p-ISSN 2087-1716
e-ISSN 2548-7779
ILKOM Jurnal Ilmiah Volume 10 Nomor 3 Desember 2018

OPTIMASI *K-MEANS CLUSTERING* UNTUK IDENTIFIKASI DAERAH ENDEMIK PENYAKIT MENULAR DENGAN *ALGORITMA PARTICLE SWARM OPTIMIZATION* DI KOTA SEMARANG

Suhardi Rustam[1], Heru Agus Santoso [2], Catur Supriyanto [3]
[1]suhardirstm@gmail.com, [2]herezadi@gmail.com, [3]catur.dinus@gmail.com
[1]Universitas Ichsan Gorontalo,[2,3]Universitas Dian Nuswantoro

Abstrak

Wilayah tropik merupakan kawasan endemik berbagai penyakit menular. Sekaligus merupakan kawasan yang berpotensi tinggi untuk hadirnya penyakit infeksi baru. Penyakit menular masih merupakan masalah utama kesehatan masyarakat Indonesia. Indentifikasi daerah endemik penyakit menular merupakan persoalan penting dibidang kesehatan, rata-rata tingkat penderita cacat fisik dan kematian bersumber dari penyakit menular. Data Mining dalam perkembangannya menjadi salah satu tren utama dalam pengolahan data. Data Mining secara efektif bisa mengidentifikasi wilayah endemik dari hubunngan antar variable. Algoritma k-means klustering digunakan dengan tujuan mengelompokkan daerah endemik sehingga pengidentifikasian endemik penyakit menular bisa tercapai dengan tingkat validasi yang maksimal dalam klastering. Penggunaan optimasi dalam mengidentifikasi daerah endemik penyakit menular menggabungkan algoritma k-means clustering dengan optimasi particle swarm optimization (PSO). hasil eksperimen endemik untuk algoritma k-means dengan iterasi =10, K-Fold =2 memiliki Indek davies bauldin = 0.169 dan algoritma k-means dengan PSO, iterasi = 10, K-Fold = 5, memiliki index davies bouldin = 0.113. k-fold = 5 memiliki performance yang lebih baik.
Kata Kunci*:* Endemik Penyakit menular, Data Mining, Klustering, K-Means, Particle Swarm Optimization

Abstract

Tropical regions is a region endemic to various infectious diseases. At the same time an area of high potential for the presence of infectious diseases. Infectious diseases still a major public health problem in Indonesia. Identification of endemic areas of infectious diseases is an important issue in the field of health, the average level of patients with physical disabilities and death are sourced from infectious diseases. Data Mining in its development into one of the main trends in the processing of the data. Data Mining could effectively identify the endemic regions of hubunngan between variables. K-means algorithm klustering used to classify the endemic areas so that the identification of endemic infectious diseases can be achieved with the level of validation that the maximum in the clustering. The use of optimization to identify the endemic areas of infectious diseases combines k-means clustering algorithm with optimization particle swarm optimization (PSO). the results of the experiment are endemic to the k-means algorithm with iteration =10, the K-Fold =2 has Index davies bauldin = 0.169 and k-means algorithm with PSO, iteration = 10, the K-Fold = 5, index davies bouldin = 0.113. k-fold = 5 has better performance.
Keywords: Endemic infectious Disease, Data Mining, Clustring, K-Means, Particle Swarm Optimization

1. Pendahuluan

wilayah tropik dan wilayah dinamik secara sosial ekonomi, merupakan kawasan endemik berbagai penyakit menular. Sekaligus merupakan kawasan yang berpotensi tinggi untuk hadirnya penyakit infeksi baru. Penyakit menular masih merupakan masalah utama kesehatan masyarakat Indonesia. Penyakit menular tidak mengenal batas-batas daerah administratif, misalnya antar propinsi, Kabupaten/kota bahkan antar negara. menghilangkan sumber penyakit dengan cara menemukan dan mencari kasus secara proaktif dan mengurangi Intervensi faktor resiko, misalnya lingkungan dan intervensi terhadap perilaku. Endemik ditandai dengan adanya kondisi/keadaan dimana penyakit secara menetap berada dalam masyarakat pada suatu tempat/populasi tertentu, suatu infeksi penyakit disebut sebagai endemik jika kondisi/keadaan setiap orang yang terinfeksi penyakit menularkannya kepada setiap orang dalam kondisi rata-rata[1]. Variabel-variabel endemik secara umum seperti Nama Penyakit, Agen/Pembawa, tubuh pasien, kondisi lingkungan dan kepadatan penduduk.

