

## **Metode Boost-K-means untuk Clustering Puskesmas berdasarkan Persentase Bayi yang Diimunisasi**

### *Boost-K-means Method for Clustering Puskesmas based on Percentage of Immunized Infant*

**Ahmad Irfan Abdullah<sup>\*1</sup>, Edi Winarko<sup>2</sup>, Aina Musdholifah<sup>3</sup>**

<sup>1</sup>Program Studi Desain Grafis, Polimedia Kreatif, Makassar

<sup>2,3</sup>Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: <sup>\*</sup>irfan@polimedia.ac.id, <sup>2</sup>ewinarko@ugm.ac.id, <sup>3</sup>aina\_m@ugm.ac.id

#### **ABSTRAK**

DOI;  
10.30595/jrst.v4i2.7546

#### **Histori Artikel:**

Diajukan:  
08/06/2020

Direvisi:  
26/09/2020

Diterima:  
30/09/2020

Dinas Kesehatan Kabupaten/Kota adalah satuan kerja pemerintahan daerah kabupaten/kota yang bertanggung jawab menyelenggarakan urusan pemerintahan dalam bidang kesehatan di kabupaten/kota. Pelayanan kesehatan adalah upaya yang diberikan oleh Puskesmas kepada masyarakat, mencakup perencanaan, pelaksanaan, evaluasi, pencatatan, pelaporan, dan dituangkan dalam suatu sistem. Pada penelitian ini, akan digunakan data persentase bayi yang diimunisasi yang merupakan salah satu layanan dari Puskesmas. Pelayanan imunisasi ini merupakan pelayanan imunisasi dasar meliputi BCG, DPT/HB1-3, polio 1-4 dan campak. Data persentase bayi yang diimunisasi belum memiliki pengelompokan sehingga pada penelitian ini akan diterapkan metode clustering untuk melakukan pengelompokan Puskesmas berdasarkan persentase bayi yang diimunisasi. Data persentase bayi dari masing-masing Puskesmas dijadikan data uji yang akan diterapkan pada proses multi-clustering dengan metode boost-clustering. Output dari penerapan metode ini akan dibandingkan dengan output metode clustering dasar k-means, hasil clustering akan diukur menggunakan metode silhouette index. Evaluasi menggunakan metode silhouette index dilakukan pada dataset puskesmas. Analisis dilakukan dengan melihat hasil evaluasi dataset yang sudah diimplementasikan kedalam algoritma cluster dasar k-means dan algoritma multiclustering boost-k-means. Berdasarkan hasil evaluasi, diperoleh nilai silhouette index 0,798102756 untuk k-means dan 0,789901932 untuk boost-k-means, dengan ini algoritma yang diusulkan memiliki kualitas hasil clustering minimal sama atau lebih baik dari single clustering k-means dengan jumlah iterasi yang lebih sedikit.

**Kata Kunci:** k-means; multi-clustering; boost-clustering; boost-k-means.

#### **ABSTRACT**

District Health Office is a working unit of local government districts/cities are responsible for organizing the affairs of government in the health sector in the district/city. Health services are efforts exerted by the health center to the community, including planning, implementation, evaluation, recording, reporting, and contained in a system. In this study, used percentage of infants immunized data which is one of the services of Puskesmas. Immunization services is a basic immunization services include BCG, DPT/HB1-3, 1-4 polio and campak. Data percentage of infants who were immunized yet have a grouping so that this research will be applied clustering methods to grouping Puskesmas by the percentage of infants immunized. Data percentage of children from each Puskesmas used as test data to be applied to multi-

clustering process with boost-clustering method. Output from the application of these methods will be compared with the output method basic k-means clustering, the clustering results will be measured using methods silhouette index. Silhouette index evaluation methods performed on Puskesmas dataset. The output of clustering have been analyzed by looking at the results of evaluation on dataset, that has been implemented into the single cluster algorithm k-means and multi-clustering algorithm boost-k-means. Based on the evaluation results, the value of silhouette index is 0,798102756 for k-means and 0,789901932 for boost-k-means, the proposed algorithm has the quality of the clustering of at least the same or better than single k-means clustering with fewer number of iterations.

