

CLUSTERING DAN DATA MINING PADA GEODATABASE KEBENCANAAN GERAKAN TANAH DI KABUPATEN SUMEDANG

Afnindar Fakhurrozi¹ dan Sukristiyanti¹

¹Puslit Geoteknologi–LIPI, Jl. Sangkuriang Bandung 40145
E-mail: afnindar.f@gmail.com

Abstrak

Penelitian ini adalah pengembangan teknik clustering dan data mining pada geodatabase kebencanaan gerakan tanah di Kabupaten Sumedang. Tujuan dari kegiatan ini adalah memperoleh sebaran spasial tingkat kerentanan gerakan tanah di Kabupaten Sumedang untuk membantu upaya mitigasi bencana gerakan tanah. Metode clustering yang digunakan adalah algoritma pencarian tetangga terdekat (*K-Nearest Neighbor*) dan spasial predikat *ST_Contains*. Clustering dilakukan untuk semua entitas obyek cluster geodatabase yang memiliki keragaman jenis entitas data (poligon, segmen garis dan titik, spasial dan non-spasial) meliputi administrasi, geologi, jenis tanah, jalan, sungai, penggunaan lahan, elevasi dan pemukiman. Hasil clustering pada geodatabase dapat memberikan informasi-informasi penting terkait dengan upaya mitigasi bencana gerakan tanah. Hubungan relasional di antara obyek-obyek dalam geodatabase merupakan salah satu karakteristik penting yang harus diperhatikan dalam memilih metoda clustering untuk mendapatkan informasi yang benar.

Kata kunci: clustering, data mining, geodatabase kebencanaan

PENDAHULUAN

Gerakan tanah ialah pergerakan massa batuan/tanah karena tarikan gaya gravitasi. Sebanyak 17 dari 26 kecamatan di Kabupaten Sumedang berada pada zona rawan gerakan tanah menengah hingga tinggi. Data Pusat Vulkanologi dan Mitigasi Bencana Geologi (PVMBG) menyebutkan, di wilayah selatan, potensi longsor terjadi di Pamulihan, Rancakalong, Sumedang Selatan, dan Ganeas. Sedangkan di sisi utara, kerawanan longsor terjadi di Cimalaka, Paseh, Buah Dua, dan Conggeang. Banyak variasi data baik spasial maupun non spasial yang berbentuk *flat file* (*.dbf, *.shp, *.xls, dan sebagainya) yang mendeskripsikan kondisi fisik Kabupaten Sumedang, serta tingginya kejadian gerakan tanah di kabupaten tersebut menjadikan penyusunan geodatabase kebencanaan gerakan tanah penting dilakukan.

Geodatabase mengandung data (spasial dan non spasial) yang terstruktur dan berelasi. Namun dengan melakukan *clustering* pada geodatabase akan diperoleh informasi yang lebih berguna bagi pengguna. Clustering merupakan suatu metode analisis data dan teknik *data mining* yang populer untuk mengelompokkan obyek/data berdasarkan ukuran kedekatan dari yang lebih besar kedalam satuan data yang lebih kecil dan bergantung pada sejumlah data yang dianalisis.

Clustering dan data mining pada geodatabase dilakukan untuk mendapatkan potensi informasi yang berguna (Marinette et. al., 2009). Clustering merupakan analisis data serta teknik data mining yang populer (Tae-Wan dan Christoph, 2005). Data mining adalah ekstraksi informasi yang berguna dari data set yang besar atau database (Das, 2011). Tujuan clustering adalah untuk mengambil data dari database yang menyimpan informasi tipe obyek tertentu dan mengidentifikasi subgroup dari obyek tersebut seperti obyek yang termasuk dalam subgroup yang sama satu dengan yang lainnya, dan obyek yang berbeda dari subgroup satu dengan yang lainnya (Tae-Wan dan Christoph, 2005). Sedangkan spatial data mining merupakan upaya menemukan hubungan dan karakteristik yang ada secara implisit pada spatial database (Ng dan Han, 1994). Clustering yang dilakukan pada geodatabase kebencanaan gerakan tanah kabupaten sumedang berbeda dengan conventional clustering yang menggunakan flat data set tanpa membutuhkan generalisasi (Tae-Wan dan Christoph, 2005).