Beberapa penyakit menular, menjadi perhatian pemerintah untuk ditangani secara komputasi. Teknologi komputasi yang telah berkembang dengan kemampuan yang dimiliki memberikan beberapa solusi untuk menganalisis permasalahan-permasalahan yang muncul dalam melihat kesehatan masyarakat. Teknologi komputasi mampu memberikan analisa pandangan dengan bentuk persentase angka dengan perhitungan tersendiri berdasarkan metode yang digunakan. Teknologi komputasi

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259

memberikan cara menganalisa akurasi terhadap jenis data, pengelompokkan data, mengatur variable dan pengaruh terhadap setiap variable. Terdapatnya daerah penyakit menular dituntut untuk mengelompokkan penyakit menular berdasarkan daerah endemik.

Secara komputasi, dalam pengolahan data komputasi memenggunakan beberapa metode seperti klustering, Analisis kluster merupakan suatu teknik klasifikasi tanpa pengawasan, menurut kriteria kesamaan tertentu untuk mengklasifikasikan dataset, sehingga objek dari kelas yang mungkin sama, tapi itu adalah keragaman mungkin antara objek yang berbeda, k-means adalah algoritma klustering klasik berdasarkan partisi, algoritma k-means memiliki banyak keuntungan mudah dimengerti, sederhana, konvergensi cepat dan sebagainya, tetapi memiliki kekurangan: sensitif untuk memilih pusat pengelompokan awal dan mudah konvergen untuk optimasi local[2]. Fungsi yang paling penting dari algoritma klustering mengukur kesamaan untuk menentukan bagaimana menutup dua pola data ke satu sama lain. Lebih dari ribuan algoritma pengelompokan dirumuskan dalam berbagai makalah penelitian termasuk metode pengelompokan tradisional dan teknik optimasi yang berbeda [3].

K-means klustering belum optimal sehingga perlu menerapkan metode *particle swarm optimization* (PSO). *particle swarm optimization* (PSO) adalah teknik pencarian dan optimasi [3]. Hal ini efisien untuk memecahkan masalah optimasi. Partisi klustering dapat dianggap sebagai masalah optimasi, karena kita perlu mengoptimalkan nilai jarak antar dan antar kluster. PSO dapat dikombinasikan dengan partisi pengelompokan untuk mengatasi keterbatasan partisi klustering, dan kombinasi ini dapat menghasilkan kluster yang berkualitas tinggi. Optimasi ini terpercaya juga dalam memastikan pusat kluster baru berdasarkan total kluste*r* yang tertentukan[4]. Algoritma ini akan dimaksimalkan dengan metode k-means klustering dalam kelompok awal, selanjutnya PSO memperbaiki kelompok data dari hasil bentukan k-means clustering.

Penelitian ini akan dilihat bagaimana Mengidentifikasi daerah endemik penyakit menular berdasarkan asal pasien dengan k-means klustering yang memungkinkan untuk mengoptimasi k-means dengan PSO sehingga dihasilkan informasi yang lebih bermutu

2. Metode

Penelitian ini, mempergunakan cara penelitian eksperimen, untuk tahap penelitian dimulai dari pengumpulan data sampai dengan mendapatkan hasil evaluasi dan hasil untuk mendapatkan tujuan yang akan dicapai[5], adapun tahapannya adalah, di mulai dari penyiapan dataset kemudian dilakukan langkah prepossessing terhadap data untuk menghilangkan outliernya, selanjutnya adalah menormalisasikan data kemudian menggunakan tool rapidminer untuk mengelola data tersebut dengan menggunakan metode PSO-*k-means clustering.*

2.1 Pengumpulan Data

Pada penelitian tersebut, yang data digunakan ini, juga diperoleh dari penelitian yang dilakuan pada Rumah Sakit/Puskesmas di kota semarang propinsi jawa tengah. Data diperoleh dari sistem rekam medis Rumah Sakit Umum Daerah (RSUD) kota Semarang. Sebagai data sekunder dengan mengajukan surat penelitian dari instansi untuk pengambilan data rekam medis pasien yang berkaitan dengan penyakit menular. Beberapa variabel yang terdapat dalam dataset tersebut adalah:

1. Variabel Kecamatan, terdiri dari
    a. Id Kecamatan
    b. Nama Kecamatan
    c. Luas Kecamatan
    d. Jumlah Penduduk