**Keywords:** k-means; multi-clustering; boost-clustering; boost-k-means.

## 1. PENDAHULUAN

Dinas Kesehatan Kabupaten/Kota adalah satuan kerja pemerintahan daerah kabupaten/kota yang bertanggung jawab menyelenggarakan urusan pemerintahan dalam bidang kesehatan di kabupaten/kota. Fasilitas pelayanan kesehatan adalah suatu tempat yang digunakan untuk menyelenggarakan upaya pelayanan kesehatan, baik promotif, preventif, kuratif maupun rehabilitatif yang dilakukan oleh pemerintah, pemerintah daerah dan/atau masyarakat. Pada data persentase bayi yang diimunisasi belum memiliki pengelompokan persentase layanan, sehingga pada penelitian ini akan digunakan sebagai data uji dalam penerapan metode *multi-clustering*, *boost-clustering* dengan algoritma *clustering* dasar *k-means*.

Diharapkan dari hasil *clustering* diperoleh pengelompokan Puskesmas yang bisa memberikan informasi mengenai Puskesmas yang sudah memenuhi target layanan imunisasi bayi, belum memenuhi target layanan imunisasi bayi dan melebihi target layanan imunisasi bayi.

Pendekatan *multi-clustering* dengan metode *boost-clustering* (Frossyniotis, D., dkk., 2004) , merupakan metode *multi-clustering* dengan tujuan menghasilkan peningkatan kualitas dan ketahanan kinerja proses *clustering*. Metode ini diharapkan dapat meningkatkan hasil *clustering* dasar *k-means* yang akan digunakan pada penelitian ini, yang akan di ukur menggunakan silhouette index.

Metode pada *clustering* telah dikembangkan oleh banyak peneliti dengan menggunakan berbagai metode untuk berbagai kasus. Berikut beberapa diantaranya, (Wang, R., dkk., 2011) penelitian ini menggunakan metode *K-means* dalam melakukan pengelompokan sumber daya perpustakaan untuk meningkatkan layanan perpustakaan sesuai kebutuhan pengguna. Hasil dari masing-masing metode yang

digunakan sudah dapat menyelesaikan permasalahan tetapi belum menyentuh pada tahap peningkatan hasil *clustering*. Pada penelitian yang akan penulis lakukan, diharapkan ada peningkatan terhadap hasil *clustering k-means* ketika dijalankan dalam proses *multi-clustering*.

Penelitian berikutnya (Ahalya, G. & Pandey, H.M., 2015) menekankan pada pemahaman beberapa metode *clustering* untuk memudahkan para pemula memahami cara kerja dari algoritma *clustering*. Dalam penelitian ini dibahas 5 metode *clustering*, yaitu algoritma *k-means*, algoritma *clustering* hirarki, algoritma *clustering* density based, algoritma self-organizing map (SOM) dan algoritma expectation maximization (EM). Pada penelitian ini algoritma *clustering* dasar seperti yang telah disebutkan akan dijalankan masing-masing, sedangkan pada penelitian yang akan penulis lakukan algoritma *clustering* dasar akan dijalankan didalam proses *multi-clustering*.

Penelitian berikut (Elssied, N.O.F., dkk., 2014) mengenai deteksi spam pada email, menggunakan metode support vector machine (SVM) dan *k-means*. Hasil yang diperoleh pada penelitian ini, yaitu dengan melakukan hybrid system menerapkan metode SVM dengan *k-means* dapat meningkatkan akurasi menjadi 98,01% dari yang tadinya 96,30% ketika hanya menggunakan metode SVM.

