

## Metode *K-Means Cluster* Untuk Mengelompokkan Kota/Kabupaten di Sumatera Selatan Berdasarkan Produksi Ikan Air Tawar

Yunita Febriani<sup>1</sup>, Yulia Puspita Sari<sup>2\*</sup>, Dina Octaria<sup>3</sup>  
\*e-mail: [yuliaaman91@gmail.com](mailto:yuliaaman91@gmail.com)

<sup>1</sup>*Program Studi Manajemen, Fakultas Ekonomi dan Bisnis  
Universitas Muhammadiyah Palembang*

<sup>2</sup>*Program Studi Budidaya Ikan, Fakultas Perikanan dan Kelautan  
Universitas PGRI Palembang*

<sup>3</sup>*Program Studi Matematika, Fakultas Keguruan dan Ilmu Pendidikan  
Universitas PGRI Palembang*

### ABSTRACT

The distribution of freshwater fish production in South Sumatra can be built into a mapping. The mapping can be done using one of the cluster analysis methods, namely the K-Means Cluster. K-Means Cluster is a clustering algorithm that groups data based on the cluster center point closest to the data. K-Means Cluster groups data by maximizing the similarity of data in one cluster and minimizing data between clusters. This study aims to classify cities/districts in South Sumatra based on freshwater fishery production using the K-Means Cluster method. This mapping in the freshwater fishery sector can assist the provincial government in the management of policy making, so as to increase the output of freshwater fishery production in every city/district in South Sumatra. The results showed that 2 clusters were formed. The determination of the number of clusters is based on the results of the curve of the Elbow, Silhouette and Gaps Statistics methods. Cluster 1 is in the low category, consisting of 12 cities/districts. While cluster 2 consists of 2 cities/districts. and is a high category cluster.

**Keywords:** Freshwater Fish Production, K-Means Cluster, South Sumatra

### ABSTRAK

Sebaran produksi ikan air tawar di Sumatera Selatan dapat dibangun menjadi sebuah pemetaan. Pemetaan tersebut dapat dilakukan dengan menggunakan salah satu metode analisis kluster yaitu metode *K-Means Cluster*. *K-Means Cluster* merupakan algoritma clustering yang mengelompokkan data berdasarkan titik pusat cluster yang paling dekat dengan data tersebut. *K-Means Cluster* mengelompokkan data dengan memaksimalkan kesamaan data dalam satu cluster dan meminimalkan data antar cluster. Penelitian ini bertujuan untuk mengklasifikasikan kota/kabupaten di Sumatera Selatan berdasarkan produksi perikanan air tawar menggunakan metode *K-Means Cluster*. Pemetaan di sektor perikanan air tawar ini dapat membantu pemerintah provinsi dalam manajemen penyusunan kebijakan, sehingga dapat meningkatkan output produksi perikanan air tawar di setiap kota/kabupaten di Sumatera Selatan. Hasil penelitian menunjukkan bahwa terbentuk 2 cluster. Penentuan jumlah cluster didasarkan pada hasil kurva metode *Elbow*, *Silhouette* dan *Gaps Statistics*. Kluster 1 termasuk kategori rendah, terdiri dari 12 kota/kabupaten. Sedangkan cluster 2 terdiri dari 2 kota/kabupaten. dan merupakan kluster kategori tinggi.

**Kata Kunci:** Produksi Ikan Air Tawar, *K-Means Cluster*, Sumatera Selatan

## PENDAHULUAN

Pada sektor perikanan, Provinsi Sumatera Selatan memiliki jenis hasil perikanan air tawar yang cukup banyak. Ikan air tawar yang ada di Sumatera Selatan antara lain terdiri dari Ikan Patin, Ikan Lele, Ikan Gurami, Ikan Emas, Ikan Nila, Ikan Sepat dan masih banyak jenis lainnya. Produksi Ikan Patin menjadi hasil perikanan air tawar yang paling tinggi di Sumatera Selatan. Salah satu faktor pengaruh jumlah produksi ikan air tawar di Sumatera Selatan yaitu luas area perikanan (Sari, 2017).