2. Variabel Endemik, terdiri dari
    a. Id Penyakit
    b. Nama Penyakit
    c. Jumlah Pasien

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259

	A	B	C	D	E	F	G	H	I	J	K	L
74	--											
75												
76	INSERT	INTO	`tbl_peny	(`kd_ende	`kd_kepal	`kode_icd	`nm_inggr	`nm_indo	VALUES			
77	(1	'A15'	'A15'	'Respirato	tuberculo	bacteriolo	and	histologic	confirmec	'TBC')		
78	(2	'A20'	'A22'	'Anthrax'	'Antrax')							
79	(3	'A90'	'A92.0'	'Chikungu	virus	disease'	'Cikungunya')					
80	(4	'B50'	'B50'	'Plasmodi	falciparun	malaria'	'Malaria')					
81	(5	'J09'	'J10'	'Influenza	due	to	other	identified	influenza	virus'	'Influenza')	
82	(6	'A75'	'A75'	'Typhus	fever'	'Tipes')						
83	(7	'A90'	'A90'	'Dengue	fever	[classical	dengue]	'	'Demam	Berdarah	Dengue	[DBD]')
84	(8	'A00'	'A01.0'	'Typhoid	fever'	'Demam	Tifoid')					
85	(9	'B65'	'B74.0'	'Filariasis	due	to	Wucherer	bancrofti'	'Kaki	Gajah')		
86	(10	'A80'	'A82'	'Rabies'	'Rabies')							
87	(11	'A30'	'A33'	'Tetanus	neonatoru	'Tetanus')						
88	(12	'B25'	'B26'	'Mumps'	'Gondong')							
89	(13	'B00'	'B05'	'Measles'	'Campak');							
90												
91	--	----------										

Gambar 1. Tabel Penyakit Endemik

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
16	--																		
17																			
18	INSERT	INTO	`tb_data_i	(`col1`	`col2`	`col3`	`col4`	`col5`	`col6`	`col7`	`col8`	`col9`	`col10`	`col11`	`col12`	`col13`	`col14`)	VALUES	
19	(3	'N'	'X'	1	'A00'	'A00.-'	'A00'	'A00'	'Cholera'	'001'	'4-002'	'3-003'	'2-001'	'1-002')					
20	(4	'T'	'X'	1	'A00'	'A00.0'	'A00.0'	'A000'	'Cholera	due	to	Vibrio	cholerae	1	biovar	cholerae'	'001'	'4-002'	'3-003'
21	(4	'T'	'X'	1	'A00'	'A00.1'	'A00.1'	'A001'	'Cholera	due	to	Vibrio	cholerae	1	biovar	eltor'	'001'	'4-002'	'3-003'
22	(4	'T'	'X'	1	'A00'	'A00.9'	'A00.9'	'A009'	'Cholera	unspecifie	'001'	'4-002'	'3-003'	'2-001'	'1-002')				
23	(3	'N'	'X'	1	'A00'	'A01.-'	'A01'	'A01'	'Typhoid	and	paratyphc	fevers'	'002'	'4-002'	'3-003'	'2-003'	'1-004')		
24	(4	'T'	'X'	1	'A00'	'A01.0'	'A01.0'	'A010'	'Typhoid	fever'	'002'	'4-002'	'3-003'	'2-003'	'1-004')				
25	(4	'T'	'X'	1	'A00'	'A01.1'	'A01.1'	'A011'	'Paratyph	fever	A'	'002'	'4-002'	'3-003'	'2-003'	'1-004')			
26	(4	'T'	'X'	1	'A00'	'A01.2'	'A01.2'	'A012'	'Paratyph	fever	B'	'002'	'4-002'	'3-003'	'2-003'	'1-004')			
27	(4	'T'	'X'	1	'A00'	'A01.3'	'A01.3'	'A013'	'Paratyph	fever	C'	'002'	'4-002'	'3-003'	'2-003'	'1-004')			
28	(4	'T'	'X'	1	'A00'	'A01.4'	'A01.4'	'A014'	'Paratyph	fever	unspecifie	'002'	'4-002'	'3-003'	'2-003'	'1-004')			
29	(3	'N'	'X'	1	'A00'	'A02.-'	'A02'	'A02'	'Other	salmonell	infections	'006'	'4-002'	'3-003'	'2-003'	'1-004')			
30	(4	'T'	'X'	1	'A00'	'A02.0'	'A02.0'	'A020'	'Salmonel	enteritis'	'006'	'4-002'	'3-003'	'2-003'	'1-004')				
31	(4	'T'	'X'	1	'A00'	'A02.1'	'A02.1'	'A021'	'Salmonel	sepsis'	'006'	'4-002'	'3-003'	'2-003'	'1-004')				
32	(4	'T'	'X'	1	'A00'	'A02.2'	'A02.2'	'A022'	'Localized	salmonell	infections	'006'	'4-002'	'3-003'	'2-003'	'1-004')			
33	(4	'T'	'X'	1	'A00'	'A02.8'	'A02.8'	'A028'	'Other	specified	salmonell	infections	'006'	'4-002'	'3-003'	'2-003'	'1-004')		
34	(4	'T'	'X'	1	'A00'	'A02.9'	'A02.9'	'A029'	'Salmonel	infection	unspecifie	'006'	'4-002'	'3-003'	'2-003'	'1-004')			
35	(3	'N'	'X'	1	'A00'	'A03.-'	'A03'	'A03'	'Shigellos	'003'	'4-002'	'3-003'	'2-003'	'1-004')					
36	(4	'T'	'X'	1	'A00'	'A03.0'	'A03.0'	'A030'	'Shigellos	due	to	Shigella	dysenteri	'003'	'4-002'	'3-003'	'2-003'	'1-004')	
37	(4	'T'	'X'	1	'A00'	'A03.1'	'A03.1'	'A031'	'Shigellos	due	to	Shigella	flexneri'	'003'	'4-002'	'3-003'	'2-003'	'1-004')	
38	(4	'T'	'X'	1	'A00'	'A03.2'	'A03.2'	'A032'	'Shigellos	due	to	Shigella	boydii'	'003'	'4-002'	'3-003'	'2-003'	'1-004')	
39	(4	'T'	'X'	1	'A00'	'A03.3'	'A03.3'	'A033'	'Shigellos	due	to	Shigella	sonnei'	'003'	'4-002'	'3-003'	'2-003'	'1-004')	
40	(4	'T'	'X'	1	'A00'	'A03.8'	'A03.8'	'A038'	'Other	shigellosi	'003'	'4-002'	'3-003'	'2-003'	'1-004')				