Penelitian selanjutnya bertema *multi-clustering*, penelitian ini (Frossyniotis, D., dkk., 2004) menggunakan metode *boosting* untuk melakukan *clustering*. Algoritma *clustering* dasar dibutuhkan pada proses *boost-clustering* , misalnya menggunakan *k-means*. Solusi *clustering* akhir dihasilkan dengan menggabungkan beberapa hasil *clustering* melalui voting pembobotan sesuai distribusi yang telah dilakukan sebelumnya. Percobaan pada kedua dataset buatan dan dunia nyata menunjukkan

bahwa solusi *boost-clustering* memberikan peningkatan kualitas terhadap hasil *clustering*. Penelitian berikut (Li C.A., 2009) , melakukan penelitian yang merupakan variant dari algortima *boosting*, proses *boosting* yang dilakukan, sama dengan yang dilakukan oleh penelitian (Frossyniotis, D., dkk., 2004) . Perbedaannya yaitu, penelitian ini distribusi data dilakukan menurut atribut yang telah melalui proses feature selection, atribut yang dianggap penting akan dihitung untuk menghasilkan nilai untuk update distribusi atribut, sehingga pada akhirnya terbentuk terbentuk cluster pada dataset yang telah diberi pembobotan.

Penelitian selanjutnya (Rashedi, E. & Mirzaei, A., 2011) , memperkenalkan metode *multi-clustering boosting* untuk mengetahui tipe *cluster* yang baik digunakan pada metode *boosting*, hasilnya tipe cluster centroid dan average yang lebih baik untuk digunakan. Sedangkan pada penelitian berikut (Li, C & Wu T., 2006) , telah dibuktikan bahwa *clustering* menggunakan metode *boosting* dapat menangani data yang sulit untuk dikelompokkan pada data yang heterogen dan dapat meningkatkan akurasi *clustering* untuk data yang terpusat.

Pada penelitian berikut akan dilakukan dengan menggunakan dua metode, penelitian (Lisangan, E.A., Musdholifah, A. & Hartati S., 2015) menggunakan metode *Fuzzy Subtractive Clustering* (FSC) dan *Self-Organizing Map* (SOM). Menggunakan pengujian yang salah satunya adalah *silhouette index*, penelitian ini menunjukkan bahwa FSC-SOM dapat memperbaiki pusat cluster yang dihasilkan oleh FSC dengan SOM sehingga dapat diperoleh kualitas eksternal dan internal hasil clustering yang lebih baik.

## 2. METODE PENELITIAN

### 2.1 K-means

Dalam statistik dan mesin pembelajaran, *clustering k-means* merupakan metode analisis *cluster* yang mengarah pada pemartisian N objek pengamatan ke dalam K *cluster* (*cluster*) di mana setiap objek pengamatan dimiliki oleh sebuah *cluster* dengan *mean* (rata-rata) terdekat, mirip dengan algoritma *Expectation-Maximization* untuk *Gaussian Mixture* di mana keduanya mencoba untuk menemukan pusat dari *cluster* dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma.

*K-means* merupakan salah satu metode *clustering* data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih *cluster*. Metode ini mempartisi data yang ada ke dalam bentuk dua atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data berkarakteristik sama dimasukkan kedalam satu *cluster* yang sama dan data yang berkarakteristik yang berbeda dikelompokkan dalam *cluster* yang lain. Adapun tujuan *clustering* data ini adalah untuk meminimalkan fungsi objektif yang di set dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*.

*Clustering* data dengan metode *k-means* ini secara umum dilakukan dengan algoritma seperti berikut (Prasetyo, E., 2012) :

1. Tentukan jumlah *cluster*.
2. Alokasikan data ke dalam *cluster* secara acak.
3. Hitung pusat kelompok (*centroid*/rata-rata) dari data yang ada di masing-masing *cluster*.
4. Alokasikan masing-masing data ke *centroid*/rata-rata terdekat
5. Kembali ke langkah 3, apabila masih ada data yang berpindah *cluster* , atau apabila ada perubahan nilai *centroid* di atas nilai ambang yang di tentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

Persamaan (1) dilakukan sebanyak  $p$  dimensi sehingga  $i$  mulai dari 1 sampai  $p$ . Pada penelitian ini peneliti akan menggunakan persamaan *Euclidian* untuk mengukur jarak data ke pusat *cluster* sehingga didapatkan jarak terpendek antara dua titik.

