

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, S.Kom., MT., MT, Drs. Djoni Dwijono, Akt., MT

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 jangan 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).

2. Rumusan Masalah

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. Sedangkan infrastruktur yang tidak kalah penting selain komputer sebagai media pengolah data adalah GPS (*Global Positioning System*). Sistem dibangun menggunakan Microsoft Visual Basic 6.0 dan Arc View GIS 3.3 sebagai media visualisasinya.

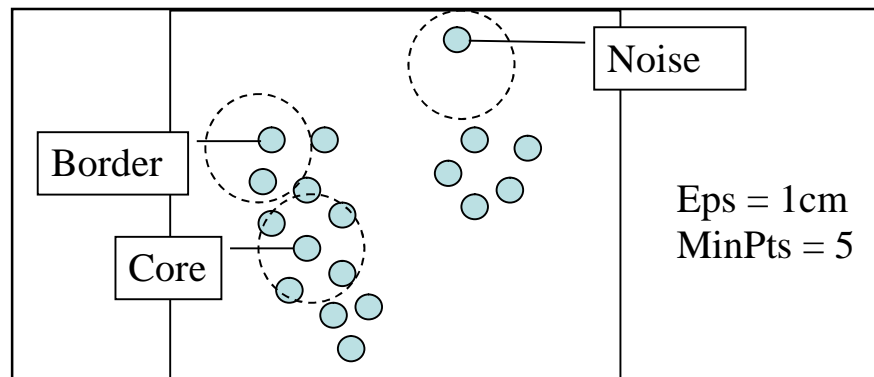
Konsep dari pembuatan sistem ini adalah melakukan pengelompokan berdasarkan tingkat kepadatan data penderita demam berdarah (berupa data koordinat UTM *-Universal Trasverse Mercator, peta dunia dalam bidang 2 dimensi-* yang diperoleh dari GPS) dengan menghitung jarak

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*.

3. Dasar Teori

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.
 - *MinPts*, adalah jumlah minimal titik sehingga sebuah cluster dapat terbentuk.
- b. DBSCAN memiliki 3 jenis tipe titik setelah proses *clusterisasi* dilakukan yaitu:
 - *Core Point*, adalah titik yang berada dalam suatu cluster dan memiliki titik tetangga (*eps-neighbourhood*) sedikitnya sebanyak *MinPts*.
 - *Border Point*, adalah titik yang berada dalam suatu cluster namaun jumlah titik tetangganya $< MinPts$.
 - *Noise Point*, adalah titik yang tidak termasuk anggota cluster manapun.

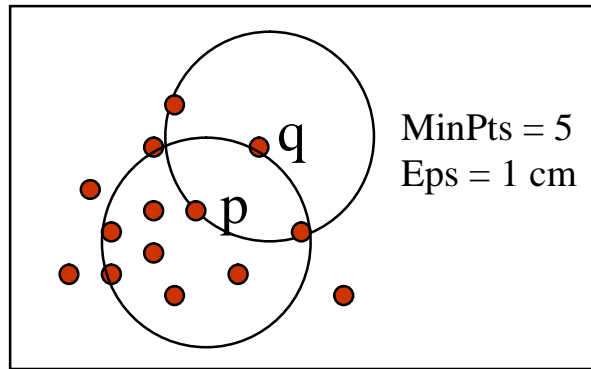


Gambar 1. Ilustrasi *core*, *border* Dan *noise*

- c. Sebuah titik *q* dikatakan tetangga dari titik *P* apabila

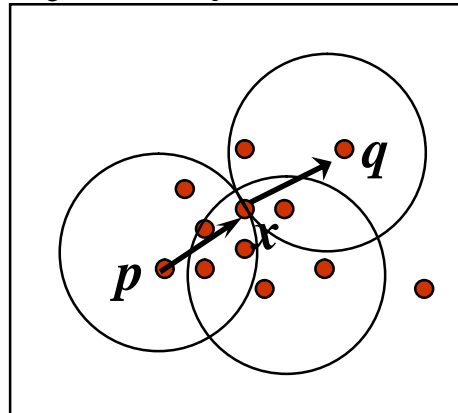
$$Neps\text{-neighbourhood}(p) = \{q \text{ belongs to } D \mid dist(p,q) \leq Eps\} \quad [2.2]$$

$$dist(p,q) = \text{jarak antara titik } p \text{ dan titik } q.$$
- 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)*
 - $Neps\text{-neighbourhood}(p) \geq MinPts$ atau dengan kata lain titik *p* merupakan *core point*.



