

Analisis *Cluster k-means* dan Autokorelasi Spasial Untuk Identifikasi Pola Penyebaran Kasus Malaria (Studi kasus : Kasus malaria di kabupaten Purworejo tahun 2015)

Dian Pravitasari*, RB. Fajriya Hakim

Departement of Statistics
Faculty of Matematics and Natural Sciences
Universitas Islam Indonesia, Yogyakarta

Lecturer at Departement of Statistics
Faculty of Matematics and Natural Sciences
Universitas Islam Indonesia, Yogyakarta

dianpravitasari12@gmail.com
hakimf@fmipa.ac.id

Abstrak

Malaria merupakan salah satu penyakit menular yang masih menjadi masalah kesehatan masyarakat di dunia termasuk Indonesia. Kabupaten Purworejo menyumbang kasus malaria terbesar di Provinsi Jawa Tengah pada Tahun 2011. Salah satu peran penting epidemiologi dalam penyakit infeksi adalah mempelajari penyebaran penyakit dalam dimensi waktu, tempat dan karakteristik orang. Penelitian ini bertujuan untuk mengetahui pengelompokan wilayah endemis penyakit malaria di Kabupaten Purworejo tahun 2015 dengan menggunakan cluster K-Means dan mengetahui autokorelasi spasial dalam penyebaran kasus malaria melalui metode Indeks Moran's dan Moran's Scatterplot. Hasil dari penelitian ini pengelompokan dibagi menjadi tiga kelompok. Kelompok 1 adalah wilayah endemis sedang, kelompok 2 wilayah endemis rendah dan kelompok 3 adalah wilayah endemis tinggi dengan nilai validasi cluster F-measure sebesar 93,75%. Untuk autokorelasi spasial melalui pengujian secara statistik nilai Z_{hitung} 2,309 yang berarti bahwa kasus malaria tidak terjadi secara random dengan kata lain terdapat pengaruh kedekatan antar lokasi pada banyaknya kasus malaria yang terjadi.

Kata kunci: Malaria, K-Means, Indeks Moran's

PENDAHULUAN

Sulistiyawati (2012) dalam penelitiannya “Spasial *Clustering* Pada Kejadian Luar Biasa Malaria di Kabupaten Purworejo” menyebutkan bahwa pada awal kemunculan dengan status KLB di Kabupaten Purworejo yaitu bulan November 2010 adalah di Kecamatan Bagelen, sementara untuk kecamatan Kaligesing memang merupakan kecamatan endemis malaria[1]. Penelitian serupa dilakukan oleh Yudianto (2009) dalam penelitiannya yang berjudul “Analisis Wilayah Sebaran Penyakit Malaria di Kabupaten Ciamis” menyebutkan bahwa bahwa faktor ketinggian lokasi, jaringan sungai, lahan basah, dan permukiman berhubungan dengan persebaran Malaria di Kabupaten Ciamis, dimana ketinggian wilayah, luas jaringan sungai, luas wilayah lahan basah dan kepadatan [2]. Penelitian selanjutnya yang menggunakan autokorelasi spasial pernah dilakukan Arrowiyah dan Sutikno (2010) yang berjudul “*Spatial Pattern Analysis* Kejadian Penyakit Demam Berdasarah *Dengue* untuk Informasi *Early Warning* Bencana di Kota Surabaya” menyebutkan bahwa perbandingan hasil pengujian dengan menggunakan indeks Moran's dan *Geary's C* memberikan informasi bahwa indeks Moran's lebih sensitif dibandingkan dengan indeks *Geary's C*. [3]

Kemestrian Kesehatan RI (2009) menjelaskan bahwa tindakan pencegahan kasus malaria dapat dilakukan dengan sistem kewaspadaan dini kejadian luar biasa malaria dengan perlindungan kelompok yang paling rentan terhadap malaria berdasarkan karakteristik masyarakat dan tindakan intervensi di daerah endemis malaria. Oleh karena itu perlu

pemanfaatan pendekatan epidemiologi spasial dalam penyelesaian masalah malaria. Berdasarkan uraian tersebut, maka perlu dilakukan pengelompokkan wilayah endemis penyakit malaria di Kabupaten Purworejo dan melihat bagaimana keterkaitan antara wilayah satu dengan wilayah lainnya dalam kasus malaria di Kabupaten Purworejo tahun 2015.

