nilai koefisien *silhouette* $= 0$ maka objek berada diantara dua *cluster* sehingga objek tersebut belum dapat ditentukan masuk didalam kelompok yang benar atau salah [6]. Untuk menentukan rentang objek secara keseluruhan dapat dirumuskan sebagai berikut:

$$SC = \frac{\sum_{i=1}^n s_i}{n} \quad (6)$$

dengan:

SC : rentang nilai dari koefisien *silhouette*

n : banyaknya data

Ketika rentang nilai dari koefisien *silhouette* $> 0,51$ menunjukkan bahwa struktur hasil pengelompokan sudah dianggap baik, sedangkan ketika rentang nilai dari koefisien *silhouette* $> 0,71$ menunjukkan bahwa struktur hasil pengelompokan dianggap sangat baik.

INTERPRETASI DAN PEMBUATAN PROFIL *CLUSTER*

Interpretasi dilakukan dengan menghitung rata-rata hasil pengelompokan sehingga memberikan gambaran yang logis tentang hasil *cluster*, perhitungan rata-rata tersebut dilakukan pada data asli, jika data telah distandarisasi atau jika analisis *cluster* dilakukan dengan menggunakan hasil analisis-analisis komponen utama, peneliti harus mengembalikan ke nilai asli, baru kemudian menghitung rata-rata menggunakan data tersebut.

APLIKASI NUMERIK

Variabel pada data yang digunakan adalah luas lahan dengan satuan hektar dan rata-rata produksi dengan satuan kg/hektar. Data memiliki satuan yang berbeda, oleh karena itu langkah pertama yang

dilakukan adalah dengan menstandarisasi data. Data yang telah distandarisasi dengan menggunakan Persamaan (2) disajikan pada Tabel 1.

Tabel 1 Data yang Telah Distandarisasi

Kabupaten/Kota	Luas areal	Rata-rata produksi
Sambas	-0,24414	-0,14502
Bengkayang	-0,19933	0,02122
Landak	0,03791	0,27765
Mempawah	-0,94860	-1,40419
Sanggau	1,52825	0,79582
Ketapang	2,48253	0,36608
Sintang	0,43734	0,61720
Kapuas Hulu	-0,23062	-1,28216
Sekadau	-0,09593	0,51286
Melawi	-0,72351	2,27075
Kayong Utara	-0,73835	-0,93200
Kubu Raya	-0,24884	-0,19807
Singkawang	-1,05673	-0,90016

Langkah kedua adalah pengecekan multikolinearitas, mengetahui ada atau tidaknya multikolinearitas dapat digunakan nilai *Variance Inflation Factor* (VIF). Dengan menggunakan Persamaan (4) nilai VIF dari variabel luas lahan dan variabel rata-rata produksi diperoleh sebesar 1,129331 yang berarti $VIF < 10$. Dengan demikian dapat disimpulkan tidak terjadi multikolinearitas antar variabel.

Langkah ketiga adalah melakukan analisis *cluster* dengan menerapkan metode *K-medoids*. Dalam melakukan analisis *cluster*, tahapan ini diawali dengan menghitung ukuran Jarak kemiripan antar objek. Salah satu perhitungan dalam penelitian ini adalah menghitung Jarak *Euclidean* Kabupaten Sambas dan Kabupaten Bengkayang. Dengan Menggunakan Persamaan (3) Jarak *Euclidean* antara Kabupaten Sambas dan Kabupaten Bengkayang adalah 0,17217.

Tahapan selanjutnya adalah memilih objek yang menjadi *medoids*, dalam penelitian ini dipilih 3 objek menjadi *medoids* awal yaitu Kabupaten Sambas, Kabupaten Bengkayang dan Kabupaten Landak. Kemudian dilanjutkan dengan memilih secara acak objek yang bukan *medoids* sebelumnya menjadi *medoids* untuk menggantikan objek yang terpilih sebagai *medoids* awal. Pada penelitian ini dipilih objek yang menjadi *medoids* untuk menggantikan *medoids* awal adalah Kabupaten Sambas, Kabupaten Landak dan Kabupaten Ketapang. Pergantian objek *non-medoids* menjadi *medoids* dilakukan terus menerus sehingga semua objek menjadi *medoids*.

Setelah dilakukan pengelompokan, langkah selanjutnya adalah membandingkan nilai *absolute error* sebelum dan sesudah pertukaran objek awal yang menjadi *medoids* dengan objek *random* yang menjadi *medoids*. Objek-objek yang menjadi *medoids* awal adalah Kabupaten Sambas, Kabupaten Bengkayang dan Kabupaten Landak. Dengan menggunakan Persamaan (1) diperoleh nilai *absolute error medoids* awal sebesar 51,21221. Sedangkan objek-objek yang menjadi *medoids random* adalah Kabupaten Sambas, Kabupaten Landak dan Kabupaten Ketapang. Dengan menggunakan Persamaan (1) diperoleh nilai *absolute error medoids random* adalah 71,40835.