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

- e. *Density-Reachable* merupakan sebuah titik yang terhubung secara tidak langsung dengan titik lain. Sebuah titik q dikatakan *density-reachable* titik p apabila ada suatu titik x yang menghubungkan antara p dan q dimana :
- titik x *directly density-reachable* dari titik p
 - titik q *directly density-reachable* titik x
 - jarak titik q dengan titik x < *Eps*

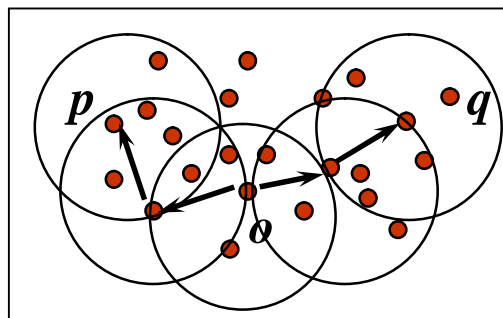


Gambar3.

Ilustrasi titik q *density-reachable* dari titik p

Titik q *density-reachable* dari titik p tapi p tidak *density reachable* dari titik q karena x tidak *directly-reachable* dari titik q.

- f. *Density-Connected* merupakan koneksi yang bersifat simetrik antara 2 titik. Titik q dikatakan *density-connected* titik p apabila ada sebuah titik o dimana titik p dan q *density reachable* dari titik o.

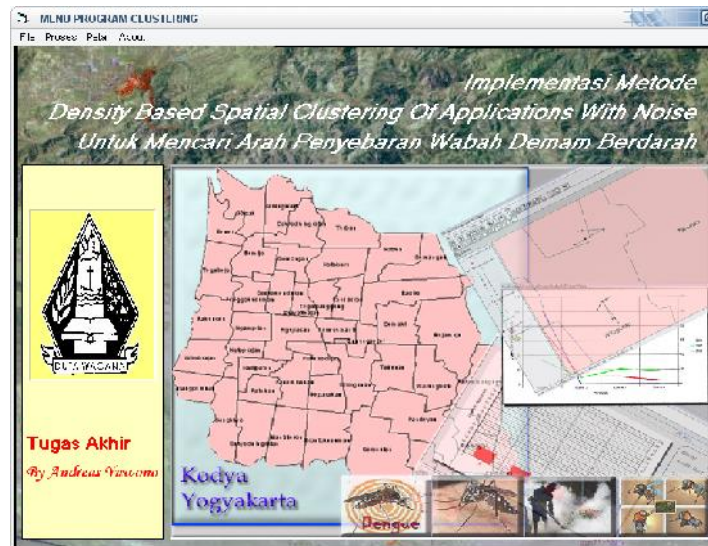


Gambar 4 titik q *density connected* dengan titik p

DBSCAN bekerja dengan cara melakukan pengumpulan data jarak secara iterasi dimulai titik pertama sampai titik terakhir, mengumpulkan jumlah titik-titik yang menjadi tetangga dengan batas jarak antar titik tetangga dengan titik acuan tidak boleh lebih dari *Eps*. Kemudian membandingkan jumlah titik tetangga tersebut dengan parameter Min Pts. Perbandingan ini dilakukan untuk menentukan tipe titik (core, border, noise) dan membentuk cluster baru. Cluster yang terbentuk akan memiliki tingkat kesamaan yang tinggi di dalam cluster itu sendiri (*High intra-class similarity*), dan tingkat kesamaan yang rendah antar cluster yang berbeda (*Low inter-class similarity*). (Kantardzic, 003).

