LONTAR KOMPUTER VOL. 6, NO.3, DESEMBER 2015 **p-ISSN 2088-1541**
DOI: 10.24843/LKJITI.2015.v06.i03.p05 **e-ISSN 2541-5832**

Implementasi Metode *Clustering* DBSCAN pada Proses Pengambilan Keputusan

Ni Made Anindya Santika Devi[1], I Ketut Gede Darma Putra[2], I Made Sukarsa[3]

Jurusan Teknologi Informasi, Universitas Udayana
Bukit Jimbaran, Bali, Indonesia,
[1]anindyasande@gmail.com
[2] ikgdarmaputra@gmail.com
[3]e_arsa@yahoo.com

Abstrak

Spatial Data Clustering merupakan salah satu teknik penting pada data mining yang digunakan untuk mendapatkan informasi atau pengetahuan pada data spasial dalam jumlah yang besar dari berbagai aplikasi [1]. Salah satu teknik yang menjadi pelopor perkembangan algoritma clustering pada data spasial adalah DBSCAN. Teknik ini dapat menentukan cluster dari bentuk data yang tidak beraturan dan dapat menangani noise secara efektif. Penelitian ini berfokus pada pengimplementasian Metode DBSCAN pada proses pengambilan keputusan untuk membantu perusahaan menentukan pelanggan potensialnya[1]. Hasil uji coba pada penelitian ini menunjukkan bahwa Metode DBSCAN telah berhasil melakukan proses clustering untuk membantu proses pengambilan keputusan dalam penentuan pelanggan potensial dengan membentuk sejumlah cluster.

Kata kunci: *Clustering, Data Mining,* DBSCAN, Data Spasial, Pengambilan Keputusan.

Abstract

Spatial Data Clustering is one of the significant techniques in data mining which used to obtain information or knowledge in a large number of spatial data from various applications. One technique that being a pioneer in the development of spatial data clustering algorithm is DBSCAN. This technique can determine cluster of irregular data shape and can handle the noise effectively. This study is focused on implementation of DBSCAN method in decision making process in order to help a company to decide its potential customer. The trial results in this study show that DBSCAN method has been successfully conduct clustering process to support decision making process in determination of potential customer by forming several number of clusters.

Keywords: *Clustering, Data Mining, DBSCAN, Spatial Data, Decision Making*.

1. Pendahuluan

Data mining merupakan sebuah langkah dalam proses *Knowledge Discovery in Database* (KDD) yang terdiri dari penerapan analisis data dan penemuan algoritma yang menghasilkan enumerasi tertentu terhadap pola pada data [1]. *Spatial Data Mining* adalah bagian dari *data mining* yang merupakan proses menemukan pola yang menarik dan sebelumnya tidak dikenal tetapi secara potensial dapat berguna dari *dataset* spasial yang besar. Penggalian pola yang menarik dan berguna dari *dataset* spasial lebih sulit daripada penggalian pola data numerik tradisional dan kategorikal dikarenakan oleh kompleksitas jenis, hubungan dan autokorelasi dari *dataset* spasial tersebut [2].

Sebagian besar penelitian terbaru pada data spasial menggunakan teknik *clustering* dikarenakan oleh sifat dari data tersebut. *Clustering* merupakan proses pengelompokan sejumlah besar data menjadi beberapa kelas sesuai dengan ciri khasnya masing-masing. Algoritma *clustering* yang paling efisien untuk menentukan *cluster* pada data dengan kepadatan yang berbeda adalah algoritma *density based clustering* [3].

DBSCAN adalah salah satu contoh pelopor perkembangan teknik pengelompokan berdasarkan kepadatan atau yang biasa dikenal dengan sebutan *density based clustering* [4]. Penelitian menggunakan Metode DBSCAN telah beberapa kali dilakukan sebelumnya.

Danu Zakrzewska menerapkan konsep *data mining* dalam proses segmentasi pelanggan (*customer segmentation*) pada sebuah bank. Penelitian ini membandingkan tiga algoritma *clustering* dalam hal *high dimensionality data with noise* yaitu DBSCAN, K-Means, dan *Two-phase Clustering* [5].

Penelitian lainnya dilakukan oleh Xiaohui Hu dengan melakukan proses segmentasi pelanggan pada sebuah maskapai penerbangan dalam hubunggannya dengan *Customer Relationship Management* (CRM). Penelitian ini menggunakan tiga metode utama yaitu *K-Means*, DBSCAN dan *Biclustering*. Metode DBSCAN dalam penelitian ini digunakan untuk mengelompokkan pelanggan ke dalam tiga grup yang berbeda [6].