Penelitian terdahulu (Das, 2011) melakukan eksplorasi dan analisis data spasial gerakan tanah dengan menggunakan fungsi K dan G *Nearest Neighbour* serta mencari kemungkinan intensitas susceptibility gerakan longsor menggunakan Strauss Model. Disamping itu (Zhang dan Huang, 2007) melakukan clustering

data spasial dengan menggunakan ekstensi SQL untuk agregasi data spasial. Zhang dan Huang (2007) menggunakan predikat spasial dan fungsi spasial pada PostGIS dalam clustering database. Penelitian kami ini mengkombinasikan ide dari kedua penelitian terdahulu yaitu dengan melakukan clustering pada data spasial gerakan tanah, namun dengan penambahan dimana clustering dilakukan pada geodatabase gerakan tanah.

TUJUAN PENELITIAN

Tujuan penelitian ini adalah melakukan clustering pada geodatabase untuk memperoleh informasi-informasi penting berkaitan dengan kerentanan gerakan tanah di Kabupaten Sumedang yang akan membantu, misalnya untuk mengidentifikasi lokasi-lokasi relokasi penduduk dari daerah rawan gerakan tanah di Kabupaten Sumedang.

METODOLOGI

Clustering geodatabase kebencanaan gerakan tanah Kabupaten Sumedang menggunakan beberapa metode analisis geospasial antara lain analisis pola titik, analisis jaringan, geostatistik, dan operasi multilayer. Perangkat lunak yang digunakan berbasis *free* dan *open source* (FOSS = Free and Open Source Software), sehingga kami dapat menggunakan perangkat lunak tersebut dan mengkustomisasi bahasa pemrograman (*script*) atau bahkan modul perangkat lunaknya dengan bebas dan legal.

Clustering pada semua entitas obyek cluster dalam geodatabase dilakukan terhadap entitas kejadian gerakan tanah di Kabupaten Sumedang. Ada dua macam metoda yang digunakan dalam clustering, yaitu:

1. clustering pada entitas (obyek cluster) administrasi, geologi, dan jenis tanah menggunakan spasial predikat *ST_Contains* pada PostGIS,
2. clustering entitas jalan, sungai, penggunaan lahan, elevasi, dan permukiman menggunakan K-Nearest Neighbor (algoritma euclidean dalam mencari jarak antar dua titik yang terdekat) dengan memanfaatkan fitur spasial agregat *ST_DWithin*.

Tahapan pelaksanaan proses clustering digambarkan dalam diagram alir Gambar 1.

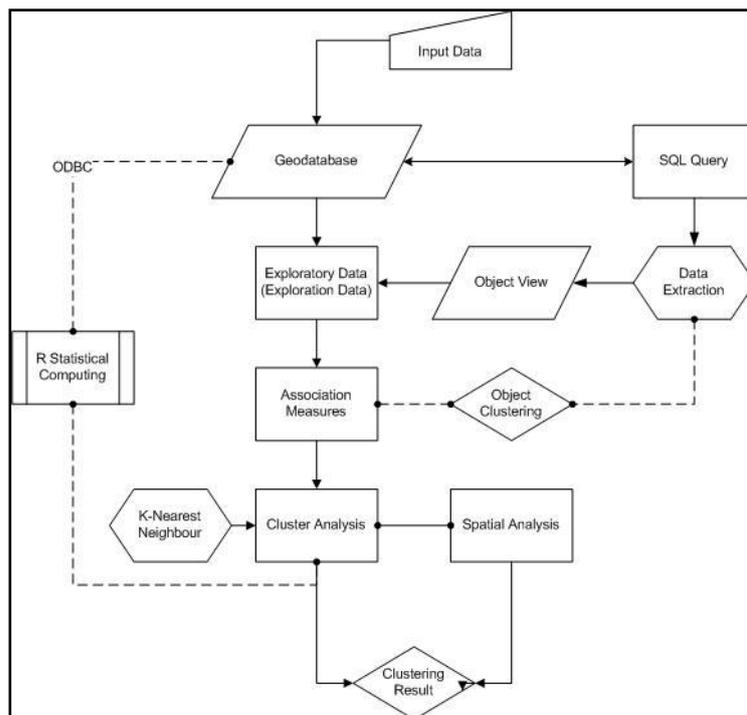
Data yang digunakan antara lain:

1. Peta Rupa Bumi Indonesia skala 1:25000, terdiri atas layer-layer jalan, sungai, kontur, penggunaan lahan administrasi, dan layer toponimi
2. Peta Geologi
3. Peta Jenis Tanah
4. Peta lokasi longsor (hasil survai lapangan)

Perangkat lunak yang digunakan antara lain:

1. PostgreSQL
2. PostGIS
3. pgAdmin
4. Quantum GIS
5. Bahasa R

Semua perangkat lunak yang digunakan dalam penelitian ini adalah berbasis Open Source.

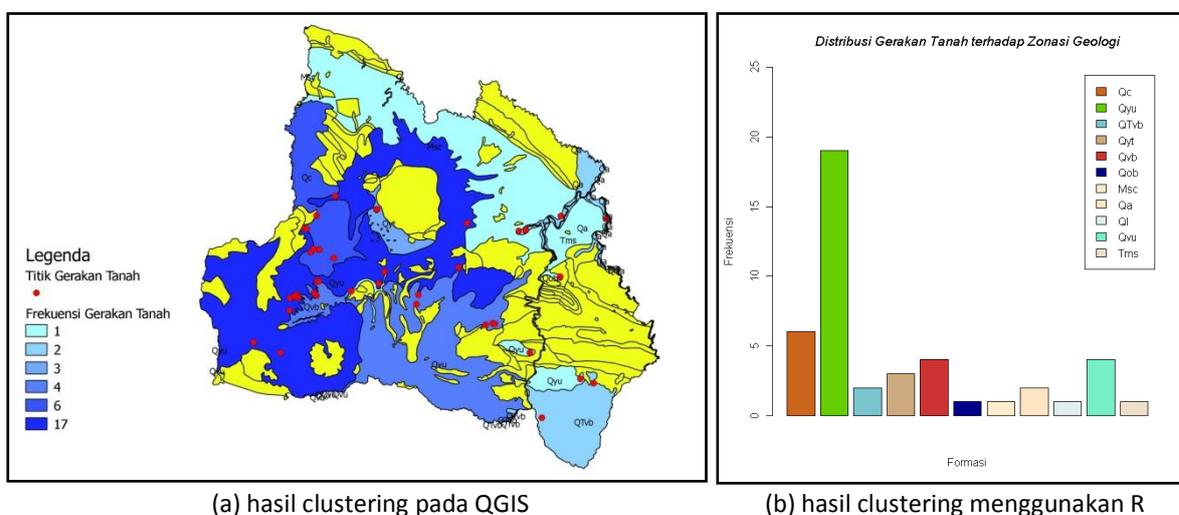


Gambar 1. Tahap-tahap Penelitian (dimodifikasi dari Borcard et. al., 2011).

HASIL

Hasil penelitian merupakan hasil clustering pada geodatabase gerakan tanah Kabupaten Sumedang untuk semua entitas obyek cluster dari database yang dibuat, meliputi administrasi, geologi, jenis tanah, jalan, sungai, penggunaan lahan, elevasi, dan permukiman.