4. Implementasi

Hasil yang dicapai dari implementasi sistem ini adalah sebuah program yang mampu menciptakan *cluster* dari suatu kumpulan data yang ada dengan parameter yang diinginkan. Kemudian menampilkan garis-garis pergerakan wabah DB pada Arc View GIS yang diharapkan dapat membantu mendeteksi arah penyebaran wabah demam berdarah. (O'Sullivan, 2003).



Gambar 1. Form Menu Utama

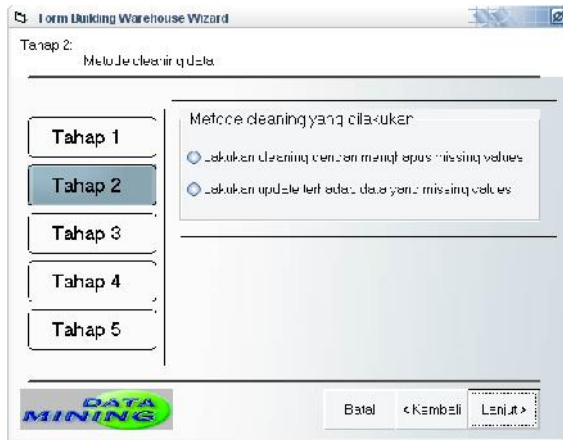
Sistem dibagi menjadi 3 bagian utama yaitu :

a. Sub Sistem Membangun Gudang Data:



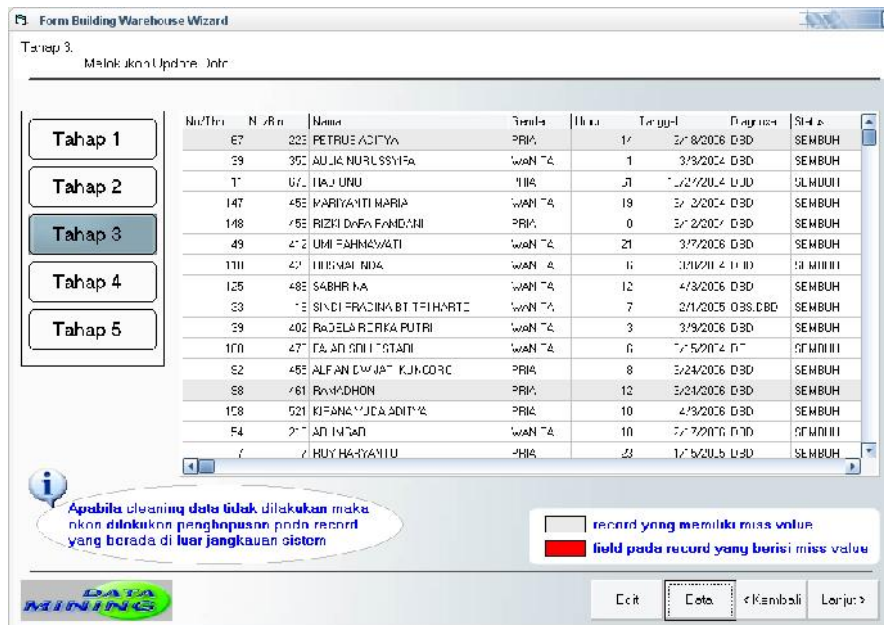
Tahap 1 :
Melakukan pemilihan file sumber data dalam kasus ini file excel.