Penelitian ini membahas implementasi Metode DBSCAN pada proses pengambilan keputusan. Metode DBSCAN dalam penelitian ini digunakan untuk membantu menentukan pelanggan potensial pada sebuah perusahaan dengan menggunakan parameter input *minimal point (minpts)* dan *epsilon (eps)*. Proses penentuan nilai parameter bersifat *trial and error*, artinya penentuan nilai parameter harus diuji coba beberapa kali hingga mendapatkan jumlah *cluster* tertentu.

2. Metodologi Penelitian

Metodologi penelitian yang digunakan untuk mendapatkan hasil sesuai dengan yang diharapkan adalah dengan terlebih dahulu melakukan studi pustaka/*literature* untuk mengetahui metode yang digunakan dan dijadikan referensi dalam penelitian ini. Pengumpulan data dan informasi mengenai objek pada penelitian ini adalah hal kedua yang harus dilakukan untuk mengetahui data terkini dan dapat menetapkan metode yang digunakan sesuai dengan studi pustaka, yang dilanjutkan dengan pemodelan sistem. Langkah selanjutnya adalah dengan mempelajari algoritma metode yang digunakan agar dapat membuat rancangan metode dengan lebih baik. Proses yang dilakukan selanjutnya adalah pembuatan program. Pengujian pada program yang sudah dibuat perlu dilakukan untuk mengetahui siap atau tidaknya program tersebut untuk digunakan. Pembuatan laporan penelitian dilakukan pada tahap akhir untuk merangkum keseluruhan proses penelitian.

3. Kajian Pustaka

Data mining merupakan sebuah langkah dalam proses *Knowledge Discovery in Database* (KDD) yang terdiri dari penerapan analisis data dan penemuan algoritma yang menghasilkan enumerasi tertentu terhadap pola pada data [1]. Tan juga mengartikan *data mining* sebagai sebuah proses ekstraksi informasi baru dari sejumlah besar data yang dapat berguna dalam proses pengambilan keputusan [7]. Proses penambangan pengetahuan dari sejumlah besar data spasial dikenal sebagai *spatial data mining* [4].

Spatial Data Mining [2] adalah bagian dari *data mining* yang merupakan proses menemukan pola yang menarik dan sebelumnya tidak dikenal tetapi secara potensial dapat berguna dari *dataset* spasial yang besar. Penggalian pola yang menarik dan berguna dari *dataset* spasial lebih sulit daripada penggalian pola data numerik tradisional dan kategorikal dikarenakan oleh kompleksitas jenis, hubungan dan autokorelasi dari *dataset* spasial tersebut.

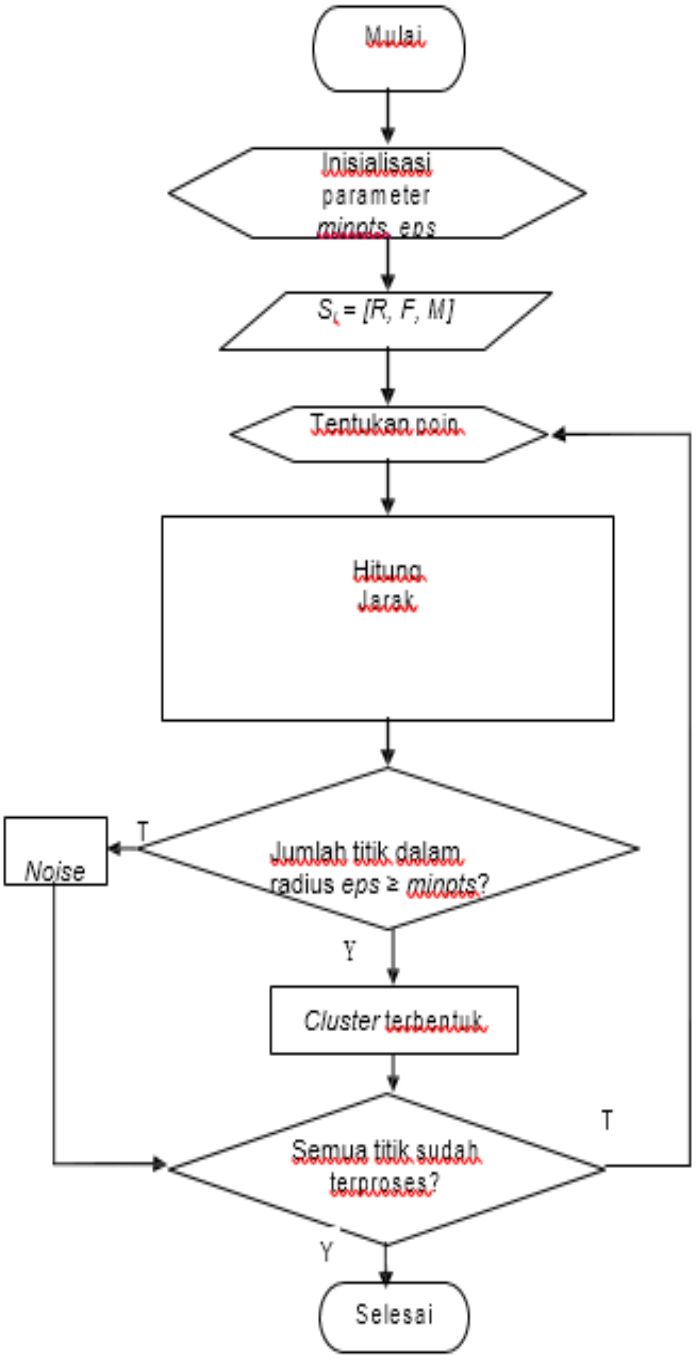
Sebagian besar penelitian terbaru pada data spasial menggunakan teknik *clustering* dikarenakan oleh sifat dari data tersebut. *Clustering* merupakan proses pengelompokan sejumlah besar data menjadi beberapa kelas sesuai dengan ciri khasnya masing-masing. Di antara berbagai jenis algoritma *clustering*, *density based clustering* lebih efisien untuk menentukan *cluster* pada data dengan kepadatan yang berbeda [3]. *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) adalah salah satu contoh pelopor perkembangan teknik pengelompokan berdasarkan kepadatan atau yang biasa dikenal dengan sebutan *density based clustering* [4].