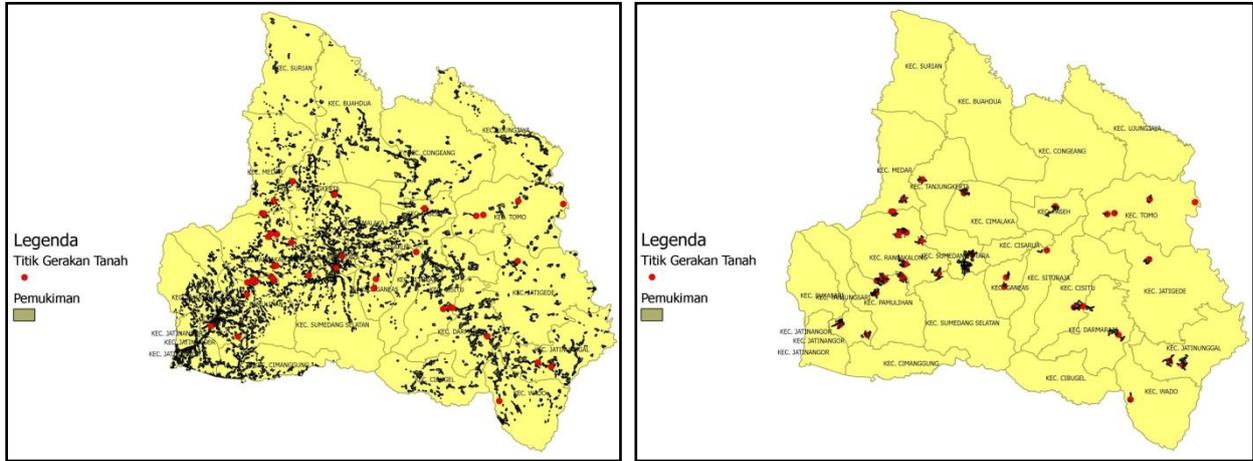
Gambar 2 adalah hasil clustering pada entitas obyek cluster geologi dengan menggunakan spasial predikat ST_Contains. Hasil clustering pada QGIS dapat divisualisasikan secara spasial, sedangkan hasil clustering menggunakan R divisualisasikan dalam bentuk grafik batang. Keduanya saling melengkapi dalam mempresentasikan hasil clustering.



Gambar 2. Hasil clustering pada entitas obyek cluster geologi.

Gambar 3a dan 3b secara berturut-turut menunjukkan tampilan entitas obyek cluster permukiman sebelum dan sesudah dilakukan clustering. Hasil clustering pada entitas obyek cluster permukiman memberikan

informasi mengenai sebaran obyek-obyek permukiman diasumsikan terkena dampak gerakan tanah. Asumsi yang dibuat adalah jarak 500 m dari titik-titik kejadian gerakan tanah akan mendapatkan dampak. Clustering ini dilakukan dengan menggunakan K-Nearest Neighbor dengan memanfaatkan fitur spasial agregat ST_DWithin.

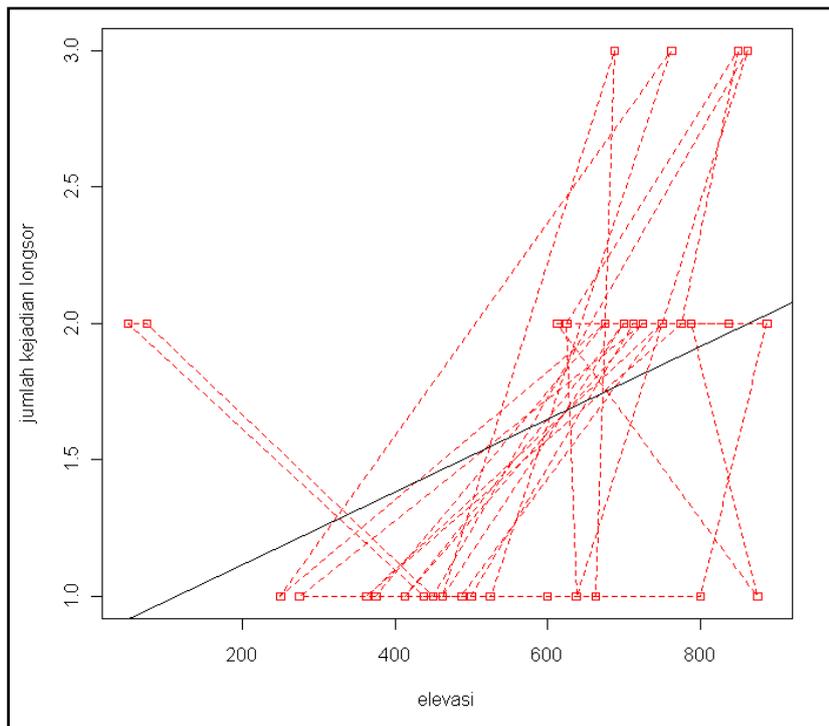


a. sebelum clustering

b. sesudah clustering

Gambar 3. Peta sebaran permukiman dan titik gerakan tanah di Kabupaten Sumedang.