Gambar 2. Form pemilihan file sumber data



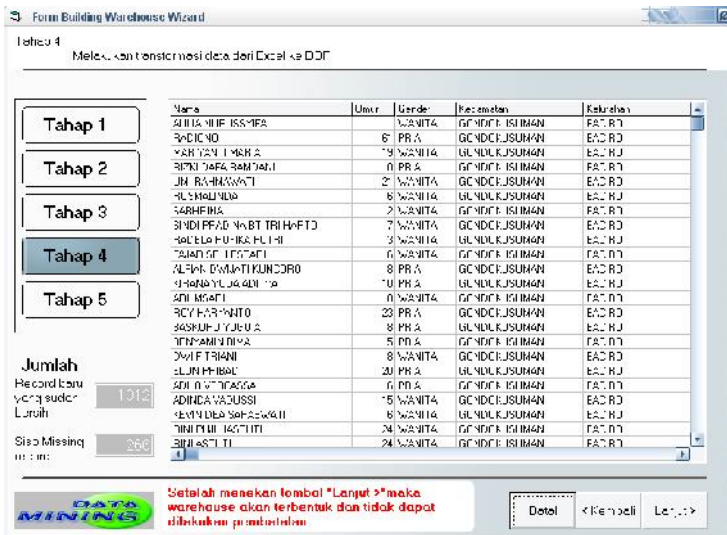
Tahap 2 :
Memilih metode cleaning yang digunakan:
- apakah akan mengupdate data yang kotor atau
- langsung menghapus data yang kotor

Gambar 3. Form pilihan untuk update data atau hapus data kotor



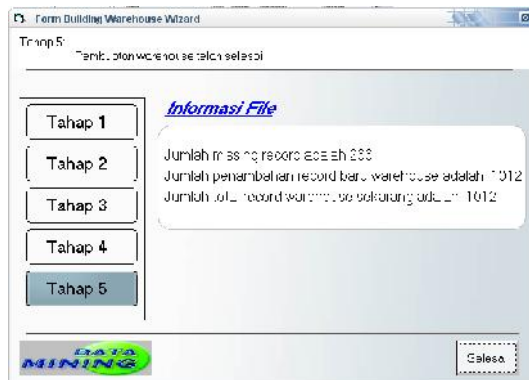
Tahap 3 :
Pada tahap 3 dilakukan proses pembersihan data sesuai dengan metode yang dipilih di tahap sebelumnya.

Gambar 4. Proses pembersihan data



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

Gambar 5. Pengecekan data yang bersih



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

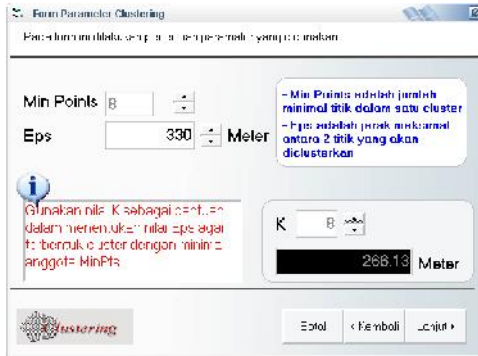
Gambar 6. Migrasi data ke gudang data

b. Sub Sistem Proses Clustering



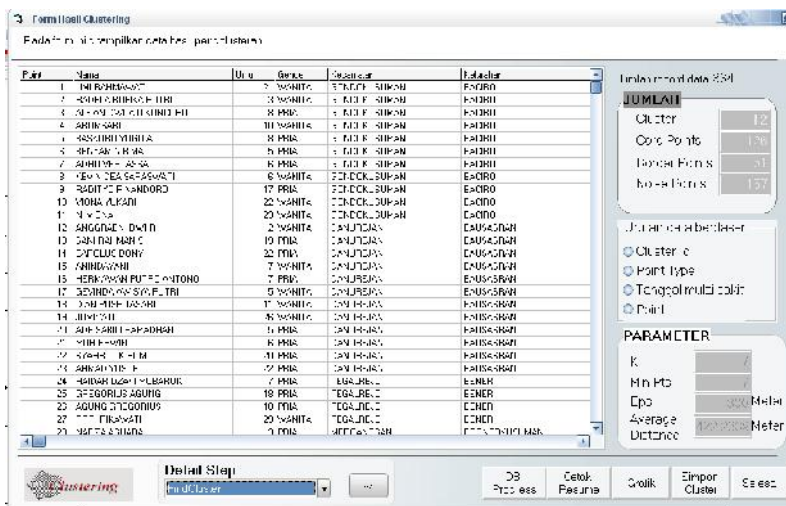
Tahap 1 :
Sistem akan melakukan seleksi data dari *warehouse*, sehingga tidak semua data diperlukan dalam proses clustering.