Setiap kota/kabupaten memiliki karakteristik yang berbeda-beda dalam sektor perikanan air tawar. Pentingnya mengelompokkan kota/kabupaten berdasarkan hasil produksi ikan air tawar dapat menjadi bahan pertimbangan bagi pemerintah dalam menentukan kebijakan dalam sektor perikanan air tawar tersebut.

Penyebaran hasil produksi ikan air tawar di Sumatera Selatan dapat dibangun menjadi pemetaan. Hasil pemetaan pada sektor perikanan air tawar ini dapat membantu pemerintah provinsi dalam membuat kebijakan, sehingga dapat meningkatkan hasil produksi sektor perikanan air tawar di Sumatera Selatan. Salah satu cara untuk melakukan pemetaan tersebut yaitu dapat dilakukan dengan menggunakan metode analisis cluster/gerombol.

Analisis cluster/gerombol adalah salah satu metode yang dapat digunakan untuk mengelompokkan  $n$  buah objek dalam  $m$  buah kelompok. Tujuan analisis gerombol adalah mengelompokkan beberapa objek berdasarkan kemiripan karakteristiknya. Objek dalam suatu gerombol memiliki kemiripan karakteristik yang tinggi, namun karakteristik antar objek antara satu gerombol dengan lainnya memiliki tingkat kemiripan yang rendah. Terdapat tiga metode umum dalam analisis

gerombol, yaitu metode hierarki, metode non-hierarki dan metode grafik (Mattjik dkk, 2011).

Teknik clustering memiliki penggunaan yang luas dan cenderung semakin meningkat seiring berkembangnya jumlah data. Salah satu teknik pengelompokan sederhana yaitu dengan menggunakan *K-Means*. Algoritma *K-Means* merupakan algoritma pengelompokan yang mengelompokkan data berdasarkan titik pusat kluster terdekat dengan data (Sharma dkk, 2012).

Tujuan dari *K-Means* adalah mengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan data antar kluster. Fungsi jarak adalah ukuran kemiripan yang digunakan dalam kluster (Asroni dkk, 2015).

Pada Algoritma *K-Means* pertama-tama dilakukan pemilihan secara acak  $K$  yang merupakan jumlah banyaknya cluster yang ingin dibentuk. Nilai-nilai  $K$  ditetapkan secara acak,  $K$  tersebut merupakan nilai means atau titik pusat dari cluster. Setiap data akan dihitung jaraknya masing-masing dari titik pusat menggunakan rumus Euclidian sampai mendapatkan jarak yang paling dekat dari setiap data dengan titik pusatnya. Klasifikasikan setiap data berdasarkan kedekatan jaraknya dengan titik pusat. Lakukan langkah-langkah tersebut hingga nilai titik pusatnya telah stabil (Darmi dkk, 2016).

Terdapat tiga metode yang dapat digunakan dalam menentukan banyaknya jumlah cluster ( $k$ ), yaitu metode *Elbow Curve*, *Silhouette*, dan *Gap Statistics*. Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara

jumlah *cluster* yang akan membentuk siku pada suatu titik. Sedangkan metode *Silhouette* digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik suatu objek ditempatkan dalam suatu *cluster*.

Metode *K-Means* telah banyak digunakan dalam berbagai penelitian, diantaranya algoritma *K-Means Cluster Analysis* digunakan dalam studi kasus penyakit menular manusia pada suatu objek. Penelitian tersebut mengkaji metode *K-Means Cluster Analysis* dalam penyakit menular pada manusia berdasarkan set variabel yang dibentuk per kecamatan tiap puskesmas (Bastian, 2018). Selain itu digunakan untuk menemukan pola pemanfaatan internet yang dilakukan oleh pegawai di lingkungan Pusbindiklat Peneliti LIPI dengan pendekatan web usage mining menggunakan algoritme clustering *K-Means* (Sukoco, 2018). Metode *K-Means* juga diterapkan dalam mengelompokkan penjualan produk (Darmi *dkk*, 2016) dan juga digunakan dalam mengelompokkan pelanggan UMS Store menjadi 3 kelompok (Ghofar *dkk*, 2018).

