

IMPLEMENTASI DENSITY BASED SPATIAL CLUSTERING APPLICATION WITH NOISE (DBSCAN) DALAM PERKIRAAN TERJADI BANJIR DI BANDUNG

Bima Aryo Putro¹, Kemas Rahmat Saleh Wiharja², Shaufiah³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

¹aryo.241090@gmail.com

Abstrak

Klaster adalah pengelompokan objek-objek yang saling memiliki kedekatan karakteristik dan memiliki aturan tertentu. Saat ini teknik-teknik klasterisasi telah banyak di implementasikan, misal pada bidang biologi, iklim, kesehatan dan sebagainya. Terdapat banyak teknik klasterisasi diantaranya Hierarchical Clustering, Partitioning Clustering, Density based clustering.

Teknik clustering yang dipakai pada penelitian TA ini adalah density based clustering dengan jenis DBSCAN (Density Based Spatial Clustering Application with Noise). DBSCAN adalah salah satu jenis density based clustering dengan karakteristik dapat menangani noise, dapat menangani bentuk dari data yang cukup bervariasi (arbitrary shape), tidak menciptakan complete clustering, dan cocok digunakan untuk data dengan jumlah besar.

Banjir menjadi satu masalah yang belum tertuntaskan, parameter penyebab terjadinya banjir yang dapat diukur secara numerik diantaranya curah hujan, temperatur dan tekanan udara yang dapat diukur setiap hari, parameter - parameter tersebut didasarkan pada petunjuk dari BMKG Kota Bandung. DBSCAN dapat memperlihatkan pola banjir yang terjadi di Kota Bandung dan sekitarnya (Bandung Selatan), pola banjir yang diperlihatkan berupa bulan yang memiliki potensi banjir, beserta pola parameter - parameter potensi banjir yang ditentukan diatas.

Kata Kunci : klaster, DBSCAN, banjir.

Abstract

Cluster is grouping object which has a similar characteristic and has a specific type. Nowadays clustering technique has implemented in many aspect such in biologys , climate , pharmacy , etc. There are many clustering techniques such as Hierarchical clustering, Partitioning Clustering, Density based Clustering.

Clustering Technique that writer use in this paper is density based clustering with specified in DBSCAN (Density Based Spatial Clustering Application with Noise). DBSCAN is one of density based clustering technique which has characteristics, they are can handle noise , can handle data with various type, doesn't make a complete clustering, suitable for high density data.

Flood is one of problem that not yet to be solve, the parameters of flood that we could count numerical are rain density, temperatures , and air pressure in particular month. DBSCAN could show the pattern of flood in Bandung City and the area around it (South Bandung). The pattern that shown here are the month with possibility of flood and the parameters.

Keywords : cluster, DBSCAN, flood.

1. Pendahuluan

1.1 Latar belakang

Klaster adalah pengelompokan objek-objek yang saling memiliki kedekatan karakteristik dan memiliki aturan tertentu^[3]. Dalam kehidupan sehari-hari teknik klasterisasi telah banyak di implementasikan, contohnya pada bidang biologi, iklim, kesehatan dan sebagainya^[3].

Terdapat banyak jenis teknik klasterisasi diantaranya *hierarchical clustering*, *partitioning clustering*, *density based clustering*, dan sebagainya^[3]. Klasterisasi dapat membantu dalam proses perolehan informasi/pengetahuan (*knowledge gained*) dari suatu data.

Karakteristik dari data adalah tidak dapat berarti apapun jika tidak dilakukan analisis atau pengolahan, didalam suatu kumpulan data dimungkinkan terdapat pencilan. Nilai pencilan ini dapat mempengaruhi informasi dari data – data yang lainnya. Oleh karena itu diperlukan teknik klasterisasi yang dapat menangani pencilan – pencilan tersebut.

Sebagai contoh terdapat 5 orang anak yang memiliki nilai, dengan rincian 4 orang mendapat nilai 50 dan 1 orang mendapat nilai 100, tentunya anak yang mendapat nilai 100 merupakan *noise* / pencilan bagi anak lainnya. Tentunya saat proses *knowledge gained* dilakukan maka nilai yang dihasilkan tidak mencerminkan keadaan sebenarnya.