Gambar 7. Seleksi data dari warehouse



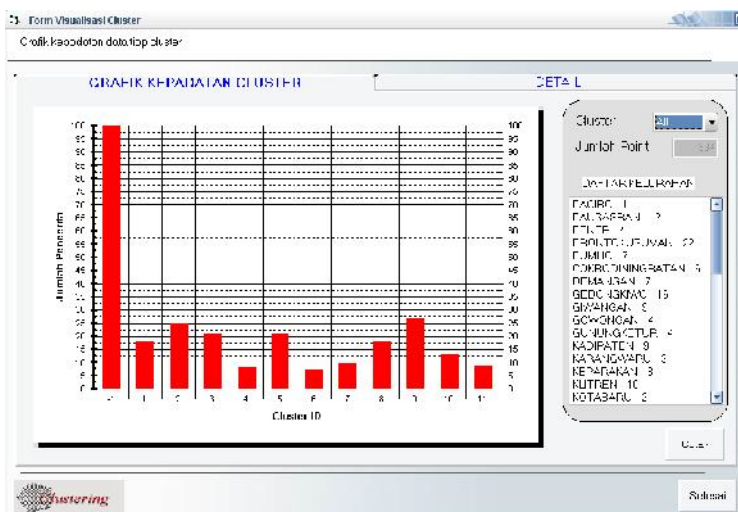
Tahap 2 :
 Dilakukan penentuan parameter yang akan digunakan.
 Meliputi
 - Min Pts
 - Eps
 - K

Gambar 8. Penentuan parameter



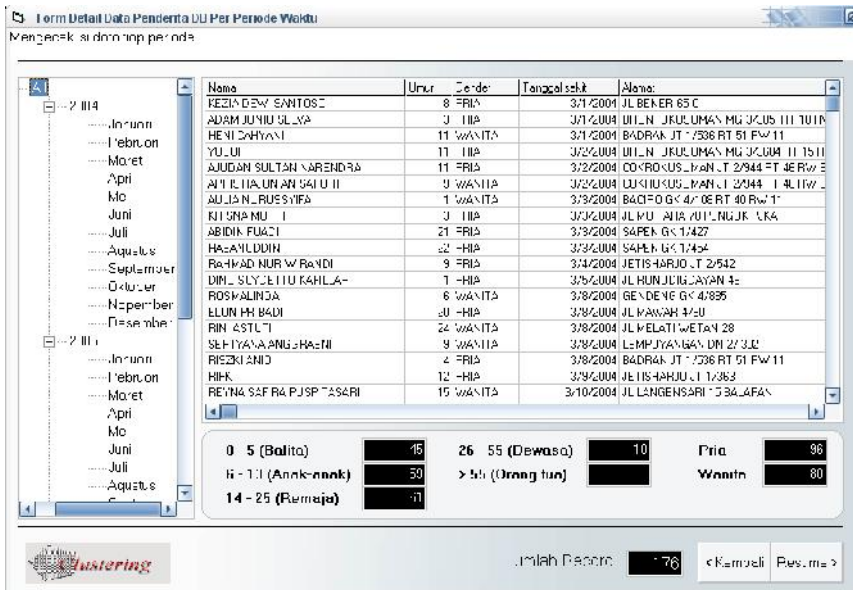
Tahap 3.1:
 Merupakan form yang berisi informasi hasil clustering

Gambar 9. Form informasi hasil clustering



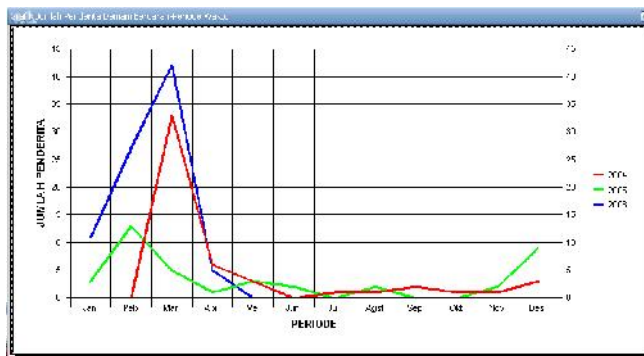
Tahap 3.2 : Form yang berisi informasi hasil cluster dalam bentuk grafik batang