TEORI

Menurut kamus Depkes, malaria merupakan suatu penyakit yang disebabkan oleh sejenis parasit genus *Plasmodium* dengan gejala utama demam yang berulang-ulang [4]. Penilaian kasus malaria berdasarkan Kemenkes tahun 2009 dengan parameter yang digunakan pada pengamatan rutin malaria adalah *Annual Parasite Incidence* (API). API merupakan jumlah dari penderita baru di suatu daerah dalam satu tahun terhitung per seribu penduduk.

$$API = \frac{\text{kasus malaria positif dalam satu tahun}}{\text{jumlah penduduk daerah tersebut}} \times 1000 \quad (1)$$

Tujuan utama analisis kelompok adalah mengklasifikasi objek seperti orang, produk, toko, perusahaan ke dalam kelompok-kelompok yang relatif homogen didasarkan pada suatu set variabel yang dipertimbangkan untuk diteliti (Supranto, 2004) [5]. Penelitian ini menggunakan ukuran kesamaan jarak dengan jarak *euclidean*.

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

Dengan:

d_{ij} : jarak antara objek i dan j

X_{ik} : nilai variabel k untuk objek i

X_{jk} : nilai variabel k untuk objek j

p : jumlah variabel yang digunakan

i dan j menyatakan objek yang akan dikelompokkan.

Menurut Suwarsa (2013) pada proses peng-*cluster*-an, prosedur untuk mengevaluasi hasilnya dikenal sebagai validitas *cluster* dan dapat dilakukan dengan berbagai macam pengukuran yang disebut dengan pengukuran validitas *cluster*. Pengujian *accuracy* dari hasil F-Measure, digunakan suatu standar yang disebut matriks *confusion*. Matriks *confusion* berisi informasi mengenai peng-*cluster*-an yang sebenarnya dan prediksi peng-*cluster*-an yang dilakukan oleh sistem. [6]

Tabel 1. Matriks Confusion

		Kelas Prediksi	
		a	b
Kelas sebenarnya	A	TP	FP
	B	FN	TN

Keterangan :

True Positive (TP), yaitu jumlah objek dari kelas a yang benar dan diklasifikasikan sebagai kelas A.

True Negative (TN), yaitu jumlah objek dari kelas b yang benar diklasifikasikan sebagai kelas B.

False Positive (FP), yaitu jumlah objek dari kelas b yang salah diklasifikasikan sebagai kelas A

False Negative (FN) yaitu jumlah objek dari kelas a yang salah diklasifikasikan sebagai kelas B.

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (3)$$

Hukum pertama tentang geografi yang dikemukakan oleh W Tobler dalam Anselin (1988) yang berbunyi : *”Everything is related to everything else, but near things are more related than distant things”*. Segala sesuatu saling berhubungan satu dengan yang lainnya, tetapi sesuatu yang dekat lebih mempunyai pengaruh daripada sesuatu yang jauh [7]. Menurut Lembo (2006) autokorelasi spasial adalah korelasi antara variabel dengan dirinya sendiri berdasarkan ruang atau dapat juga diartikan suatu ukuran kemiripan dari objek di dalam suatu ruang (jarak, waktu dan wilayah), jika terdapat pola sistematis di dalam penyebaran sebuah variabel, maka terdapat autokorelasi spasial. Adanya autokorelasi spasial mengindikasikan bahwa nilai atribut pada daerah tertentu terkait oleh nilai atribut tersebut pada daerah lain yang letaknya saling berdekatan (bertetangga). [8]

Indeks Moran’s adalah sebuah tes statistik lokal untuk melihat nilai autokorelasi spasial, yang mana digunakan untuk mengidentifikasi suatu lokasi dari pengelompokkan spasial atau autokorelasi spasial. Nilai statistik ini mewakili kondisi rata-rata keterkaitan diseluruh wilayah yang diformulasikan dengan :

$$I = \frac{N \sum_i^N \sum_j^N \tilde{w}_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i^N \sum_j^N w_{ij} \sum_i^N (Y_i - \bar{Y})^2} \tag{3.8}$$

dimana:

- I = Indeks Moran's,
- N = banyak lokasi kejadian,
- x_i = nilai pengamatan pada daerah i ,
- Y_j = nilai pengamatan pada daerah j ,
- \bar{Y} = rata-rata dari nilai pengamatan diseluruh wilayah,
- \tilde{w}_{ij} = elemen pada bobot matriks yang terstandarisasi antara daerah i dan j ,

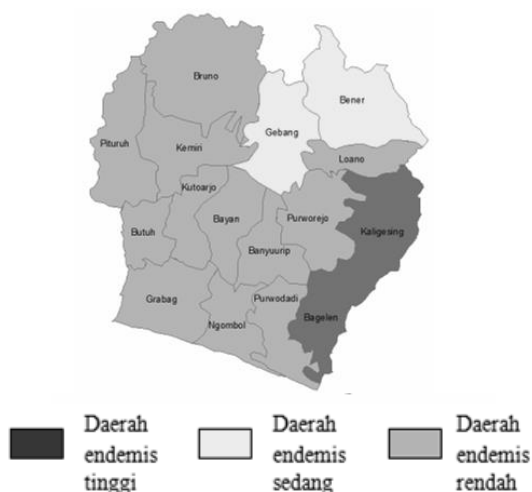
Pencarian nilai indeks Moran's juga dapat dilakukan dengan menggunakan *software* OpenGeoDa. OpenGeoDa membantu pengecekan korelasi spasial dan heterogenitas suatu data. Melalui nilai Indeks Moran's dapat diketahui pola pengelompokan wilayah yang terjadi. Rentang nilai Indeks Moran's berada pada interval $-1 < 0 < 1$. Hasil Indeks Moran's mendekati 1 menandakan bahwa terjadi pengelompokan wilayah dengan karakteristik yang sama atau dapat dikatakan terjadi autokorelasi spasial positif. Lee dan Wong (2001) menyebutkan bahwa *Moran's Scatterplot* adalah salah satu cara untuk menginterpretasikan statistik Indeks Moran's. *Moran's Scatterplot* merupakan alat untuk melihat hubungan antara (nilai pengamatan yang sudah distandarasi) dengan (nilai rata-rata daerah tetangga yang telah distandarasi). [9]

HASIL DAN DISKUSI

Populasi dalam penelitian ini adalah data laporan penemuan kasus malaria pada seluruh kecamatan di Kabupaten Purworejo, yang berjumlah 16 kecamatan. Data sekunder kasus malaria tahun 2015 digunakan untuk melihat trend/kecenderungan kasus malaria. Variabel yang digunakan adalah faktor penentu wilayah endemis malaria antara lain jumlah kasus malaria positif, curah hujan, ketinggian, luas lahan basah, luas jaringan sungai, kepadatan penduduk, dan jumlah fasilitas kesehatan. Pemilihan variabel didasarkan dari beberapa faktor yang mempengaruhi kejadian malaria.

Tiap kecamatan melaporkan kejadian malaria yang tercatat oleh petugas kesehatan melalui laporan puskesmas ke dinas kesehatan. Dengan menggunakan ketentuan dari Ditjen PP dan PL (2009) tentang pembagian wlayah endemisitas yakni :

1. Daerah Non Endemis dengan nilai API 0
2. Daerah endemis rendah dengan nilai API 0 sampai 1 per 1000 penduduk
3. Daerah endemis sedang dengan nilai API 1 sampai 5 per 1000 penduduk
4. Daerah endemis tinggi dengan nilai API lebih dari 5 per 1000 penduduk



Gambar 1. Pemetaan Endemisitas Malaria Per-Kecamatan di Kabupaten Purworejo tahun 2015 (Menurut Ditjen PP dan PL)



Gambar 2. Pemetaan Endemisitas Malaria di Kabupaten Purworejo tahun 2015 (Cluster K-Means (k=3)).