Gambar 2.Tabel icd Penyakit Endemik

2.2 Pengolahan Data Awal

Penelitian ini dibatasi hanya pada tahapan ini melakukan *preprocessing* terhadap data untuk menghilangkan outlier, merapikan setiap variable yang terkait sehingga siap untuk ke tahap penggunaan selanjutnya

2.3 Eksperimen

Ekperimen pada tahan ini adalah menjalankan seluruh prosedur yang telah ditentukan/disusun dalam sebuah kerangka pemikirian untuk menyelesaikan masalah, adapun tahapannya adalah sebagai berikut.:

Tahap 1, Preprosessing

Tahap 2, Algoritma k-means [6]

Setiap tahapan dalam melakukan clustering dengan algoritma K-Means yaitu:

a. Awali dengan memilih jumlah cluster *k*
b. Berikan Inisialisasi *k* pusat *cluster* tersebut maka akan dilakukan dengan berbagai cara. tetapi paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random,
c. Juga alokasikan semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut.halnya juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. tahap ini juga perlu dihitung jarak tiap data ke tiap pusat *cluster*. jarak paling dekat antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana dengan menggunakan teori jarak Euclidean.

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259

d. Juga Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang terkini. Pusat *cluster* merupakan rata-rata dari semua data/objek dalam *cluster* tersebut. Jika dikehendaki bisa juga digunakan median dari *cluster* tersebut. Jadi rata-rata(mean) bukan satu-satunya ukuran yang bisa dipakai.
e. Akhirnya tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi

Tahap 3, optimasi dengan PSO [7]

Langkah-langkah dalam Algoritma PSO untuk klastering

Langkah 1 : Pendefinisian pusat klaster awal.

Pada langkah ini, ditentukan posisi titik awal sebagai pusat klaster data sebanyak *k* titik data secara random.

Langkah 2 : Pengelompokan data ke dalam klaster

Pada langkah ini, data dimasukkan ke dalam salah satu klaster yang mempunyai pusat klaster terdekat dengan data tersebut. Besar kecilnya nilai jarak sangat menentukan letak data tersebut akan dimasukkan dalam klaster mana. Dalam menghitung jarak antara data ke pusat *cluster*, dapat digunakan rumus jarak Euclidean. Perhitungan jarak pada algoritma ini dilakukan untuk masing-masing data ke setiap pusat *cluster*. sehingga jika terdapat N data dan *k* pusat *cluster* maka akan dihasilkan sebanyak (N x *k*) perhitungan jarak. Selanjutnya, dari hasil perhitungan jarak data dengan pusat *cluster* akan dicari nilai minimun untuk masing-masing data. Nilai minimun menunjukkan bahwa data lebih dekat dengan pusat klaster tersebut, sehingga data akan lebih tepat ditempatkan kedalam klaster tersebut.

Langkah 3 : Perhitungan nilai *sum of squared error* (SSE)

Dalam penelitian ini, SSE merupakan *fitness function* yang akan dicari nilainya dalam algoritma *clustering*. SSE inilah yang akan dicari nilai optimalnya (minimun) dengan menggunakan algoritma PSO.