Penelitian ini bertujuan untuk mengelompokkan kota/kabupaten di Sumatera Selatan berdasarkan produksi perikanan air tawar menggunakan metode *K-Means Cluster*. Hasil pemetaan pada sektor perikanan air tawar ini diharapkan dapat membantu pemerintah provinsi dalam membuat kebijakan, sehingga dapat meningkatkan hasil produksi sektor perikanan air tawar pada masing-masing kota/kabupaten di Sumatera Selatan

## BAHAN DAN METODE

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data produksi ikan air tawar di kota/kabupaten Sumatera Selatan tahun 2007-2014 yang dipublikasi oleh Badan

Pusat Statistik (BPS) Sumatera Selatan. Data yang diolah adalah data rata-rata produksi ikan air tawar di Provinsi Sumatera Selatan pada tahun 2007-2014..

Ikan air tawar adalah ikan yang menghabiskan sebagian atau seluruh hidupnya di air tawar, seperti sungai dan danau, dengan salinitas kurang dari 0,05% (Efendi, 2009). Ikan Air tawar diantaranya adalah Ikan Patin, Ikan Lele, Ikan Emas, Ikan Nila dan Ikan Sepat[11].

Adapun kota/kabupaten yang diikutsertakan dalam penelitian hanya 14 kota/kabupaten. Variabel yang digunakan dalam penelitian ini adalah rata-rata produksi ikan air tawar per tahun. Ikan air tawar yang dijadikan variabel yaitu Ikan Patin (X1), Ikan Lele (X2), Ikan Mujair (X3), Ikan Emas (X4), Ikan Gurame (X5), Ikan Nila (X6). Seluruh variabel yang digunakan memiliki satuan Ton.

Data dieksplorasi dengan membuat boxplot untuk melihat karakteristik data rata-rata produksi ikan air tawar. Selanjutnya data dianalisis menggunakan metode *K-Means cluster* dengan bantuan *Software SPSS 21* dan *Software R*.

## HASIL DAN PEMBAHASAN

### Eksplorasi Data

Tabel 1 menyajikan data eksplorasi dari masing-masing jenis ikan air tawar. Produksi ikan air tawar tertinggi di Sumatera Selatan adalah Ikan Nila (X6) dengan rata-rata sebesar 20.049,11 ton/tahun. Sedangkan jenis ikan air tawar dengan produksi terendah yaitu Ikan Mujair (X3), yaitu sebesar 602,9643 ton/tahun.

Karakteristik sebaran masing-masing variabel jenis ikan air tawar terlihat dari boxplot pada Gambar 1. Secara keseluruhan jenis ikan air tawar memiliki karakteristik yang berbeda-beda

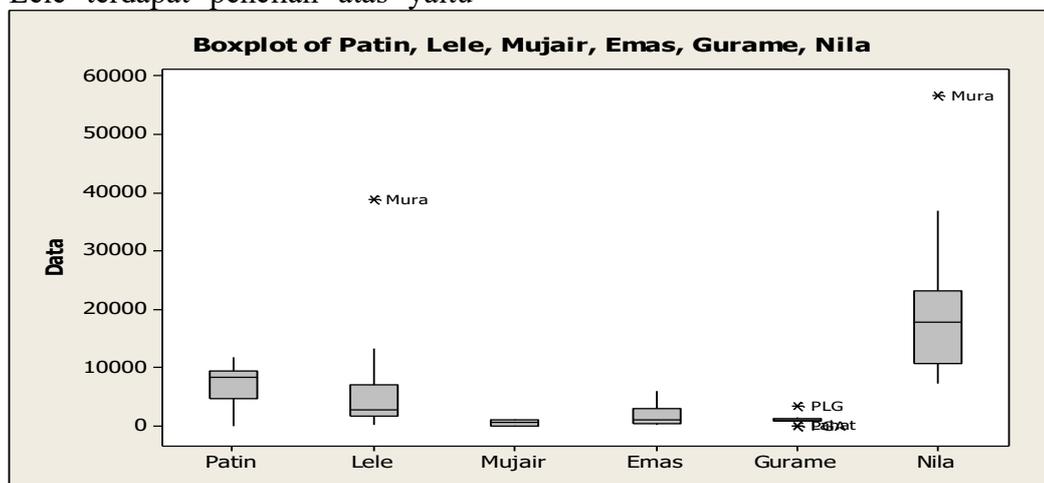
jika dilihat dari range masing-masing variabel.