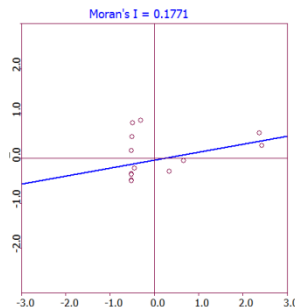
Berbeda hasil dengan pengelompokan menurut Duitjen PP dan PL, berdasarkan pengelompokan wilayah endemis dengan menggunakan metode *K-means*, terdapat 2 kecamatan dengan kategorikan wilayah endemis tinggi, 3 kecamatan dengan kategori wilayah endemis sedang dan 11 kecamatan dengan kategori rendah

Dilakukan evaluasi untuk melihat seberapa baik hasil *clustering* dengan menggunakan sekumpulan kelas acuan dimana kelas acuan diperoleh dari hasil pengelompokan yang telah dikelompokkan berdasarkan nilai API Kabupaten Purworejo Tahun 2015.

$$\text{Akurasi} = \frac{(2+2+11)}{16} \times 100\% = 93,75\%$$

Karena nilai akurasi dengan *F-measure* yang dihasilkan 93,75% semakin mendekati 100% menunjukkan bahwa *clustering* memberikan hasil yang semakin baik.

Analisis spasial dengan menggunakan Indeks Moran's dapat memperlihatkan apakah terdapat autokorelasi spasial pada penyebaran kasus malaria di Kabupaten Purworejo yang tersebar di 16 Kecamatan. Penelitian ini digunakan *Queen Contiguity* dalam penentuan tetangga, artinya wilayah tetangga ditentukan berdasarkan persinggungan sisi perbatasan atau persinggungan ujung (*vertex*) perbatasan dengan wilayah lain.



Gambar 3. Moran's I dan Moran's Scatterplot

Nilai Moran's yang dihasilkan *software OpenGeoDa* hasilnya sama dengan perhitungan manual yakni sebesar 0.1771, angka ini mengarah pada autokorelasi positif diperkuat dengan garis lurus pada pola *moran's scatterplot* dalam penyebaran kasus malaria di Kabupaten Purworejo atau dengan kata lain terdapat pengaruh kedekatan antar lokasi pada banyaknya jumlah kasus malaria di Kabupaten Purworejo, kemudian dilakukan pengujian secara statistik dengan membandingkan nilai Z_{hitung} dengan nilai Z_{tabel} .

Perhitungan Z_{hitung} secara manual juga dapat dilakukan seperti di bawah ini.

$$I = 0.1771$$

$$E(I) = -\frac{1}{N-1} = -\frac{1}{16-1} = -\frac{1}{15} = -0.06667$$

$$Var(I) = \frac{N^2 S_1 - N S_2 + 3S_0^2}{S_0^2(N^2 - 1)} = \frac{16^2(3.959) - 16(65.863) + 3(16^2)}{16^2(16^2 - 1)}$$

$$= 0.01114$$

$$Z_{hitung} = \frac{I - E(I)}{\sqrt{Var(I)}}$$

$$Z_{hitung} = \frac{0.1771 - (-0.06667)}{\sqrt{0.01114}}$$

$$= 2.309$$

Selanjutnya dilakukan uji hipotesis untuk mengetahui autokorelasi spasial pada kasus malaria di Kabupaten Purworejo, dengan tingkat signifikansi $\alpha = 5\%$ dan H_0 akan ditolak jika $|Z_{hitung}| > Z_{tabel}$,

Hipotesis yang digunakan adalah :

$H_0 : I = 0$ (tidak ada autokorelasi antar lokasi)

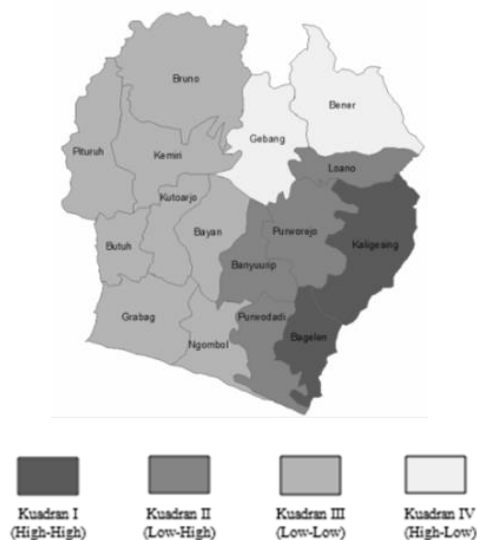
$H_1 : I \neq 0$ (ada autokorelasi antar lokasi)

