

IMPLEMENTASI METODE DENSITY BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE UNTUK Mencari ARAH PENYEBARAN WABAH DEMAM BERDARAH

Studi Kasus: Data Dinas Kesehatan Kodya Jogjakarta

Andreas Yuwono
Yetli Oslan, Djoni Dwijono

Abstrak

Penyakit demam berdarah yang biasa disebut DB merupakan salah satu wabah penyakit yang menyumbang angka kematian yang cukup besar di Indonesia. Cara pencegahan yang banyak dilakukan saat ini adalah dengan metode preventif yaitu melakukan pengasapan. Dalam penelitian ini, dikembangkan suatu sistem untuk mencari arah penyebaran wabah demam berdarah dengan memanfaatkan metode *Density Based Spatial Clustering Of Applications With Noise* dan teknologi GIS untuk membantu visualisasinya.

Kata kunci: GIS, Spatial Clustering, DBSCAN, Demam Berdarah

1. Pendahuluan

Sekitar 2-3 tahun terakhir, di Indonesia khususnya di kodya Jogja memiliki suatu permasalahan di bidang kesehatan. Permasalahan tersebut adalah mengenai penyebaran penyakit demam berdarah yang biasa disebut DB. Jumlah penderita DB tidak hanya puluhan namun mencapai ratusan hanya dalam kurun waktu yang singkat (Januari-Maret, periode penyebaran wabah DB yang tinggi). Salah satu penyebab cepatnya penularan penyakit ini adalah vektor penularnya yaitu nyamuk *Aedes Aegypti* betina. Penanganan terbaik untuk mengatasi penyebaran wabah DB adalah dengan metode preventif atau pencegahan jangsan sampai nyamuk *Aedes Aegypti* berkembang biak salah satunya dengan melakukan pengasapan. Sebelum melakukan pengasapan, terlebih dahulu harus ditentukan di mana wilayah yang memiliki banyak jumlah penderita demam berdarah dan wilayah yang memiliki jumlah penderita sedikit secara akurat. (Agushyana, 2005).

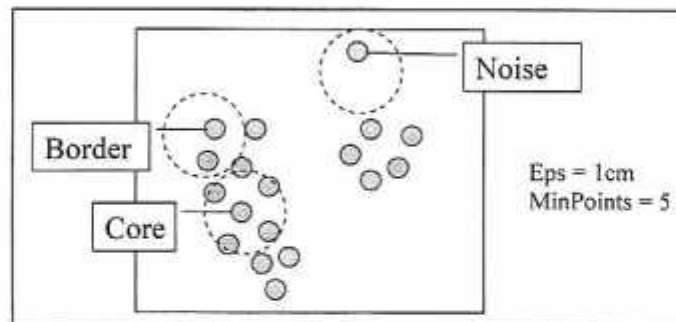
Solusi untuk mengatasi masalah tersebut adalah dengan membuat sistem yang dapat memberikan informasi mengenai pengelompokan data penderita demam berdarah kedalam *cluster-cluster* sehingga diketahui wilayah mana yang jumlah penderitanya banyak dan sedikit. Sistem tidak akan dapat dibangun bila tidak memiliki dasar-dasar pengolahan data yang baik serta tanpa didukung infrastruktur yang memadai. Dasar-dasar pengolahan data dapat diperoleh selama perkuliahan antara lain ilmu *Data Mining* dan Sistem Informasi Geografi (SIG). Sedangkan infrastruktur yang tidak kalah penting selain komputer sebagai media pengolah data adalah *Global Positioning System* (GPS). Sistem dibangun menggunakan *Microsoft Visual Basic 6.0* dan *Arc View GIS 3.3* sebagai media visualisasinya.

2. Konsep dasar

Konsep dari pembuatan sistem ini adalah melakukan pengelompokan berdasarkan tingkat kepadatan data penderita demam berdarah (berupa data koordinat *Universal Transverse Mercator* - UTM, peta dunia dalam bidang 2 dimensi- yang diperoleh dari GPS) dengan menghitung jarak koordinat titik-tempat tinggal penderita demam berdarah- satu dengan titik yang lain. Setelah pengelompokan data dilakukan, selanjutnya dilakukan proses untuk mendapatkan arah penyebaran data yang diasumsikan sebagai pergerakan nyamuk *Aedes Aegypti*.