Tabel 1 Nilai Rerata dan Ragam Ikan Air Tawar

Variabel	Rerata	Ragam
Patin (X1)	7239.98	14869424.9
Lele (X2)	6616.07	102000095.3
Mujair (X3)	602.96	217186.5
Emas (X4)	1968.14	3485833.3
Gurame (X5)	1136.94	612069.4
Nila (X6)	20049.11	171765158.2

Untuk Ikan Mujair dan Ikan Gurame memiliki range yang sangat sempit, artinya produktivitas untuk masing-masing ikan tersebut pada tiap kota/kabupaten tidak terlalu beragam. Pada variabel Ikan Gurame terdapat pencilan bawah yaitu Kabupaten Lahat dan terdapat pula pencilan atas yaitu Kota Palembang. Sedangkan pada variabel Ikan Lele terdapat pencilan atas yaitu

Kabupaten Musi Rawas. Untuk variabel Ikan Nila memiliki range yang sangat besar, hal berarti produktivitas Ikan Nila pada masing-masing kota/kabupaten memiliki keragaman yang besar dan memiliki pencilan atas yaitu Kabupaten Musi Rawas.



Gambar 1. Boxplot Untuk Masing-Masing Jenis Ikan Air Tawar

### Pengelompokan K-Means Cluster

Jumlah  $k$  yang akan dibentuk akan ditentukan berdasarkan kurva yang diperoleh dengan metode *Elbow Curve*, metode *Silhouette*, dan metode *Gap Statistics*.

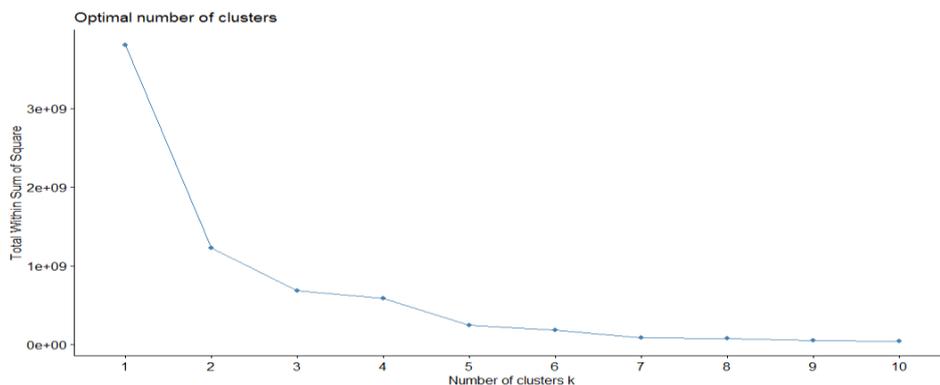
Pada metode *Elbow Curve*, banyaknya  $k$  yang akan dibentuk ditentukan titik  $k$  dimana garis mengalami patahan yang membentuk

*elbow* atau siku. Pada Gambar 2 terlihat garis yang paling mendekati bentuk siku yaitu berada di  $k=2$ . Artinya berdasarkan metode ini, jumlah cluster/gerombol yang terbaik yaitu sebanyak 2 cluster.