Cluster yang terbentuk dengan nilai *absolute error* yang paling kecil yaitu *cluster* dengan Kabupaten Sambas, Kabupaten Bengkayang dan Kabupaten Kubu Raya sebagai *medoids* dari *cluster* tersebut. Nilai *absolute error* dari *cluster* tersebut adalah 51,10284. Hasil *cluster* yang terbentuk dengan Kabupaten Sambas, Kabupaten Bengkayang dan Kabupaten Kubu Raya sebagai *medoids*,

Cluster pertama dengan anggota kelompoknya Kabupaten Sambas. *Cluster* kedua dengan anggota kelompoknya adalah Kabupaten Bengkayang, Kabupaten Landak, Kabupaten Sanggau, Kabupaten Ketapang, Kabupaten Sintang, Kabupaten Sekadau dan Kabupaten Melawi. *Cluster* ketiga dengan anggota kelompoknya adalah Kabupaten Mempawah, Kabupaten Kapuas Hulu, Kabupaten Kayong Utara, Kabupaten Kubu Raya dan Kota Singkawang

Setelah melakukan analisis *cluster* dengan menggunakan metode *K-medoids*, langkah keempat adalah mengukur validitas dengan menggunakan validasi indeks *Silhouette*. Pada penelitian ini kriteria hasil pengelompokan yang dipilih adalah hasil pengelompokan yang memenuhi syarat nilai koefisien *silhouette* (s_i) dan memiliki nilai *absolute error* paling kecil. Pengelompokan dengan Kabupaten Sambas, Kabupaten Bengkayang dan Kabupaten Kubu Raya sebagai *medoids* tidak memenuhi syarat nilai koefisien *silhouette* (s_i), karena nilai koefisien *silhouette* objek kesatu (s_1), objek ketiga (s_3), objek ketujuh (s_7), objek kesembilan (s_9) dan objek keduabelas (s_{12}) kurang dari 0 sehingga dapat disimpulkan objek-objek tersebut berada dalam kelompok yang salah.

Pengelompokan dengan Kabupaten Sambas, Kabupaten Landak dan Kabupaten Sekadau sebagai *medoids* merupakan hasil pengelompokan yang memenuhi syarat nilai koefisien *silhouette* (s_i) dan memiliki nilai *absolute error* paling kecil. Dengan menggunakan Persamaan (1) nilai *absolute error* hasil pengelompokan tersebut sebesar 52,00999. Anggota dari *cluster* pertama adalah Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Mempawah, Kabupaten Kapuas Hulu, Kabupaten Kayong Utara, Kabupaten Kubu Raya dan Kota Singkawang. *Cluster* kedua anggotanya adalah Kabupaten Landak, Kabupaten Sanggau, Kabupaten Ketapang dan Kabupaten Sintang. *Cluster* ketiga anggotanya adalah Kabupaten Sekadau dan Kabupaten Melawi.

Dengan menggunakan Persamaan (5) didapat nilai koefisien *silhouette* (s_i) masing-masing *cluster*. Nilai koefisien *silhouette* (s_i) *cluster* kesatu adalah 0,61369. Nilai koefisien *silhouette* (s_i) *cluster* kedua adalah 0,13698. Nilai koefisien *silhouette* (s_i) masing-masing objek hasil pengelompokan terbaik disajikan pada Tabel 2.

Tabel 2 Tabel Nilai Koefisien *Silhouette* (s_i) Hasil Pengelompokan Terbaik

Objek ke-	Koefisien <i>Silhouette</i> (s_i)
1	0,55884
2	0,42447
3	0,0039
4	0,69311
5	0,57034
6	0,56008
7	0,2803
8	0,65205
9	0,26008
10	0,64698
11	0,7396
12	0,58255
13	0,69756

Berdasarkan Tabel 2 nilai koefisien *silhouette* (s_i) dari masing-masing objek lebih dari 0 sehingga dapat disimpulkan objek sudah berada dalam kelompok yang benar. Langkah selanjutnya adalah menghitung rentang nilai koefisien *silhouette* (SC). Dengan menggunakan Persamaan (6) rentang nilai dari koefisien *silhouette* (SC) yang diperoleh yaitu sebesar 0,51307, maka dapat disimpulkan bahwa

struktur pengelompokan sudah baik karena rentang nilai dari koefisien *silhouette* (SC) lebih besar dari 0,51.

Langkah kelima adalah menentukan interval untuk masing-masing kategori. Variabel luas lahan suatu *cluster* masuk kedalam kategori rendah jika berada pada interval 6.493 ha sampai 121.061,33 ha. Luas lahan suatu *cluster* masuk kedalam kategori sedang jika berada pada interval 121.062,33 ha sampai 235.630,6667 ha. Luas lahan suatu *cluster* masuk kedalam kategori tinggi jika berada pada interval 235,631,663 ha sampai 350.200 ha. Sedangkan untuk variabel rata-rata produksi, suatu *cluster* masuk kedalam kategori rendah jika berada pada interval 1.218 kg/ha sampai 1.909,67 kg/ha. Rata-rata produksi suatu *cluster* masuk kedalam kategori sedang jika berada pada interval 1.910,67 kg/ha sampai 2.602,33 kg/ha. Rata-rata produksi suatu *cluster* masuk kedalam kategori tinggi jika berada pada interval 2.603,336 kg/ha sampai 3.295 kg/ha.