Langkah 4 : Optimasi *cluastering* dengan PSO

Dalam algoritma PSO *clustering* pada penelitian ini, pusat *cluster* mewakili partikel. Set solusi yang diperoleh nantinya adalah pusat *cluster* yang baru, yang diharapkan akan menghasilkan perhitungan SSE yang lebih kecil daripada SSE sebelumnya. Sebelum memulai optimasi klastering untuk iterasi awal perlu didefenisikan terlebih dahulu kecepatan awal partikel (V_0), dengan memperhatikan batas kecepatan. Jika nilai V_0 lebih besar dari batas maksimun atau lebih kecil dari batas minimun, nilai kecepatan ditetapkan sama dengan batas. Dalam algoritma klastering pada penelitian ini, nilai rata-rata yang digunakan adalah nilai rata-rata terbaik dari semua solusi, pada penelitian ini adalah pencarian nilai optimal pada pengelompokkan data ke tiap klaster dengan memperhitungkan nilai terbaik dari semua data. Langkah selanjutnya adalah meng-*update* posisi pusat *cluster*, dengan cara menjumlahkannya dengan nilai kecepatan. Setelah pusat klaster secara random di-*update*, maka akan diperoleh k titik pusat *cluster* dilakukan tahapan dalam langkah 2 dan langkah 3, sehingga diperoleh nilai SSE dan *cluster* terbaik untuk tiap-tiap data.

Langkah 5: Meng-*update* nilai SSE

Nilai SSE yang diperoleh paling akhir kemudian dibandingkan dengan nilai SSE sebelumnya, jika nilai SSE tersebut lebih kecil dari nilai SSE sebelumnya, maka pusat *cluster* yang dihasilkan akan menjadi pusat *cluster* yang baru.

Langkah 6: Kembali ke langkah 4

Pengulangan dilakukan sampai terjadi *stopping condition*, yaitu setelah beberapa kali iterasi sesuai yang telah ditentukan sebelumnya.

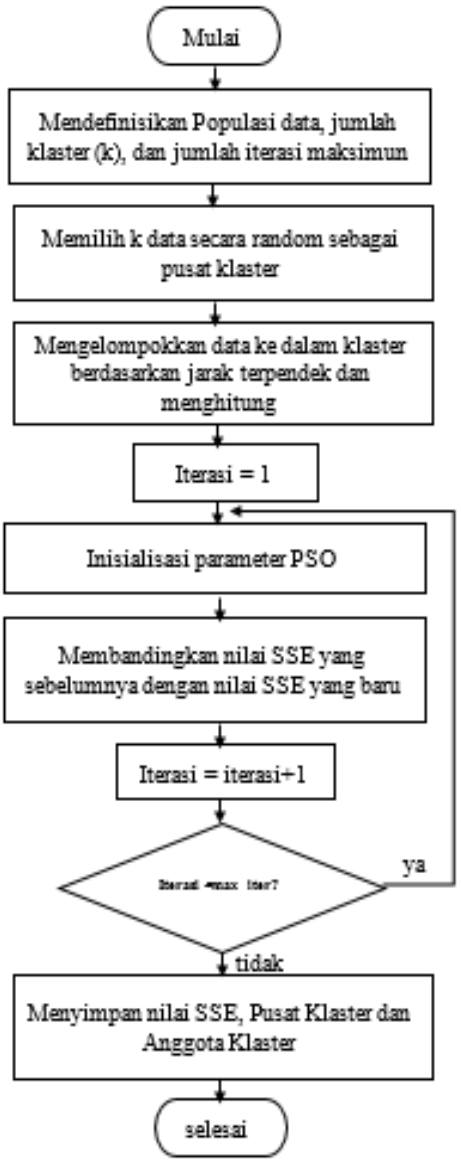
Berikut ini adalah *Pseudo-Code* untuk algoritma PSO *clustering :*

1. Mulai
2. Mendefenisikan populasi data yang akan diklasterkan, jumlah klaster (*k)*, serta jumlah iterasi maksimum (maxiter).
3. Memilih sebanyak k data secara random dari populasi sebagai pusat *cluster*.
4. Mengelompokkan data berdasarkan jarak terdekat data ke pusat klaster. Perhitungan jarak menggunakan rumus jarak Euclidean
5. Menghitung nilai *sum of squared error* (SSE) klaster menggunakan persamaan SSE

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259

6. Inisialisasi parameter PSO, me-*generate* nilai kecepatan awal (V_0) dengan memperhatikan batas kecepatan yang telah ditentukan
7. Iterasi mulai dilakukan
8. Untuk i = 1 sampai *k*, lakukan proses *update* kecepatan;
    a. Mencari X_{gi}, yaitu nilai rata-rata data dari semua solusi
    b. Jika *cluster i* tidak memiliki anggota, kecepatan = 0, sebaliknya jika *cluster i* memiliki anggota, dilakukan proses *update* kecepatan dengan memperhatikan batas kecepatan.
    c. Meng-*Update* posisi pusat *cluster*
9. Selesai *looping i*, diperoleh pusat *cluster* baru
10. Mengulangi langkah 4 dan 5 untuk pusat *cluster* baru.
11. Satu iterasi selesai, mengulangi langkah 7-9 hingga jumlah iterasi mencapai iterasi maksimun
12. Selesai.