Metode yang digunakan untuk mengelompokkan data yaitu *Density Based Spatial Clustering Of Applications With Noise* yang biasa disingkat DBSCAN. Beberapa konsep penting dalam metode DBSCAN (Grizaite, 2006):

- a. DBSCAN memerlukan 3 parameter untuk melakukan proses *clusterisasi*. 3 Parameter tersebut adalah :
 - K , lebar atau jumlah titik tetangga.
 - Eps , jarak maksimal antara 2 titik yang diijinkan dalam satu *cluster*.
 - $MinPoints$, adalah jumlah minimal titik sehingga sebuah *cluster* dapat terbentuk.
- b. DBSCAN memiliki 3 jenis tipe titik setelah proses *clustering* dilakukan yaitu:
 - *Core Point*, adalah titik yang berada dalam suatu *cluster* dan memiliki titik tetangga (*eps-neighbourhood*) sedikitnya sebanyak $MinPoints$.
 - *Border Point*, adalah titik yang berada dalam suatu *cluster* namun jumlah titik tetangganya $< MinPoints$.
 - *Noise Point*, adalah titik yang tidak termasuk anggota *cluster* manapun.

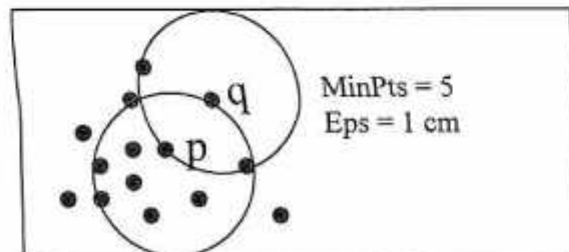


Gambar 1. Ilustrasi *core*, *border* Dan *noise*

- c. Sebuah titik q dikatakan tetangga dari titik p apabila

$$Neps-neighbourhood(p) = \{q \text{ belongs to } D \mid dist(p,q) \leq Eps\} \quad [2.2]$$

$$dist(p,q) = \text{jarak antara titik } p \text{ dan titik } q.$$
- d. *Directly Density-Reachable* merupakan suatu titik yang terhubung langsung dengan titik lain. Sebuah titik q dikatakan *directly density-reachable* titik p apabila :
 - q merupakan anggota $Neps-neighbourhood(p)$
 - $Neps-neighbourhood(p) \geq MinPts$ atau dengan kata lain titik p merupakan *core point*.



Gambar 2. Ilustrasi Titik q *Directly Density-Reachable* Dari Titik p

3. Implementasi Sistem



Gambar 5. Form Menu Utama

Sistem dibagi menjadi 3 bagian utama yaitu :

a. Sub Sistem Membangun Gudang Data:



Gambar 6. Pilihan Tahap 1

Tahap 1 :
Melakukan pemilihan file sumber data dalam kasus ini file Excel.



Gambar 7. Pilihan Tahap 2

Tahap 2 :
Memilih metode *cleaning* yang digunakan: apakah akan meng-*update* data yang kotor atau langsung menghapus data yang kotor



Gambar 8. Pilihan Tahap 3

Tahap 3: Pada tahap 3 dilakukan proses pembersihan data sesuai dengan metode yang dipilih di tahap sebelumnya.



Gambar 9. Pilihan Tahap 4

Tahap 4: Sistem akan mengecek data yang telah bersih, sebelum data-data tersebut dipindahkan ke *warehouse*



Gambar 10. Pilihan Tahap 5

Tahap 5: Sistem akan melakukan migrasi data yang telah bersih untuk disimpan ke dalam gudang data (*warehouse*).

b. Sub Sistem Proses Clustering



Gambar 11. Tahap 1 Sub Sistem Proses Clustering

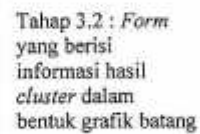
Tahap 1 : Sistem akan melakukan seleksi data dari *warehouse*, sehingga tidak semua data diperlukan dalam proses *clustering*.

Tahap 2 :
Dilakukan penentuan parameter yang akan digunakan.
Meliputi MinPoints, Eps dan K

Gambar 12. Tahap 2 Sub Sistem Proses *Clustering*

Tahap 3.1: Tahap ini penyajian berupa *form* yang berisi informasi hasil *clustering*

Gambar 13. Form informasi hasil *clustering*

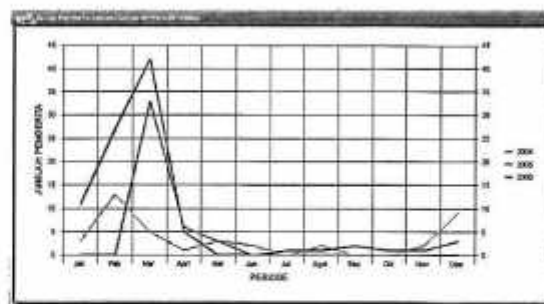


Gambar 14. From informasi hasil *cluster* dalam bentuk grafik batang



Tahap 3.3 :
Form yang berisi informasi rinci data hasil *clustering* dalam bentuk *Tree*