Gambar 10. Form hasil cluster dalam bentuk grafik batang



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

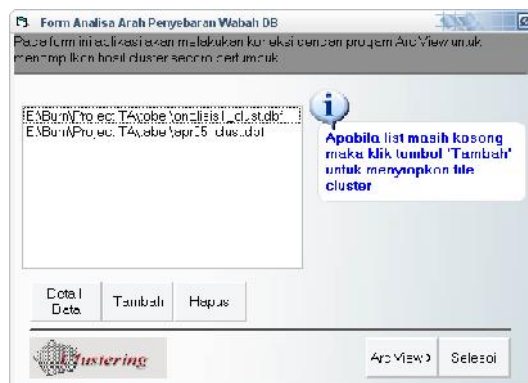
Gambar 11. Form informasi detail data hasil clustering dengan Tree



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

Gambar 12. Form visualisasi informasi waktu

c. Sub Sistem Visualisasi Hasil Clustering Dengan Arc View GIS



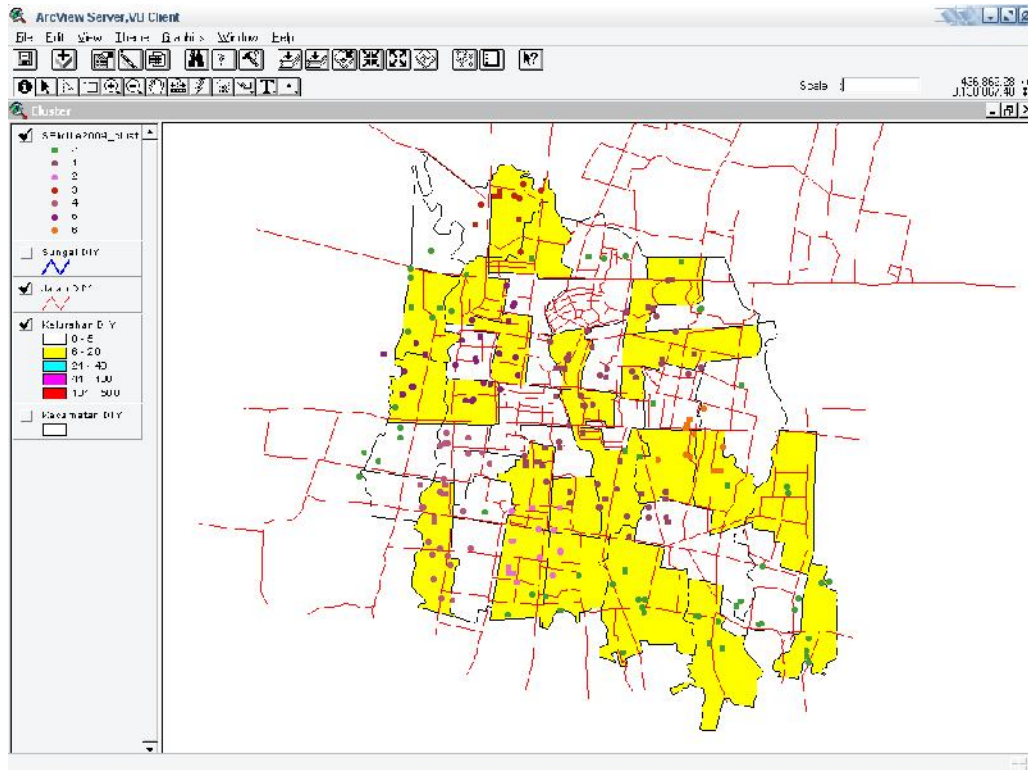
Sebelum melakukan koneksi dengan Arc View, dilakukan pemilihan file hasil clustering terlebih dahulu. Hasil Visualisasi ditunjukkan pada Gambar 1 dan 2 di bawah.