Klaster berbasis densitas memiliki karakteristik yaitu terdapat klaster yang padat dan dikelilingi oleh klaster / bagian lain yang tidak padat^[3]. Teknik klasterisasi berbasis densitas memiliki kelebihan yaitu dapat menangani klaster dari bentuk yang tidak tetap (*arbitrary shape*), dapat menangani *noise* atau data pencilan, dapat menangani data dalam jumlah besar dengan waktu yang lebih cepat, serta memiliki akurasi *clustering* yang lebih baik dibanding algoritma *clustering hierarchy* dan *partitioning*^{[9][7]}.

DBSCAN (Density Based Spatial Clustering Application with Noise) adalah salah satu teknik *clustering* berbasis densitas, yang memanfaatkan fungsi lokal konektivitas yang memanfaatkan parameter Eps dan MinPts serta memiliki karakteristik dapat menangani *noise* dan menangani bentuk dari data yang cukup bervariasi^[7]. DBSCAN tidak menciptakan *complete clustering*, hal ini dikarenakan DBSCAN hanya memperdulikan bagian dengan kepadatan yang tinggi^[9], sedangkan bagian lainnya di golongan sebagai *border* dan *noise*.

Faktor penyebab terjadinya banjir dapat digolongkan menjadi 3 bagian, yaitu faktor global, faktor lokal, dan faktor regional. Faktor global adalah faktor iklim secara universal, faktor lokal adalah faktor topologi daerah tersebut, faktor regional adalah faktor iklim di daerah tersebut. Dari ketiga jenis faktor tersebut, dalam penelitian ini adalah faktor regional dipilih sebagai parameter.

Berdasarkan arahan BMKG Kota Bandung, terdapat 3 unsur klimatologi yang bisa dijadikan parameter yang dapat digunakan sebagai parameter banjir pada suatu daerah. Yaitu tingkat curah hujan, temperatur, tekanan udara yang dapat diukur setiap harinya yang terjadi di daerah tersebut. Untuk mendapatkan pola banjir tersebut, digunakanlah DBSCAN.

Bandung adalah salah satu kota di Jawa Barat dengan intensitas curah hujan yang tinggi. Akhir-akhir ini sering terjadi banjir di wilayah Bandung, terutama di daerah Bandung selatan. Dari penelitian ini akan dibangun sistem yang diharapkan dapat pola banjir (potensi banjir pada bulan tersebut dan parameter-parameter) di daerah Bandung serta (Bandung selatan) berdasarkan parameter-parameter yang telah ditentukan.

1.2 Perumusan masalah

Dari penjelasan di atas disimpulkan permasalahan pokok diantaranya:

1. Bagaimana menerapkan metode *density based clustering* DBSCAN untuk menciptakan klusterisasi pola banjir berdasarkan parameter-parameter yang telah ditentukan.
2. Bagaimana melakukan validasi dan analisis kluster yang baik guna memastikan bahwa kluster yang dibentuk merepresentasikan pola banjir sebenarnya.

Batasan Masalah

Sedangkan batasan dari TA adalah:

1. Parameter yang menjadi ukuran adalah curah hujan, temperatur, dan tekanan udara.
2. Data klimatologi tahun 2006-2009 yang menjadi masukan adalah curah hujan perhari, temperatur, dan tekanan udara.
3. Pola banjir yang dilakukan analisis adalah bulan yang memiliki potensi banjir dan tidak banjir, beserta parameter-parameter.
4. Status potensi terjadi banjir dibagi atas 2 bagian yaitu banjir dan tidak banjir dimana akan terdapat beberapa profil didalam masing-masing kelas.
5. Keterangan lapangan didapatkan dari DPSDA Kota Bandung, PMI IT TELKOM, dan media online.
6. Sistem hanya dapat memperlihatkan pola banjir berdasarkan parameter yang dimiliki.

1.3 Tujuan

Tujuan penelitian ini adalah:

1. Menerapkan metode *density based clustering* DBSCAN untuk menghasilkan klusterisasi pola banjir dari parameter penyebab banjir di Bandung serta (Bandung Selatan) akibat parameter yang ditentukan.
2. Membuat klusterisasi dengan tingkat validasi dan analisis yang baik.