Pseudo-code diatas dapat digambarkan dalam bentuk sebagai berikut :



Gambar 3. *Flowchart* Algoritma PSO *Clustering*

2.4 Evaluasi

Pengujian eksperimen yang dilakukan berupa berbandingan Perbandingan dengan *cosine similarity* dalam kluster serta menampilakan uraian analisi daerah endemik dengan validasi davis bouldin index (IDB).

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259

p-ISSN 2087-1716
e-ISSN 2548-7779
ILKOM Jurnal Ilmiah Volume 10 Nomor 3 Desember 2018

3. Hasil dan Pembahasan

3.1 Hasil Pengumpulan Data

Pengumpulan data dilakukan dengan mendapatkan variabel yang sesuai dari dataset endemik, dengan variabel. Data yang digunakan pada penelitian ini dikumpulkan berdasarkan data Penderita Penyakit di kota Semarang RS.Karyadi sebagai Rumah Sakit umum dan rujukan di kota Semarang Jawa Tengah.

Tabel 1. Data Endemik (Contoh)

No	Kd_Endemik	nm_endemik	kd_jnsKlmin	jns_klamin	kd_wilyh	nm_wilyah	Luas_wil	jml_pnddk	Kpdtn_Pndduk	nm_penyakit	jml_pasien
1	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Cholera	1
2	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Cholera due to Vibrio cholerae 01, biovar cholerae	1
3	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Cholera due to Vibrio cholerae 01, biovar eltor	1
4	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Cholera, unspecified	1
5	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Typhoid and paratyphoid fevers	2
6	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Typhoid fever	2
7	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid fever A	2
8	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid fever B	2
9	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid fever C	2
10	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid fever, unspecified	2
11	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Other salmonella infections	6
12	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Salmonella enteritis	6
13	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Salmonella sepsis	6
14	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Localized salmonella infections	6
15	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Other specified salmonella infections	6
16	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Salmonella infection, unspecified	6
17	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Shigellosis	3
18	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Shigellosis due to Shigella dysenteriae	3
19	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Shigellosis due to Shigella flexneri	3
20	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Shigellosis due to Shigella boydii	3
21	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Shigellosis due to Shigella sonnei	3
22	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Other shigellosis	3
23	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Shigellosis, unspecified	3
24	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Other bacterial intestinal infections	6
25	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Enteropathogenic Escherichia coli infection	6

3.2 Preprocessing

Preprosessing dataset endemik dengan menggunakan *fitur Replace Missing Value*, fitur ini menghasilkan data yang telah siap diujikan, adapun hasilnya adalah sebagai berikut :

Tabel 2. Data Endemik tahap Preprocessing

ExampleSet (2536 examples, 1 special attribute, 11 regular attributes) Filter (2,536 / 2,536 examples): all

Row No.	No	Kd_Endemik	nm_endemik	kd_ jnsKlmin	jns_klamin	kd_wilyh	nm_wilyah	Luas_wil	jml_pnddk	Kpdtn_Pnd...	nm_penyakit	jml_pasien
1	1	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Cholera	1
2	2	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Cholera due	1
3	3	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Cholera due	1
4	4	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Cholera, uns	1
5	5	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Typhoid and	2
6	6	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Typhoid feve	2
7	7	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid	2
8	8	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid	2
9	9	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid	2
10	10	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Paratyphoid	2
11	11	3	Cinkunya	N	Laki-laki	1	Banyumanik	25.69	130494	5079	Other salmo	6
12	12	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Salmonella	6
13	13	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Salmonella	6
14	14	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Localized sa	6
15	15	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Other specifi	6
16	16	4	Malariya	T	Perempuan	1	Banyumanik	25.69	130494	5079	Salmonella i	6

Pemilihan atribut dengan memilih variabel yang saling berkaitan yaitu, variabel nama endemik, luas wilayah, jumlah penduduk, kepadatan penduduk dan jumlah penderita/pasien. Dalam pemilihan nama endemik yang memiliki kesamaan rata-rata disetiap wilayah yaitu endemik cinkunya dan malaria sehingga nama endemik yang difokuskan adalah cinkunya.

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259

Tabel 3. Jumlah Masing-masing Endemik

No	Wilayah	Luas	penduduk	Kepadatan	Jml pasien
1	Banyumanik	25.69	130494	5079	5605
2	Candisari	6.54	79706	12187	10598
3	Gayamsari	6.177	73745	11938	627
4	Genuk	27.39	93439	3411	686
5	Smg Barat	21.74	158668	7298	830
6	Smg Tengah	6.14	71200	11596	990
7	Gajah Mungkur	9.07	63599	7012	393
8	Pedurungan	20.72	177143	8549	575
9	Gunung Pati	54.11	75885	1402	360
10	Smg Selatan	5.928	82293	13882	542
11	Smg Timur	7.7	128026	10210	211
12	Tembalang	44.2	147564	3338	708
13	Tugu	31.78	31279	984	735
14	Ngaliyan	37.99	122555	3226	142
15	Smg Utara	10.97	128026	11670	606
16	mijen	57.55	57887	1005	0
	Total	373.695	1621509	112787	23608