Gambar 13. Form pemilihan file hasil clustering

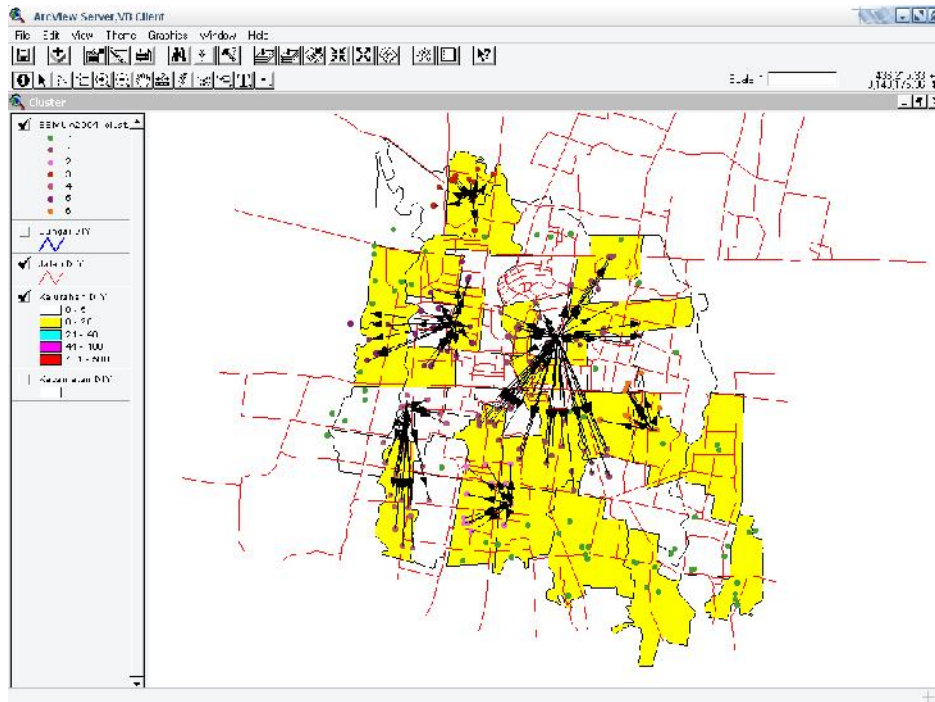


Merupakan form yang digunakan untuk mengontrol aplikasi Arc View GIS 3.3 melalui VisualBasic