Gambar 15. Form informasi rinci dalam bentuk *Tree*



Tahap 3.4 :
Form yang berisi visualisasi informasi waktu terjadinya peledakan wabah DB dalam bentuk grafik

Gambar 16. Form berisi visualisasi informasi ledakan wabah DB

c. Sub Sistem Visualisasi Hasil Clustering Dengan Arc View GIS

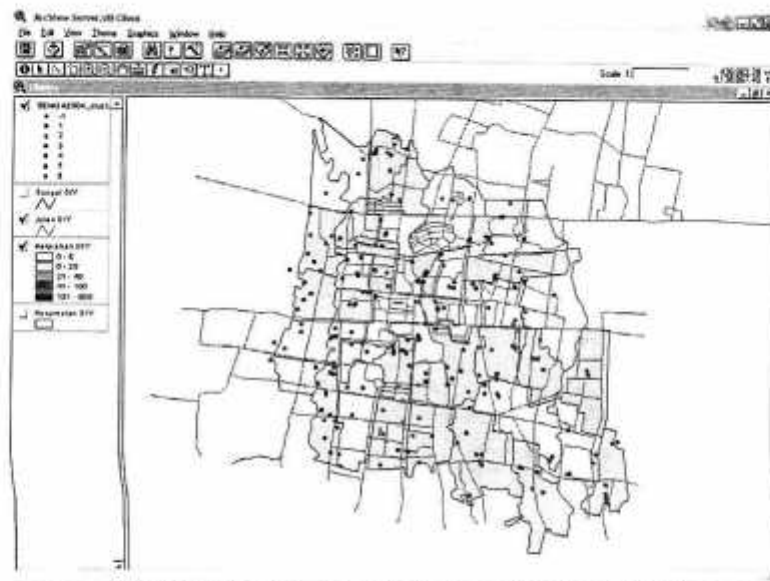


Sebelum melakukan koneksi dengan Arc View, dilakukan pemilihan file hasil *clustering* terlebih dahulu. Hasil Visualisasi ditunjukkan pada Gambar 19 dan 20 di bawah.

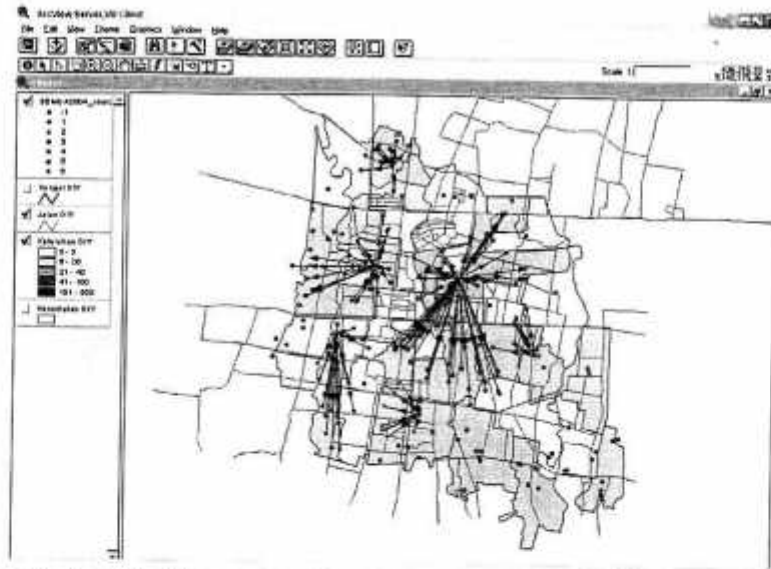
Gambar 17. Pemilihan file hasil *clustering*



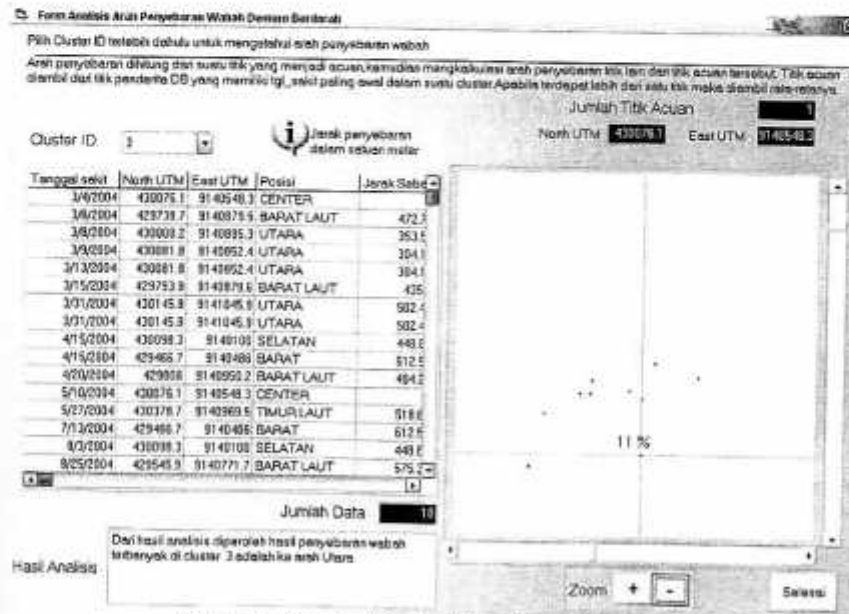
Gambar 18. Form pengontrol aplikasi Arch View GIS 3.3