Pada eksperimen ini algoritma klustering k-means dengan optimasi PSO (particle swarm optimization) digunakan sebagai pembanding dalam pengelompokkan daerah endemik cinkunya, pengujian menggunakan model *Cluster Number index* (keseluruhan nilainya memiliki kesaman) dan *davies bouldin index* (IDB). Dengan bobot awal w = 0.36 yang di set denga K=2 dengan pertimbangan nilai validasi ini perbandingakan dengan IDB k-means dan IDB k-means+PSO dengan Lbest 1.0 dan Gbest 1.0 menghasilkan IDB 0.165. dengan nilai tersebut menjadi nilai terukur yang bermutu, adapun tabel kedua perbandingan validasi adalah sebagai berikut

Tabel 4 Evaluasi K-Means, K-Means dengan PSO

Metode	Lbest	Gbest	K-Fold	IDB
K-Means	0.25	0.25	5	0.136
	0.25	0.5	5	0.174
	1.0	1.0	5	0.165
PSO	-	-	2	0.164
	-	-	3	0.119

Tabel 5. Evaluasi Hasil K-Means, K-Means dengan PSO

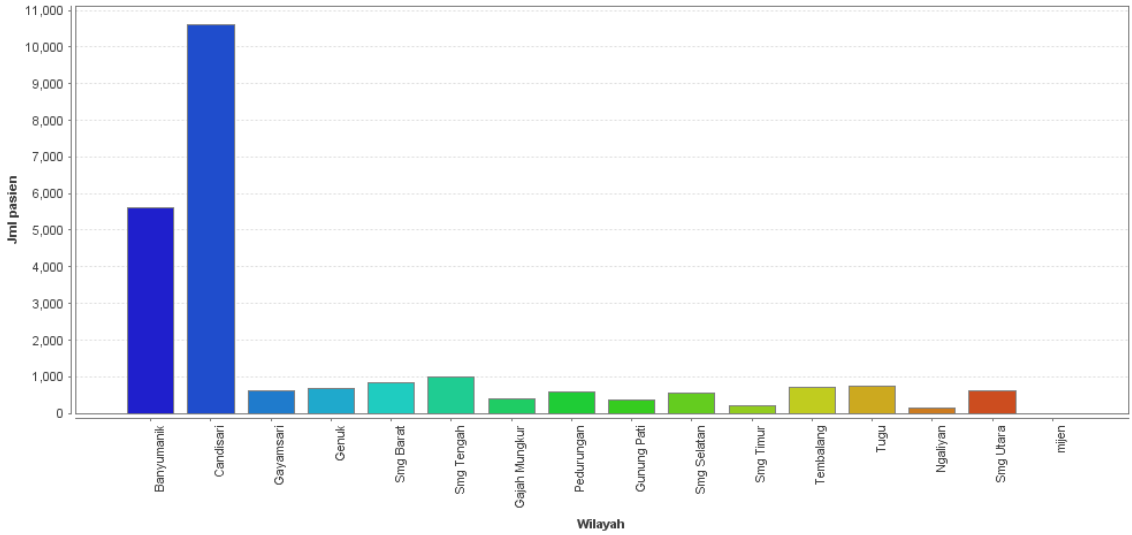
Validasi	Iterasi	K-Means	K-Means dengan PSO
Davie Bouldin Index	10	0. 168	0.136
		0.180	0.174
		-	0.165

Tabel 6. Hasil Eksperimen Data Endemik

Dataset	Metode	Iterasi	K-Fold	Index Davies Bouldin
Endemik (Cikunya)	K-Means	10	2	0.169
	K-Means+PSO Lbest = 1.0 Gbest = 1.0	10	5	0.113