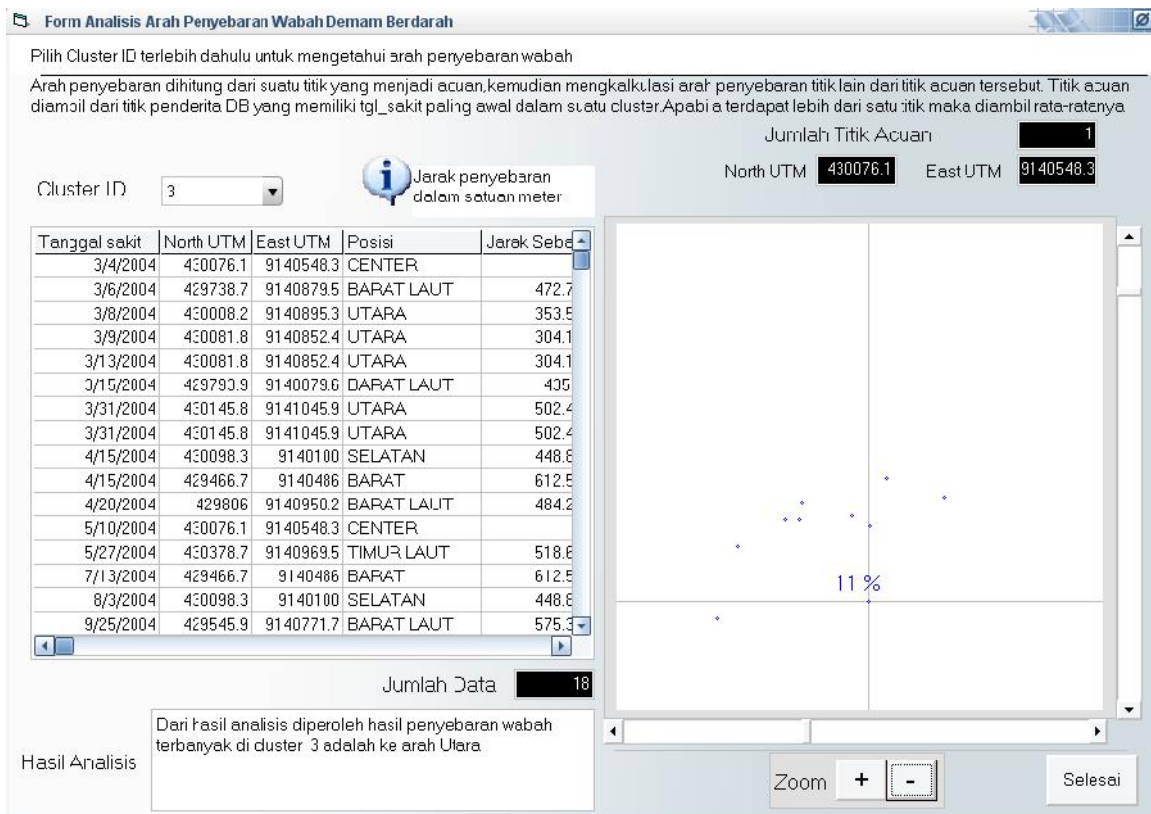
Gambar 14. Form kontrol aplikasi Arc View GIS 3.3



Gambar 15. Hasil Visualisasi Cluster dengan Arc View GIS 3.3



Gambar16. Hasil Pemetaan Arah Penyebaran Wabah DB dengan Arc View GIS 3.3



Gambar 17. Form Analisis Arah Penyebaran Wabah DB

5. Kesimpulan

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

- a. 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*.
- b. 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 *-MinPts-* dan seberapa jauh jarak maksimal antara 2 titik agar sebuah *cluster* diijinkan terbentuk *-Eps-*.
- c. Semakin besar nilai *Min Pts* yang diberikan maka cluster yang terbentuk semakin sedikit.
- d. Semakin besar nilai *Eps* maka cluster yang terbentuk semakin sedikit, begitu pula sebaliknya.
- e. 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*.
- f. 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.
- g. Wabah demam berdarah cenderung terjadi di bulan Januari, Pebruari dan Maret .

7. Daftar Pustaka

- Agushybana, 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
- Grizaite, Giedre dan Roland Innerhofer-Oberperfler, *DBSCAN Clustering Algorithm*. [http://www.inf.unibz.it/dis/teaching/DWDM05/reports/dbscan .pdf](http://www.inf.unibz.it/dis/teaching/DWDM05/reports/dbscan.pdf), 2006
- Han, Jiawei dan Micheline Kamber, *Data Mining : Concept and Techniques*. San Fransisco. Morgan Kauffman Publishers, 2001
- Kantardzic, Mehmed, *Data Mining : Concepts, Models, Methods, and Algorithms*. A John Willey & Sons, Inc., Publication., IEEE Press., 2003
- Kimball, Ralph dan Margy Ross, *The Data Warehouse Toolkit*. New York. John Wiley and Sons, 2nd edition, 2002
- Kusumo, Drs. Ario Suryo, *Buku Latihan Pemrograman Database Dengan Visual Basic 6.0*. Jakarta. Elex Media Komputindo, Cetakan Ketiga, September 2003
- Moreira, Adriano dan Carneiro, Sofia - Maribel Y.Santos, *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), 2006
- O'Sullivan, David dan Unwin, David, *Geographic Information Analysis*, Danvers: John Wiley and Sons, Inc., 2003 .
- Prahasta, Eddy, *Pemrograman Bahasa Script Avenue*, Bandung: Informatika, 2004.