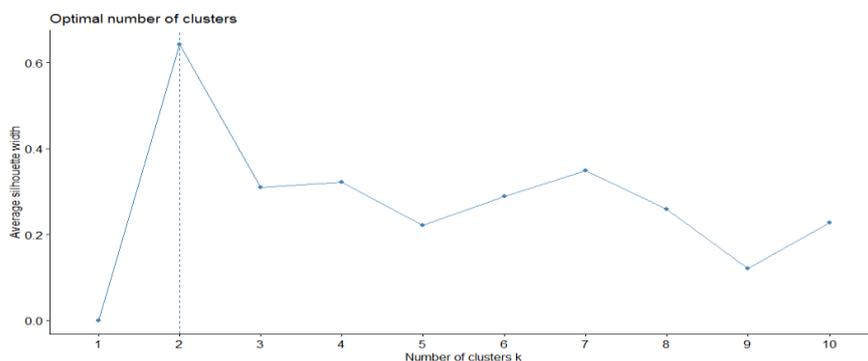
Pendekatan rata-rata nilai metode *silhouette* untuk menduga kualitas dari cluster yang terbentuk. Semakin tinggi nilai rata-rata nya maka akan semakin baik. Pada Gambar 3 terlihat bahwa pada  $k=2$  nilai rata2 silhouette

paling tinggi, dan diikuti oleh k=7 sebagai nilai tertinggi kedua. Maka

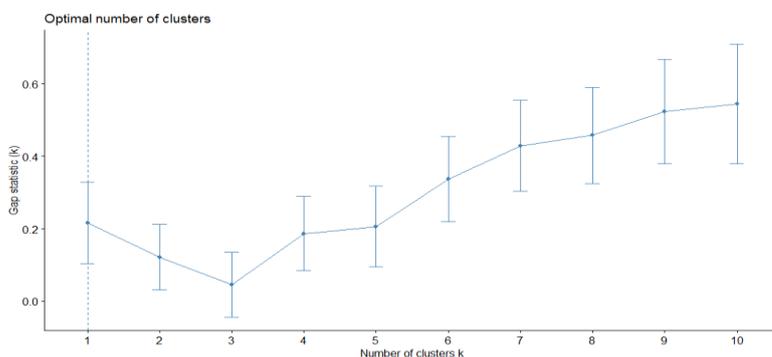
kemungkinan jumlah cluster yang paling baik yaitu sebanyak 2 cluster.



Gambar 2 Output R : Kurva Metode *Elbow*



Gambar 3 Ouput R : Kurva Metode *Silhouette*



Gambar 4 Output R : Kurva Metode *Gaps Statistics*

Hasil kurva menggunakan metode Gap Statistics terlihat pada Gambar 4, terlihat nilai optimal cluster di k=1, tapi menjadi tidak menarik karena nilai k=4 mempunyai nilai yang sama dengan k=1, sehingga tidak terdapat kekonsistenan pada hasil metode ini. Asumsi pada metode ini menjadi tidak menarik yaitu ketika dibentuk sebanyak 4 kelompok

namun rata-ratanya akan sama saja dengan membuat 1 kelompok saja.

Berdasarkan uraian ketiga metode tersebut, maka pada penelitian menggunakan teknik cluster dengan *K-Means Cluster* ini digunakan 2 cluster dalam mengelompokkan kota/kabupaten di Sumatera Selatan berdasarkan variabel produksi ikan air tawar. Jumlah cluster

sebanyak dua yaitu dapat dikelompokkan berdasarkan dua kategori. Cluster 1 merupakan kota/kab. yang memiliki rata-rata produksi ikan air tawar yang rendah, sedangkan pada cluster 2 merupakan kota/kab. yang memiliki rata-rata produksi ikan air tawar yang tinggi.

Berdasarkan hasil pengelompokan akan diketahui anggota cluster manakah yang menjadi rekomendasi kota/kab. manakah yang potensial untuk mendapatkan perlakuan lebih lanjut oleh pemerintah dalam bidang produksi ikan air tawar. Hasil pengelompokan dapat