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \quad (1)$$

Pada langkah ke 4 menggunakan persamaan (2), pengalokasian kembali data ke dalam masing-masing *cluster* dalam metode *k-means* didasarkan pada perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada. Data dialokasikan ulang secara tegas ke *cluster* yang mempunyai *centroid* dengan jarak terdekat dari data tersebut. Pengalokasian ini dapat dirumuskan sebagai berikut :

$$a_{i1} = \begin{cases} 1 & d = \min\{D(x_i, C_1)\} \\ 0 & \text{lainnya} \end{cases} \quad (2)$$

Apabila masih ditemukan data yang berpindah *cluster*, maka digunakan persamaan (3) :

$$J = \sum_{i=1}^N \sum_{j=1}^K a_{ij} D(x_i, C_j)^2 \quad (3)$$

Dimana  $N$  adalah jumlah data,  $K$  adalah jumlah *cluster*,  $a_{ij}$  adalah nilai keanggotaan titik data  $x_i$  ke pusat *cluster*  $C_j$ ,  $C_j$  adalah pusat *cluster* ke- $j$ , dan  $D(x_i, C_j)$  adalah jarak titik  $x_i$  ke *cluster*  $C_j$  yang diikuti.  $a$  mempunyai nilai 0 atau 1. Apabila suatu data merupakan anggota suatu *cluster*, nilai  $a_{ij} = 1$ . Jika tidak, nilai  $a_{ij} = 0$ .

Proses *multi-clustering* melakukan pembentukan *cluster* pada *dataset* dengan menggunakan metode *cluster* dasar yang dijalankan pada metode *multi-clustering* dalam satu proses. Sehingga dalam pembentukan *cluster* ada perbedaan pada komponen-komponen input dan metode yang digunakan untuk menghasilkan *cluster* pada setiap iterasi. Proses *multi-clustering* terdiri dari algoritma *multi-clustering* dan sebuah algoritma *cluster* dasar, pada penelitian ini digunakan algoritma *multi-clustering boosting* dan algoritma *cluster* dasar *k-means*.

## 2.2 Boost-clustering

*Multi-clustering* merupakan teknik untuk menggabungkan beberapa algoritma *clustering* dasar untuk menghasilkan sebuah data *cluster* (Ashlock, D. dan Guo, L., 2007). Algoritma yang dapat digunakan untuk *multi-clustering* salah satunya dengan menggunakan *boost-clustering* (Frossyniotis, D. dkk., 2004).

## 2.3 Algoritma Boost-clustering

Algoritma *boost-clustering* merupakan pendekatan *multi-clustering* yang secara iteratif mendaur ulang contoh pelatihan, menyediakan beberapa *clustering* dan menghasilkan partisi umum.

Diberikan data input yang berurutan seperti contoh  $N$  berikut,  $(x_1, \dots, x_N), (x_i \in R^d), i = 1, \dots, N$ , sebuah algoritma dasar *clustering*, jumlah  $C$  *cluster* untuk partisi data set dan jumlah maksimum iterasi  $T$ . Pada algoritma *boost-clustering* (Frossyniotis, D., dkk., 2004) *clustering* akan dilakukan dengan langkah-langkah seperti berikut :

1. Inisialisasi distribusi bobot awal  $w_i^1 = \frac{1}{N}$  untuk  $i = 1, \dots, N$ . Tetapkan,  $\epsilon_{max} = 0$  dan  $ic = 0$

Dimana :  $N$  = Jumlah Data

$\epsilon_{max}$  = Nilai *pseudoloss* awal

$ic$  = iterasi untuk *stopping criteria pseudoloss*

2. Pada kondisi iterasi  $t \leq T$ 
  - a. Lakukan proses *clustering* dengan algoritma *clustering* dasar, untuk mempartisi *dataset* sehingga didapatkan partisi dari  $H^t$  yang merupakan hasil dari proses pembentukan *cluster* dasar.  
Dimana :  $t$  = jumlah iterasi  
 $T$  = iterasi maksimum

- b. Pada tahapan ini *cluster hypothesis* awal  $H_i^t$  telah diperoleh,  $H_i^t = (h_{i,1}^t, h_{i,2}^t, h_{i,3}^t, \dots, h_{i,C}^t)$  untuk semua  $i, i = 1, \dots, N$ , dimana  $h_{i,j}$  merupakan derajat keanggotaan dari data  $i$  terhadap *cluster*. Simbol  $h_{i,j}$  merupakan hasil perbandingan nilai jarak dari *cluster* dasar pada masing-masing data dan *cluster*. Persamaan (4) berikut digunakan untuk memperoleh nilai  $h_{i,j}$  :

$$h_{i,j} = \frac{1}{\sum_{k=1}^C \frac{d(x_i, \mu_k)}{d(x_i, \mu_j)}} \quad (4)$$

Dimana :

$d(x_i, \mu_j)$  &  $d(x_i, \mu_k)$  = Nilai *eucliden distance* hasil proses *k-means*.