Matriks Confusion

	C _{tinggi}	C _{sedang}	C _{rendah}
C _{tinggi}	2	0	0
C _{sedang}	0	2	1
C _{rendah}	0	0	11

Berdasarkan perhitungan, karena nilai Z_{hitung} (2.309) $> Z_{tabel}$ (1.96) maka tolak H_0 dengan kata lain terdapat autokorelasi spasial pada penyebaran kasus malaria di Kabupaten Purworejo.

Tahapan selanjutnya adalah memvisualisasikan hasil tersebut kedalam bentuk *plot. Moran's scatterplot*, membagi daerah menjadi 4 kuadran.



Gambar 4. Peta Sebaran Kasus Malaria di Kabupaten Purworejo

KESIMPULAN

Berdasarkan pada pengujian yang telah dilakukan, dapat disimpulkan bahwa menggunakan *cluster K-Means* Kelompok 1 adalah wilayah endemis sedang, kelompok 2 wilayah endemis rendah dan kelompok 3 adalah wilayah endemis tinggi dengan nilai validasi *cluster F-measure* sebesar 93,75%. Untuk autokorelasi spasial melalui pengujian secara statistik nilai Z_{hitung} 2,309 yang berarti bahwa kasus malaria tidak terjadi secara random dengan kata lain terdapat pengaruh kedekatan antar lokasi pada banyaknya kasus malaria yang terjadi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Allah SWT. Penulis juga mengucapkan terima kasih kepada seluruh Dosen Statistika Universitas Islam Indonesia, Yogyakarta atas seluruh dedikasi dan peran yang sangat berarti dan bermanfaat selama proses pembelajaran. Terima kasih kepada Bapak RB. Fajriya Hakim bimbingannya sehingga dapat terselesaikan penelitian ini. Kemudian yang terakhir penulis mengucapkan terima kasih kepada orang tua dan rekan-rekan yang telah membantu dan selalu memberikan dukungan secara moril yang tak terhingga. Semoga Allah membalas semua kebaikan yang telah kalian berikan, Aamiin.

REFERENSI

1. Sulistyawati. 2012. Spasial Clustering Pada Kejadian Luar Biasa Malaria di Kabupaten Purworejo. Tesis. Universitas Gadjah Mada.
2. Yudianto. 2009. Analisis Wilayah Sebaran Penyakit Malaria di Kabupaten Ciamis. Penelitian. Universitas Indonesia.
3. Arrowiyah dan Sutikno. 2010. *Spatial Pattern Analysis Kejadian Penyakit Demam Berdasarah Dengue untuk Informasi Early Warning Bencana di Kota Surabaya*. Paper.pdf. <http://digilib.its.ac.id/public/TTS-Undergraduate-15561-Paper-pdf.pdf> diakses tanggal 15 April 2016 Pukul 10:03 WIB.
4. Kemenkes RI. 2009. *Pedoman Penemuan Kasus Malaria*. Jakarta : Kementrian Kesehatan Republik Indonesia.
5. Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta:Penerbit Rineka Cipta
6. Suwarsa, R. Yiska Deviarani., Ratnawati, Dian Eka., Muflikhah Lailil. 2013. *Implementasi K-Modes Pada Clustering Data Kategori Menggunakan New Dissimilarity Measure*. Malang : Universitas Brawijaya Malang.
7. Anselin, L. (1988). *Spatial Data Analysis with GIS : An Introduction to Application in the Social Sciences*. National Center for Geographic Information and Analysis of California Santa Barbara, CA93106.
8. Lembo, A.J. 2006^a. Spatial Autocorrelation. Cornell University. <http://www.css.cornell.edu/courses/620/lectute9.ppt>
9. Lee J. and Wong S.W.D. 2001. *Statistical Analysis with Arcview GIS*, John Willey & Sons, Inc., United Stated of America.