Gambar 19. Hasil Visualisasi Cluster dengan Arc View GIS 3.3



Gambar 20. Hasil Pemetaan Arah Penyebaran Wabah DB dengan Arc View GIS 3.3



Gambar 21. Form Analisis Arah Penyebaran Wabah DB

4. Kesimpulan

Dengan memperhatikan secara keseluruhan hasil penelitian ini, maka penulis menarik beberapa kesimpulan sebagai berikut:

- Metode DBSCAN dapat menjawab kebutuhan akan informasi area manakah yang memiliki tingkat kepadatan penderita demam berdarah tinggi ditunjukkan oleh adanya *cluster* dan area manakah yang memiliki tingkat kepadatan penderita wabah demam berdarah rendah - ditunjukkan oleh adanya *noise*.
- Metode DBSCAN membentuk sejumlah *N* buah *cluster* secara otomatis, sedangkan acuan yang dibutuhkan untuk membentuk *cluster* adalah seberapa padat jumlah data dalam satu *cluster* yang diinginkan -*MinPoints*- dan seberapa jauh jarak maksimal antara 2 titik agar sebuah *cluster* diijinkan terbentuk -*Eps*-.
- Semakin besar nilai *Min Points* yang diberikan, maka *cluster* yang terbentuk semakin sedikit.
- Semakin besar nilai *Eps* maka *cluster* yang terbentuk semakin sedikit, begitu pula sebaliknya.
- Apabila dibutuhkan informasi berapakah jarak maksimal antara 2 titik agar sebuah *cluster* dengan kepadatan data sejumlah *N* dapat terbentuk, sistem memberikan bantuan dengan nilai *K*. Nilai *K* diisi dengan jumlah kepadatan data yang diinginkan, dan akan menghasilkan rata-rata jarak yang dapat digunakan sebagai acuan penentuan *eps*.
- Hasil visualisasi *cluster* dengan menggunakan Metode DBSCAN tidak dapat secara tepat 100 % menunjukkan arah penyebaran wabah demam berdarah. Hal ini dikarenakan pergerakan nyamuk yang tidak menentu.
- Wabah demam berdarah cenderung terjadi di bulan Januari, Pebruari dan Maret .

Daftar Pustaka

- Agusthybana, Farid dan Cahya Tri Purnami. *Sistem Surveilans Demam Berdarah Dengue Berbasis Komputer untuk Perencanaan, Pencegahan dan Pemberantasan DBD di Kota Semarang*. Inovasi, Vol. 4/ XVII/ Agustus 2005
- Grizaitė, Giedrė dan Roland Innerhofer-Oberperfler, 2006, *DBSCAN Clustering Algorithm*. <http://www.inf.unibz.it/dis/teaching/DWDM05/reports/dbscan.pdf>, 5 September 2006.
- Han, Jiawei dan Micheline Kamber, 2001, *Data Mining : Concept and Techniques*. San Fransisco. Morgan Kauffman Publishers
- Kantardzic, Mehmed, *Data Mining : Concepts, Models, Methods, and Algorithms*. A John Willey & Sons, Inc., Publication., IEEE Press., 2003
- Kimball, Ralph dan Margy Ross, 2002, *The Data Warehouse Toolkit*. New York. John Wiley and Sons, 2nd edition
- Kusumo, Drs. Ario Suryo, 2003, *Buku Latihan Pemrograman Database Dengan Visual Basic 6.0*. Jakarta. Elex Media Komputindo, Cetakan Ketiga
- Moreira, Adriano dan Carneiro, Sofia - Maribel Y.Santos, 2006, *Density Based Clustering Algorithms -DBSCAN and SNN*. [http:// get.dsi.uminho.pt/local /download/SNN&DBSCAN.pdf](http://get.dsi.uminho.pt/local/download/SNN&DBSCAN.pdf), 5 September 2006.
- O'Sullivan, David dan Unwin, David, 2003, *Geographic Information Analysis*, Danvers: John Wiley and Sons, Inc.
- Prahasta, Eddy, 2004, *Pemrograman Bahasa Script Avenue*. Bandung: Informatika.