- c. Jika iterasi sudah lebih dari satu kali ( $t > 1$ ), lakukan penggantian anggota *cluster*  $H^t$  berdasarkan hasil *cluster aggregate hypothesis*  $H_{ag}^{t-1}$ .  $H_{ag}^{t-1}$  merupakan hasil pembentukan *cluster hypothesis aggregate* iterasi sebelumnya.
  - d. Proses selanjutnya hitung *Pseudoloss* dengan persamaan (5)

$$\epsilon_t = \frac{1}{2} \sum_{i=1}^N w_i^t CQ_i^t \quad (5)$$

sebelum persamaan (5) terlebih dahulu dilakukan perhitungan untuk mendapatkan nilai kulit *clustering CQ* dengan persamaan (6)

Dimana :

$w_i^t$  = Nilai bobot masing-masing data

$$CQ_i^t = - \sum_{j=1}^c h_{i,j}^t \log(h_{i,j}^t) \quad (6)$$

Dimana :

$CQ_i^t$  = Nilai *cluster quality* masing-masing data

$h_{i,j}^t$  = Nilai *cluster hypothesis* awal

- e. Tetapkan nilai  $\beta_t$  dengan persamaan (7),

$$\beta_t = \frac{1 - \epsilon_t}{\epsilon_t} \quad (7)$$

Hasil dari perhitungan ini akan digunakan pada proses *update* bobot.

- f. Diberikan kondisi seperti berikut :

- (i) Jika  $\epsilon_t > 0.5$  maka,

$$T = t - 1$$

lakukan langkah ke 3  
(iterasi *boost-clustering*  
berhenti)

- (ii) Jika  $\epsilon_t > \epsilon_{max}$  maka

$$ic = ic + 1$$

Jika  $ic = 3$  maka  $T = t$   
lakukan langkah ke 3  
(iterasi *boost-clustering*  
berhenti)

Jika tidak

$$ic = 0$$

$$\epsilon_{max} = \epsilon_t$$

$ic$  merupakan iterasi dengan

kondisi  $\epsilon_t > \epsilon_{max}$

- g. Proses berikutnya lakukan *update* distribusi bobot dari  $W$  dengan menggunakan persamaan (8):

$$w_i^{t+1} = \frac{w_i^t \beta_t^{CQ_i^t}}{Z_t} \quad (8)$$

Pada persamaan ini,  $Z_t$  merupakan normalisasi yang konstan yang akan selalu menghasilkan nilai  $w_i^{t+1} = 1$ .

Dimana :

$w_i^{t+1}$  = *update* distribusi bobot dengan jumlah iterasi lebih dari 1 kali

- h. Menggunakan persamaan (9) hitung *aggregate* dari *cluster hypothesis* awal :

$$H_{ag}^t = \arg \max_{k=1, \dots, C} \sum_{\tau=1}^t \left[ \frac{\log(\beta_\tau)}{\sum_{j=1}^t \log(\beta_j)} h_{i,k}^\tau \right] \quad (9)$$

- i. Nilai iterasi, sesuai dengan kondisi  $t := t + 1$ .

Dimana :

$h_{i,k}^\tau$  = Nilai *cluster hypothesis* awal

3. *Output* pada jumlah iterasi  $T$  (akhir) akan menjadi *cluster final*  $H^f$  yang sebenarnya merupakan nilai dari *cluster aggregate hypothesis* atau  $H^f = H_{ag}^t$ .

Dimana :

$H^f$  = *Cluster aggregate hypothesis*  $H_{ag}^t$  pada iterasi terakhir (maksimum).