Langkah keenam adalah menginterpretasi dan membuat profil dari *cluster* yang terbentuk. Hasil dari rata-rata luas lahan dan jumlah produksi tiap *cluster* dengan anggota masing-masing *cluster* disajikan pada Tabel 3.

Tabel 3 Hasil Rata-Rata Luas Lahan dan Rata-Rata Produksi Tiap *Cluster*

Cluster Ke-	Kabupaten/Kota	Rata-rata	
		Luas lahan	Rata-rata produksi
1	Sambas, Bengkayang, Mempawah, Kapuas Hulu, Kayong Utara, Kubu Raya dan Singkawang	58.247,43 ha	1.621 kg/ha
2	Landak, Sanggau, Ketapang, Sintang dan	194.382 ha	2.303 kg/ha
3	Sekadau dan Melawi	69.326 ha	2.799 kg/ha

Dari Tabel 3 dapat disimpulkan bahwa Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Kapuas Hulu, Kabupaten Kayong Utara, Kabupaten Kubu Raya dan Kota Singkawang memiliki rata-rata luas lahan yang termasuk kedalam kategori rendah yaitu sebesar 58.247,43 ha. Kabupaten Sekadau dan Kabupaten Melawi juga merupakan kabupaten yang memiliki rata-rata luas lahan yang termasuk kedalam kategori rendah yaitu sebesar 69.326 ha. Sedangkan Kabupaten Landak, Kabupaten Sanggau, Kabupaten Ketapang dan Kabupaten Sintang memiliki rata-rata luas lahan yang termasuk kedalam kategori sedang yaitu sebesar 194.382 ha.

Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Kapuas Hulu, Kabupaten Kayong Utara, Kabupaten Kubu Raya dan Kota Singkawang memiliki rata-rata dari rata-rata produksi yang termasuk kedalam kategori rendah yaitu sebesar 1.621 kg/ha., Kabupaten Landak, Kabupaten Sanggau, Kabupaten Ketapang dan Kabupaten Sintang memiliki rata-rata dari rata-rata produksi yang termasuk kedalam kategori sedang yaitu sebesar 2.303 kg/ha. Sedangkan Kabupaten Sekadau dan Kabupaten Melawi memiliki rata-rata dari rata-rata produksi yang termasuk kedalam kategori tinggi yaitu sebesar 2.799 kg/ha.

PENUTUP

Cluster kesatu dengan anggota kelompoknya yaitu Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Mempawah, Kabupaten Kapuas Hulu, Kabupaten Kayong Utara, Kabupaten Kubu Raya dan Kota Singkawang untuk variabel luas lahan dan rata-rata produksi termasuk kedalam kategori rendah. dan *Cluster* kedua dengan anggota kelompoknya yaitu Kabupaten Landak, Kabupaten Sanggau, Kabupaten Ketapang dan Kabupaten Sintang untuk variabel luas lahan dan rata-rata

produksi termasuk kedalam kategori sedang. *Cluster* ketiga dengan anggota kelompoknya yaitu Kabupaten Sekadau dan Kabupaten Melawi untuk variabel luas lahan termasuk kedalam kategori rendah dan rata-rata produksi termasuk kedalam kategori tinggi.

DAFTAR PUSTAKA

- [1]. Hair, J.F., Black, W.C., Babin, B.J., dan Anderson, R.E.. *Multivariate Data Analysis, Seventh Edition*. New Jersey: Prentice Hall; 2010
- [2]. Han, J dan Kamber, M. *Data Mining: Concepts and Techniques, Second Edition*. San Fransisco; Elvesier Inc; 2006
- [3]. Kaufman, L., dan Rousseeuw, Peter J. *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons; 2005
- [4]. Supranto, J. *Analisis Multivariat: Arti dan Interpretasi*. Jakarta: Rineka Cipta; 2010
- [5]. Rousseeuw, Peter J. *Silhouette: A Graphical Aid to The Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics*, **20**:53-65. 1987
- [6]. Lewis, Paul D. *R for Medicine and Biology*. Massachusets: Jones and Bartlett Publisher; 2010

EWALDUS OKTA : Jurusan Matematika FMIPA UNTAN
okta.ewaldus2012@gmail.com

NEVA SATYAHADEWI : Jurusan Matematika FMIPA UNTAN Pontianak,
neva.satya@math.untan.ac.id

NAOMI NESSYANA DEBATARAJA : Jurusan Matematika FMIPA UNTAN Pontianak,
naominessyana@math.untan.ac.id