Gambar 4 menyajikan hasil clustering pada entitas obyek cluster elevasi dengan menggunakan K-Nearest Neighbor dengan memanfaatkan fitur spasial agregat ST_DWithin. Hasil clustering ini disajikan dalam bentuk grafik pencar elevasi (sumbu x) versus kejadian gerakan tanah (sumbu y) yang tidak memberikan hasil yang informatif.



Gambar 4. Hasil clustering pada entitas obyek cluster elevasi.

ANALISIS/DISKUSI

Beberapa entitas obyek cluster pada geodatabase dapat dilakukan clustering dengan menggunakan spasial predikat ST_Contains dan sebagian yang lain dengan menggunakan K-Nearest Neighbor, dan memberikan hasil clustering yang informatif. Hal ini tidak untuk clustering pada entitas obyek cluster elevasi. Kegagalan yang terjadi pada clustering entitas obyek cluster elevasi yang berupa garis kontur adalah tidak semua wilayah diketahui nilai elevasinya sehingga menyulitkan proses clustering. Wilayah-wilayah yang tidak dilalui garis kontur merupakan wilayah-wilayah yang tidak memiliki nilai elevasi karena pada proses clustering tidak dilakukan proses interpolasi nilai elevasi. Dengan demikian, clustering pada entitas obyek cluster elevasi tidak dapat dilakukan pada elevasi dalam format vektor yang berupa polyline. Entitas obyek cluster elevasi harus diubah dahulu ke format raster, baik berupa aspek maupun kemiringan lereng, sehingga semua wilayah yang dilakukan clustering memiliki nilai elevasi masing-masing dan tidak mengganggu dalam proses clustering.

Metode ST_Contains hanya dapat diterapkan untuk clustering pada entitas obyek cluster yang berupa vektor polygon. Pada geodatabase yang dibangun, clustering dengan metode ST_Contains dilakukan pada entitas obyek cluster geologi, litologi, dan administrasi kecamatan. Hal ini berbeda dengan clustering pada entitas obyek cluster penggunaan lahan. Meskipun obyek cluster penggunaan lahan merupakan feature polygon, obyek cluster ini menggunakan metode clustering K-Nearest Neighbor. Hal ini dikarenakan ada asumsi yang dibuat pada saat eksplorasi data, dimana kejadian gerakan tanah tidak dipengaruhi hanya oleh 1 jenis penggunaan lahan saja. Disamping itu, obyek-obyek cluster lain yang berupa feature polygon (geologi, litologi, dan administrasi kecamatan), masing-masing satuannya terdapat lebih dari 1 kejadian longsor. Hal ini berbeda dengan obyek cluster penggunaan lahan, dimana 1 kejadian longsor dipengaruhi oleh banyak satuan lahan penggunaan lahan. Dengan metode clustering K-Nearest Neighbor, dapat diketahui tipe-tipe penggunaan lahan yang mempengaruhi dan terkena dampak kejadian-kejadian gerakan tanah.

Dibandingkan dengan clustering pada flat file, clustering pada geodatabase lebih menemui kendala, antara lain: (1) geodatabase mengandung obyek spasial dan non spatial sehingga sulit mendefinisikan obyek cluster, (2) geodatabase mempunyai sifat relasional diantara obyek-obyeknya, (3) kendala dalam ekstraksi, akses, dan manipulasi pada obyek spasial yang tersimpan dalam geodatabase, (4) kendala dalam pengujian kesamaan (*similarity*) dan ketidaksamaan (*dissimilarity*) pada obyek spasial dalam geodatabase.