## 2.4 Silhouette Index

Pengujian adalah faktor penting untuk mengetahui validitas dari hasil penelitian yang sudah dilakukan. Hasil pengujian akan memberikan informasi jumlah *cluster* yang baik untuk *dataset* yang telah melalui proses *clustering*. Untuk menghitung nilai *silhouette index* dari sebuah data ke- $i$  ada 2 komponen yaitu  $a_i$  dan  $b_i$ .  $a_i$  adalah rata-rata jarak data ke- $i$  terhadap semua data lainnya dalam satu *cluster*, sedangkan  $b_i$  didapatkan dengan menghitung rata-rata jarak data ke- $i$  terhadap semua data dari *cluster* yang lain tidak dalam satu *cluster* dengan data ke- $i$ , kemudian diambil yang terkecil (Tan, Pang-Ning., dkk. 2006).

Nilai *silhouette index* dari data ke- $i$  ( $S(i)$ ) didapatkan dengan perhitungan dengan menggunakan persamaan (10).

$$S(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))} \quad (10)$$

Pada persamaan (10),  $a(i)$  merupakan jarak rata-rata data ke- $i$  terhadap data yang terdapat dalam *cluster* yang sama sedangkan  $b(i)$  merupakan jarak rata-rata terkecil data ke- $i$  terhadap data yang terdapat pada masing-masing *cluster* yang berbeda.

$$a(i) = \frac{\sum d(i, j)}{n_{Ci}}, i, j \in C_i \quad (11)$$

$$b(i) = \min_{C_k \neq C_i} \left\{ \frac{\sum dist(i, j)}{n_{Ck}} \right\}, i \in C_i \text{ dan } k \in C_k \quad (12)$$

Nilai  $a(i)$  diperoleh dengan menggunakan persamaan (11) dan untuk nilai  $b(i)$  digunakan persamaan (12),  $dist(i, j)$  merupakan jarak antara titik ke- $i$  terhadap titik ke- $j$ ,  $n_{Ci}$  merupakan jumlah titik yang ada dalam *cluster* ke- $i$  sedangkan  $n_{Ck}$  merupakan jumlah titik yang ada dalam *cluster* ke- $k$ .



Nilai rata-rata *silhouette index* dihitung menggunakan persamaan (13), dengan menjumlahkan keseluruhan nilai *silhouette index* untuk setiap data.  $S(i)$  selanjutnya dibagi dengan jumlah data keseluruhan ( $n$ ).

$$S = \frac{1}{n} \sum_{i=1}^n S(i) \quad (13)$$

Nilai  $a_i$  akan menjadi tolak ukur tingkat kemiripan sebuah data dengan *cluster* yang diikutinya, nilai yang semakin kecil menandakan tingkat ketepatan data tersebut berada pada sebuah *cluster*. Nilai  $b_i$  yang besar menandakan tingkat keburukan data terhadap *cluster* lain (Prasetyo, E., 2012). *Silhouette index* memiliki rentang nilai  $[-1, +1]$ . Nilai *silhouette index* yang mendekati 1 menandakan tingkat ketepatan data tersebut berada pada sebuah *cluster*.

### 3. HASIL DAN PEMBAHASAN

Pengujian hasil *clustering* dengan metode *silhouette index* dilakukan dengan rentang jumlah *cluster* 2 (dua) sampai 10 (sepuluh) pada *dataset* puskesmas. Pada *dataset* puskesmas akan diimplementasikan kedalam model aplikasi dan dijalankan masing-masing sebanyak 3 (tiga) kali, kemudian akan diamati hasil pengujian terbaik dari proses yang sudah dilakukan. Pada Tabel 1 akan ditampilkan daftar hasil pengujian terbaik, dari *silhouette index* untuk kedua metode yang digunakan pada *dataset* puskesmas sebagai data uji yang telah di evaluasi.