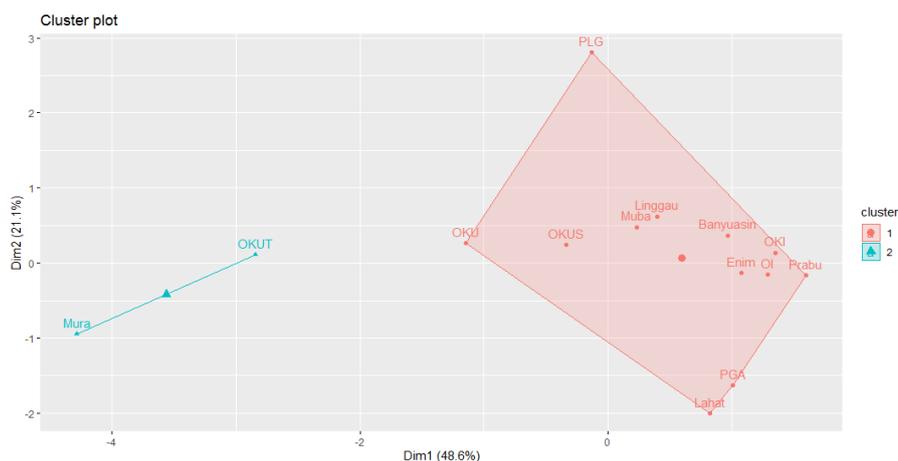
menentukan kota/kabupaten yang ada di Sumatera Selatan yang berpotensi untuk mensuplai ikan air tawar ke kota/kabupaten yang masih masuk dalam kategori rendah. Selain itu dapat pula digunakan sebagai bahan evaluasi untuk dilakukan upaya peningkatkan produksi ikan air tawar pada kota/kabupaten yang masih tergolong dalam kategori rendah tersebut.

Adapun kota/kab. yang menjadi anggota pada masing-masing cluster dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2 Pembagian Kota/Kabupaten Berdasarkan Cluster

Cluster	Kota/Kab. Anggota Cluster
1	Palembang, Linggau, OKU, OKI, Muara Enim, Lahat, Musi Banyuasin, Banyuasin, OKU Selatan, Ogan Ilir, Prabumulih, Pagaralam
2	OKU Timur, Musi Rawas

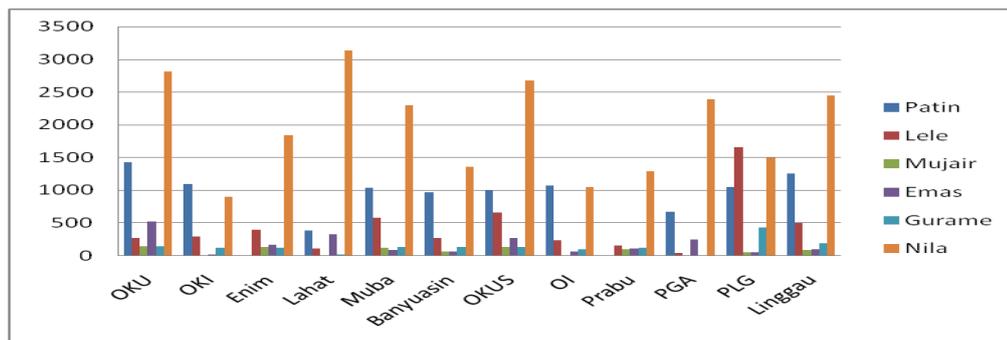
Cluster Patin Lele Mujair Emas Gurame Nila  
 <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
 1 1 6523. 3431. 567. 1374. 1094. 15603.  
 2 2 11540. 25729. 817. 5531. 1394. 46725



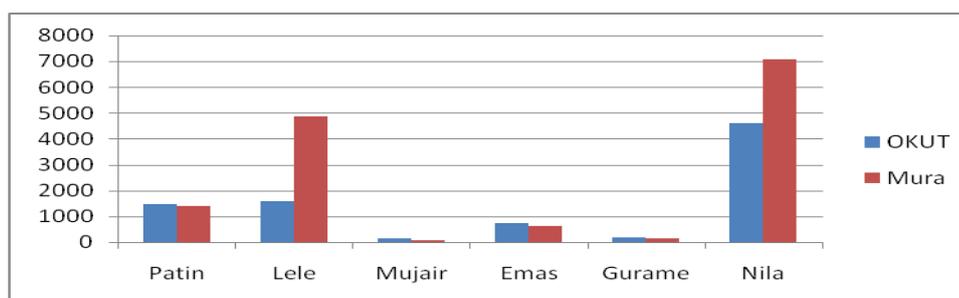
Gambar 5 Output R : Cluster Plot Dengan K=2

Pada Gambar 5 terlihat pemetaan 2 cluster yang terbentuk dengan menggunakan metode K-Means cluster. Berikut nilai mean setiap variabel per cluster.