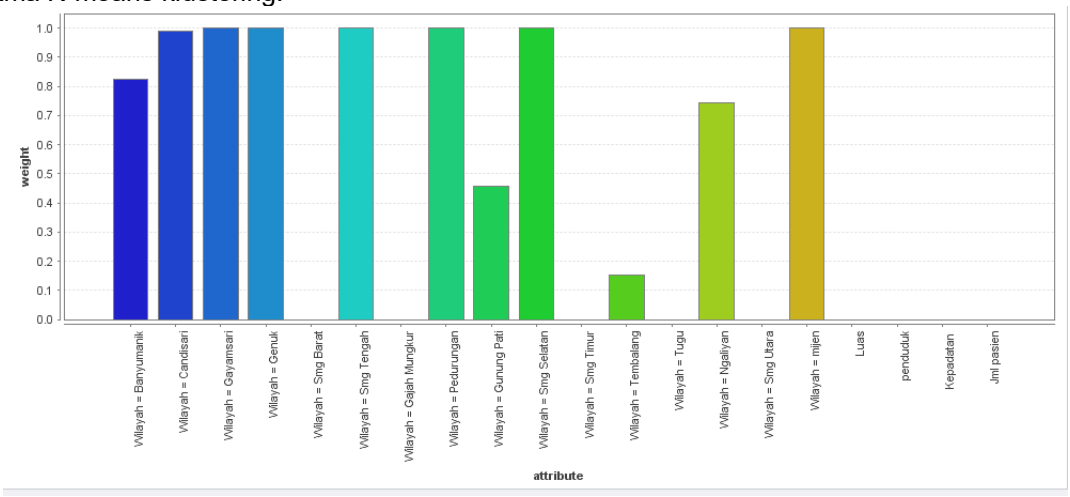
Dari tabel 6 menunjukkan hasil eksperimen endemik untuk algoritma k-means dengan iterasi =10, k-fold =2 memiliki indek davies baouldin = 0.169 dan algoritma k-means dengan PSO, iterasi = 10, k-fold = 5, memiliki index davies bouldin = 0.113, hasil ini memberikan perbedaan k-mean dalam pemilihan kluster adalah random dan k-means dengan PSO memberikan nilai pada hasil sum of square pada hasil random acak, dengan k-fold = 2, memiliki IDB = 0.169 dan k-fold = 5, memiliki IDB 0.113. maka k-fold = 5 memiliki performansi yang lebih baik.

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259



Gambar 4. K-Means Klustering

Gambar 4 menunjukkan wilayah-wilayah endemik dengan jumlah pasien dengan menggunakan algoritma K-means klustering.



Gambar 5. K-Means dengan PSO

Gambar 5. menunjukan wilayah-wilayah endemik dengan didasarkan pada bobot, untuk algoritma k-means dengan optimasi PSO.

4. Kesimpulan dan Saran

Berdasarkan hasil eksperimen dan pembahasan, maka dengan ini dapat simpulkan bahwa Dalam penelitian ini dilakukan pengujian model dengan menggunakan algoritma k-means dengan optimasi PSO (particle swarm optimization) untuk data endemik. Model yang dihasilkan diuji untuk mendapat nilai SSE (sum of square error) dan dari nilai ini digunakan sebagai bobot awal optimasi untuk mendapatkan Nilai IDB (davies bouldin index) sebagai nilai ukur terhadap validasi data endemik. Untuk selanjutnya nilai IDB k-means dengan optimasi PSO dibandingkang dengan k-mean klustering. Maka dapat disimpulkan pengujian model daerah sebaran endemik dengan menggunakan k-means dengn optimasi PSO (particle swarm optimization) Lebih baik dari pada K-Means sendiri, bahwa K-Means dengan optimasi PSO (*particle swarm optimization*) memberikan pemecahan untuk permasalahan identifikasi daerah endemic penyakit menular yang lebih akurat dan bermutu saran yaitu Penambahan variabel yang lebih banyak dan saling berkaitan dalam mengidentifikasi daerah endemik penyakit menular sangat dibutuhkan, sehingga dapat menghasilkan pola klustering daerah endemik dengan tingkat validasi yang lebih akurat.

5. Terima Kasih

Saya dengan ini mengucapkan banyak terima kasih atas sharing ide ilmu dari rekan sesama dosen dan peneliti di fakultas ilmu komputer universitas ichsan gorontalo, semoga di bidang ilmu komputer selalu dikembangkan dan diteliti terus-menerus

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259

Daftar Pustaka

[1] Achmadi UF,2009. Manajemen Penyakit berbasis wilayah. Jurnal Kesehatan Masyarakat Nasional Vol. 3,No.4
[2] Van der Merwe DW, Engelbrecht AP. 2003. Data Clustering using Practicle Swarm Optimization, University of Pretoria, IEEE 0-7803-7804-0/03, hlm 215-220.
[3] Fan Y. *Do "smart" places have less urban health penalty?*.2009. 47th International Making Cities Livable Conference; Portland, Oregon
[4] Han Jiawei, 2006. Kamber M.Data mining: Concepts and techniques.2nd ed.Beijing: China Machine Press
[5] Hasibuan, A Zaenal .2007. Metodelogi Penelitian Pada Bidang Ilmu Komputer dan Teknologi Informasi. 20 Juli 2018
[6] Agusta, Y.2007, *K-means-Penerapan, Permasalahan dan Metode Terkait.* Jurnal Sistem dan Informatika Vol.3 (Februari):47-60
[7] Utami Nanik. Dkk,. Aplikasi Metode *Particle Swarm Optimization*(PSO) dalam *Clustering*. ITS Surabaya

DOI: http://dx.doi.org/10.33096/ilkom.v10i3.342.251-259