**Tabel 1.** Daftar hasil evaluasi *silhouette index* terbaik

Nama Dataset	Jumlah Cluster / Class	Jumlah Data	K-means	Boost-K-means
Puskesmas	2	379	0,798102756	0,789901932
	3		0,548522359	0,604693087
	4		0,538086698	0,324902421
	5		0,431282336	0,411477976
	6		0,42423929	0,304520716
	7		0,395997746	0,330222033
	8		0,354574449	0,251810304
	9		0,329903844	0,183443325
	10		0,349047852	0,274783902

Hasil clustering pada *dataset* puskesmas kemudian diuji menggunakan *silhouette index*, pada *dataset* puskesmas diperoleh hasil pengujian terbaik pada jumlah *cluster* 2 untuk kedua algoritma yang digunakan, dengan nilai rata-rata 0,798102756 untuk *clustering* dengan *k-means* dan 0,789901932 untuk *clustering boost-k-means*.

Pada *dataset* puskesmas untuk jumlah *cluster* 3 nilai yang lebih besar diperoleh pada metode *boost-k-means* dibandingkan *k-means*. Perbandingan jarak rata-rata intracluster dengan jarak rata-rata intercluster, menggunakan alat ukur *silhouette index* untuk mengukur kualitas hasil *clustering* menunjukkan bahwa, pada *dataset* puskesmas hasil *clustering* dari metode *boost-k-means* dan *k-means* memiliki kualitas minimal sama.

Pada Tabel 2 ditunjukkan nilai rata-rata evaluasi *silhouette index k-means* yang dihitung berdasarkan hasil percobaan sebelumnya yang dilakukan masing-masing sebanyak 3 kali untuk setiap jumlah *cluster*.

**Tabel 2.** Rata-rata evaluasi *silhouette index k-means*

Nama Dataset	Jumlah Cluster / Class	Jumlah Data	Rata – rata Sihouette Index	
			K-means	
Puskesmas	2	379	0,798102756	± 0
	3		0,458437654	± 0,084262217
	4		0,517049787	± 0,036436999
	5		0,420574668	± 0,009273113
	6		0,423568512	± 0,001161821
	7		0,357605228	± 0,033716458
	8		0,318945046	± 0,048430895
	9		0,271917243	± 0,051937356
	10		0,270924631	± 0,091413321

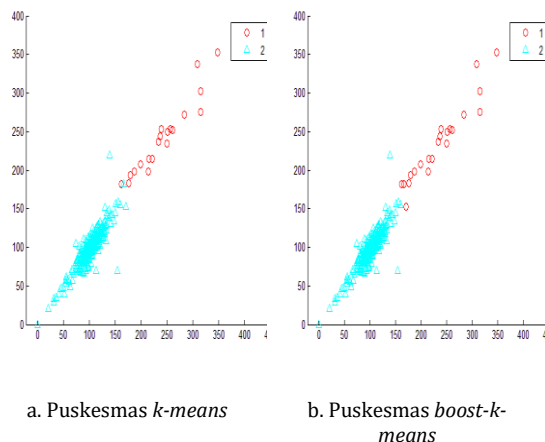
Tabel 3 ditunjukkan nilai rata-rata evaluasi *silhouette index boost-k-means* yang dihitung berdasarkan hasil percobaan sebelumnya yang dilakukan masing-masing sebanyak 3 kali untuk setiap jumlah *cluster*.

**Tabel 3.** Rata-rata evaluasi *silhouette index boost-k-means*

Nama Dataset	Jumlah Cluster/ Class	Jumlah Data	Rata – rata Silhouette Index	
			Boost-K-means	
Puskesmas	2	379	0,224769957 ±	0,584610864
	3		0,120820436 ±	0,471813181
	4		0,016064497 ±	0,314932836
	5		0,094360467 ±	0,334757734
	6		0,054818794 ±	0,247125716
	7		0,068949479 ±	0,272730902
	8		0,034349551 ±	0,223164993
	9		0,014251309 ±	0,16880279
	10		0,034860031 ±	0,243963218

Hasil clustering dari 3 kali percobaan pada *dataset* puskesmas dan masing-masing jumlah *cluster* memiliki perbedaan dalam jumlah anggota dari masing-masing *cluster* sehingga hasil pengujian yang diperoleh juga berbeda-beda, seperti yang ditampilkan pada Tabel 2 dan Tabel 3. Keberagaman hasil clustering paling minimal = 0 diperoleh pada metode *k-means* dengan jumlah *cluster* 2, sedangkan keberagaman hasil *clustering* paling minimal pada metode *boost-k-means* dengan nilai satandar deviasi = 0,014251309 pada jumlah *cluster* 9.