Gambaran masing-masing cluster dapat dilihat pada Gambar 6 dan Gambar 7. Pada gambar-gambar tersebut terlihat nilai rata-rata produksi ikan air tawar masing-masing kota/kab pada masing-masing cluster.



Gambar 6 Grafik Rata-Rata Produksi Ikan Air Tawar Di Cluster 1



Gambar 7 Grafik Rata-Rata Produksi Ikan Air Tawar Di Cluster 2

Dua belas kota/kabupaten termasuk pada cluster 1 merupakan cluster pada kategori rendah, hal ini menunjukkan bahwa kedua belas kota/kabupaten tersebut masuk kedalam kategori rendah dalam hal rata-rata produksi ikan air tawar.

Kabupaten OKU Timur dan Kabupaten Musi Rawas termasuk pada cluster 2, hal ini menunjukkan bahwa kedua kota tersebut masuk kedalam kategori tinggi dalam rata-rata produksi ikan air tawar.

## KESIMPULAN

Berdasarkan analisis dan pembahasan, maka diperoleh simpulan hasil pengelompokan kota/kabupaten di Provinsi Sumatera Selatan berdasarkan data rata-rata produksi ikan air tawar dengan menggunakan metode *K-Means* Cluster, yaitu : Jumlah cluster yang digunakan adalah 2, hal ini didasarkan pada hasil penentuan jumlah cluster menggunakan metode elbow, silhouette dan gaps statistics.

Adapun dengan diperolehnya dua cluster tersebut, kedua cluster tersebut dapat dikategorikan menjadi kategori tinggi dan rendah dalam hal rata-rata produksi ikan air tawar. Cluster 1 terdiri dari 12 kota/kab., sedangkan cluster 2 terdiri dari 2 kota/kab.

## Ucapan Terima Kasih

Ucapan Terima Kasih penulis sampaikan kepada Universitas PGRI Palembang yang telah memberikan kesempatan bagi penulis untuk mendapatkan Dana Hibah Penelitian Dosen di lingkungan Universitas PGRI Palembang.

## DAFTAR PUSTAKA

Asroni., Ronal, A. (2015). Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Malang, *Jurnal Ilmiah Semesta Teknika*, Vol. 18, No. 1, hal. 76-82.

- Bastian, A. (2018). Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka), *Jurnal Sistem Informasi*, Vol. 14, No.2.
- BPS, Badan Pusat Statistik. (2017). *Sumatera Selatan Dalam Angka 2016*, BPS Sumatera Selatan, Palembang.
- Darmi, Y., Agus, S. (2016). Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk, *Jurnal Media Infotama*, Vol.12, No.2, hal. 148-157
- Efendi, I. (2009). *Pengantar Akuakultur*, Peneber Swadaya, Jakarta
- Ghofar, MA., Kurniawan, YI. (2018). Aplikasi Pengelompokan Pelanggan UMS Store Menggunakan Algoritma K-Means, *Jurnal Teknologi & Manajemen Informatika*, Vol. 4, No.1
- Mattjik, AA., Sumertajaya, IM. (2011). *Sidik Peubah Ganda Dengan Menggunakan SAS*, Institut Pertanian Bogor Press, Bogor
- Sari, Y.P. (2017). Analisis Faktor-Faktor Yang Mempengaruhi Jumlah Produksi Ikan Air Tawar Di Sumatera Selatan, *Jurnal Agr IBA*, Vol. 5, No. 1, hal. 58-62
- Sharma, R., Alam, MA., Rani., Anita. (2012). K-Means Clustering In Spatial Data Mining Using Weka Interpace, *Proceeding Of International Journal Of Computer Application (IJCA)*
- Sukoco, SH. (2018) Analisis Kinerja Pegawai PUSBINDIKLAT Peneliti LIPI Berdasarkan Pola Pemanfaatan Internet Melalui Pendekatan WEB USAGE MINING. Master Thesis, Institut Pertanian Bogor