Asumsi yang dibangun dalam clustering dengan menggunakan metode K-Nearest Neighbor adalah jarak 500 m, dimana jarak 500 m dari titik-titik kejadian gerakan tanah akan mendapatkan dampak. Asumsi tersebut tidak berdasarkan referensi apapun, hanya contoh jarak yang digunakan dalam clustering ini. Dalam penelitian lanjut untuk tujuan mitigasi bencana, asumsi jarak dapat disesuaikan berdasarkan kondisi yang sebenarnya di lapangan.

Clustering pada entitas obyek cluster administrasi, geologi, dan jenis tanah memberikan informasi mengenai frekuensi kejadian gerakan tanah pada masing-masing administrasi kecamatan dan satuan lahan geologi dan jenis tanah. Informasi ini penting dalam mitigasi bencana gerakan tanah dalam rangka pemberian perhatian yang lebih dan pengurangan resiko bencana pada wilayah-wilayah administrasi kecamatan, satuan lahan geologi, dan jenis tanah yang memiliki frekuensi tinggi terjadinya gerakan tanah.

Clustering pada entitas obyek cluster jalan, sungai, penggunaan jalan, dan permukiman memberikan informasi mengenai obyek-obyek tersebut yang terkena dampak dan atau mempengaruhi kejadian gerakan tanah dalam radius 500 m, sesuai dengan asumsi yang digunakan dalam clustering. Informasi ini penting untuk proses evakuasi, rehabilitasi, pembuatan kebijakan dalam relokasi, dan juga dalam analisis lebih lanjut mengenai efek keberadaan sungai terhadap kejadian gerakan tanah.

KESIMPULAN

Clustering pada geodatabase gerakan tanah dapat menghasilkan informasi penting (data mining) yang bermanfaat untuk membantu mengurangi resiko bencana (mitigasi). Tingkat kesulitan clustering pada geodatabase lebih tinggi daripada clustering pada database biasa (flat data) karena obyek data pada geodatabase dapat berupa data spasial dan non-spasial, dan di antara obyek-obyek datanya dapat memiliki hubungan relasional. Hubungan relasional di antara obyek-obyek dalam geodatabase merupakan salah satu

karakteristik penting yang harus diperhatikan karena mempengaruhi pemilihan metoda clustering untuk mendapatkan informasi yang benar. Beberapa obyek geodata harus melalui pra-processing untuk dapat menghasilkan informasi yang diharapkan, misalnya untuk obyek elevasi perlu mengkonversi data kontur poligon ke bentuk raster.

UCAPAN TERIMAKASIH

Ucapan terimakasih kepada Dedi Mulyadi, MT selaku ketua tim Sumedang, kepada Hilda Lestiana, S.Si, Andarta F. Khoir, S.Kom, Bp. Adde Tatang, dan Bp. Firmansyah selaku sesama anggota tim.

DAFTAR PUSTAKA

- Borcard, D., Gillet, F., Legendre, P., 2011. *Numerical Ecology with R*. Springer Science and Bussiness. New York, USA.
- Das, C.I., 2011. *Spatial Statistical Modelling For Assessing Landslide Hazzard and Vulnerability. Dissertation*. ITC, University Of Twente.
- Marinette, P., Gancarski, P., Auefere, M.A., Boussaid, O., 2009. *Database Technologies : Concept, Methodologies, Tools and Application, Chapter 1.5 Pattern Mining and Clustering on Image Databases*. IGI Global. New York, USA.
- Ng, R.T., Han J., 1994. *Efficient and Effective Clustering Methods for Spatial Data Mining*. Proceedings of the 20th VLDB Conference. Santiago, Chile.
- Tae-Wan, R., Christoph, F., Eick., 2005. *A database clustering methodology and tool*, Information Sciences-Informatics and Computer Science: An International Journal, v.171 n.1-3, p.29-59.
- Zhang, C., Huang, Y., 2007. *Cluster By: A New SQL Extension for Spatial Data Aggregation*. GIS '07 Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems. ACM. New York, USA.