Hasil proses *clustering* juga dapat dilihat dalam bentuk visualisasi grafik *scatter*, pada Gambar 1 akan ditampilkan visualisasi sebaran masing-masing *dataset* dengan hasil pengujian *silhouette index* terbaik.

**Gambar 1.** Visualisasi sebaran *dataset* hasil *clustering*

Visualisasi yang ditunjukkan oleh Gambar 1a dan Gambar 1b merupakan sebaran *dataset* puskesmas dengan jumlah *cluster* 2. Gambar 1a merupakan sebaran *dataset* puskesmas hasil *clustering* dengan *k-means* dan Gambar 1b sebaran *dataset* puskesmas hasil *clustering boost-k-means*. Pada *dataset* puskesmas nilai kerapatan

*intracluster* lebih baik pada *boost-k-means* dibanding *k-means* sedangkan nilai kerenggangan *intercluster* lebih baik pada *k-means* dibanding *boost-k-means*.

#### 4. KESIMPULAN

Setelah melakukan serangkaian penelitian mulai dari pengumpulan data, perancangan, implementasi dengan metode *single* dan *multi-clustering* serta melakukan pengujian pada hasil *clustering* dengan dua alat ukur yang berbeda masing-masing *dataset*, maka dapat disimpulkan bahwa, metode *multi-clustering boosting* yang menjalankan metode *cluster* dasar *k-means (boost-k-means)* untuk *dataset* puskesmas memiliki kualitas hasil *cluster* terbaik yang sama dengan hasil *clustering* oleh metode *single clustering k-means*, yaitu pada jumlah *cluster* 2 (dua), dengan jumlah iterasi yang lebih sedikit pada metode *multi-clustering boost-k-means*.

#### UCAPAN TERIMA KASIH

Terima kasih kepada Dinas Kesehatan Provinsi Sulawesi Selatan yang telah bersedia membantu terkait data yang dijadikan objek pada penelitian ini.

#### DAFTAR PUSTAKA

- Frossyniotis, D., Likas, A. & Stafylopatis, A. (2004). A clustering method based on boosting, *Pattern Recognition Letters*, 25, Elsevier B.V., pp.641-654.
- Wang, R., dkk. (2011). K-means Clustering Algorithm Application in University Libraries, *Cognitive IEEE Informatics & Cognitive Computing (ICCI\*CC)*, pp.419-422.
- Ahalya, G. & Pandey, H.M. (2015). Data Clustering Approaches Survey and Analysis, *IEEE Futuristic Trends on Computational Analysis and Knowledge Management ABLAZE*, pp.532-537.
- Elssied, N.O.F., Ibrahim, O. & Osman, A.H. (2014). Enhancement of spam detection mechanism based on hybrid k-mean clustering and support vector machine, *Journal of Soft Computing*, vol.19, pp 3237-3248.
- Li C.A. (2009). A new clustering algorithm using attribute boosting, *IEEE International*

- Conference on Management and Service Science (MASS '09).*
- Rashedi, E. & Mirzaei, A. (2011). A Novel Multi-clustering Method for Hierarchical Clusterings Based on Boosting, *IEEE 19th Iranian Conference on Electrical Engineering*.
- Li, C & Wu T. (2006). A Boosted Clustering Algorithm for Distributed Homogeneous Data Mining, *IEEE The 6th World Congress on Intelligent Control and Automation*.
- Lisangan, E.A., Musdholifah, A. & Hartati S. (2015). Two Level Clustering for Quality Improvement using Fuzzy Subtractive Clustering and Self-Organizing Map, *TELKOMNIKA Indonesian Journal of Electrical Engineering*, II(15), pp.373-380.
- Prasetyo, E. (2012). *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB, ANDI*, Yogyakarta.
- Ashlock, D. dan Guo, L. (2007). Evolutionary Parameter Setting of Multi-clustering, *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*.
- Tan, Pang-Ning., Steinbach M. & Kumar V. (2006). *Introduction to Data Mining*, Pearson Inc., New York.