Hipotesa : Dengan Algoritma DBSCAN, teknik validasi entropy dan presisi serta penentuan parameter banjir (curah hujan, temperatur, dan tekanan udara), dapat memperlihatkan pola banjir yang terjadi di Bandung.

1.4 Metodologi penyelesaian masalah

- a. Studi literatur
Tahap menambah wawasan dari buku-buku, artikel, jurnal dan sumber-sumber lain yang layak, seperti informasi-informasi yang tersedia di internet. Hal tersebut meliputi diantaranya:
 - a. Mempelajari tentang teknik *density based clustering* DBSCAN.
 - b. Mempelajari implementasi teknik *density based clustering* DBSCAN.
 - c. Mempelajari implementasi validasi kluster entropy dan presisi.
 - d. Pencarian data parameter dengan meminta permohonan data ke Badan meteorologi dan geofisika (BMKG) Bandung, PSDA Kota Bandung.
- b. Pemodelan sistem
Pemodelan sistem yang digunakan adalah melakukan *preprocessing* pada data, mengkonversikannya kedalam file excel lalu mengimplementasikan rumus DBSCAN dalam PHP, Javascript, MySQL untuk proses klusterisasi.
- c. *Testing* dan Analisis Hasil
Testing yang dilakukan adalah:
 1. Pengujian dilakukan dengan 2 bentuk yaitu *whitebox testing* dan *blackbox testing*.
Whitebox testing dilakukan melihat pengaruh fungsi-fungsi utama terhadap sistem dengan dilakukan 3 skenario. Skenario 1 (Menghilangkan `SetOfPoints.changeClid(Point,Noise)`) Skenario 2 (Menghilangkan `if result.size>=MinPts then` beserta syntax didalamnya sampai akhir `if`) Skenario 3 (Menghilangkan `if seeds.size> minpts` beserta syntax lain didalamnya).
Blackbox testing dilakukan guna membandingkan hasil klusterisasi DBSCAN, kaitannya dengan pola banjir, serta dilakukan validasi kluster dengan teknik *supervised validation* (entropy, presisi).
 2. Perhitungan akurasi didasarkan dengan membandingkan potensi banjir yang dihasilkan sistem dengan fakta dilapangan.
 3. Perhitungan akurasi dihitung dengan *Jaccard Coefficient*.

Telkom
University

5. Kesimpulan dan Saran

5.1 Kesimpulan

1. Berdasarkan hasil penelitian percobaan keempat dimana memiliki pola klasterisasi paling baik dihasilkan 9 buah klaster dari pola banjir yang terjadi di Bandung, hal ini dikarenakan data banjir tahun 2006-2009 yang digunakan sebagai data latih memiliki profil-profil tertentu meski masih termasuk kedalam 1 jenis kelas banjir atau tidak banjir, dengan DBSCAN profil – profil tersebut dapat terlihat sehingga dapat membantu ahli memahami pola banjir yang terjadi.
2. DBSCAN mampu memperlihatkan pola banjir (bulan dan parameter - parameter nya). Namun, DBSCAN kurang cocok digunakan untuk melakukan prediksi hal ini terbukti pada *blackbox testing* yang dibandingkan hasil fakta dilapangan DBSCAN memperoleh akurasi 80.6%
3. Parameter-parameter yang digunakan dalam penelitian ini yaitu curah hujan, temperatur dan tekanan udara dapat digunakan untuk memperlihatkan pola banjir yang terjadi di Bandung dari sisi klimatologi.

5.2 Saran

Di masa mendatang, untuk memperoleh hasil klasterisasi dan penggambaran pola yang lebih baik, perlu diperhatikan faktor – faktor lain yang turut berpengaruh terhadap terjadinya banjir. Perlu adanya metode lain untuk mengkalkulasi / menghitung perkiraan lamanya terjadi banjir, sehingga akan meningkatkan akurasi dari sistem.

Pengembangan lebih lanjut dalam proses klasterisasi bisa menggunakan algoritma RDBC^[6] atau algoritma pengembangan DBSCAN lainnya, guna menghemat waktu proses untuk data yang lebih banyak.

Telkom
University