Density-Based Spatial Clustering of Application with Noise (DBSCAN) merupakan sebuah metode *clustering* yang membangun area berdasarkan kepadatan yang terkoneksi (*density-connected*). Setiap objek dari sebuah radius area (*cluster*) harus mengandung setidaknya sejumlah minimum data. Semua objek yang tidak termasuk di dalam *cluster* dianggap sebagai

noise. Komputasi dari Algoritma *Density Based Spatial Clustering of Application with Noise* (DBSCAN) adalah sebagai berikut:

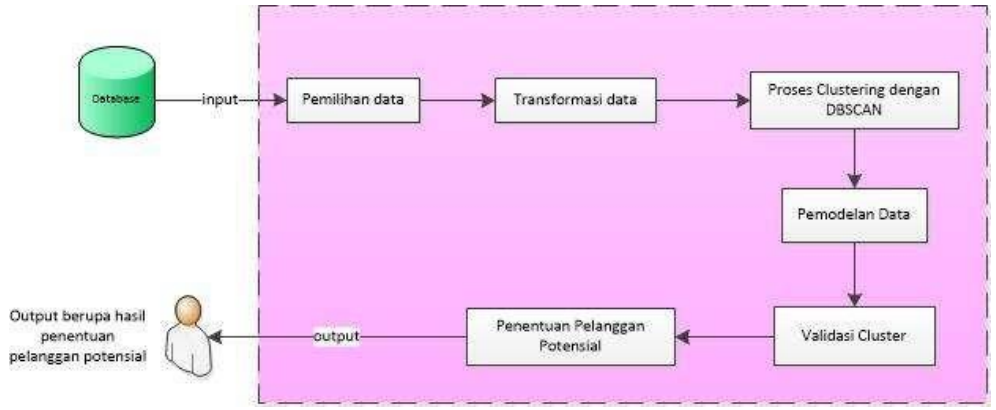
a. Inisialisasi parameter *minpts, eps*.
b. Tentukan titik awal atau *p* secara acak.
c. Ulangi langkah 3 – 5 hingga semua titik diproses.
d. Hitung *eps* atau semua jarak titik yang *density reachable* terhadap *p*.
e. Jika titik yang memenuhi *eps* lebih dari *minpts* maka titik *p* adalah *core point* dan *cluster* terbentuk.
f. Jika *p* adalah *border point* dan tidak ada titik yang *density reachable* terhadap *p*, maka proses dilanjutkan ke titik yang lain.

Flowchart langkah-langkah *clustering* menggunakan algoritma DBSCAN dapat dilihat pada Gambar 1.



Gambar 1. *Flowchart* Komputasi Algoritma DBSCAN

Gambaran umum implementasi Metode DBSCAN pada proses penentuan keputusan untuk mencari pelanggan potensial ditunjukkan pada Gambar 2.

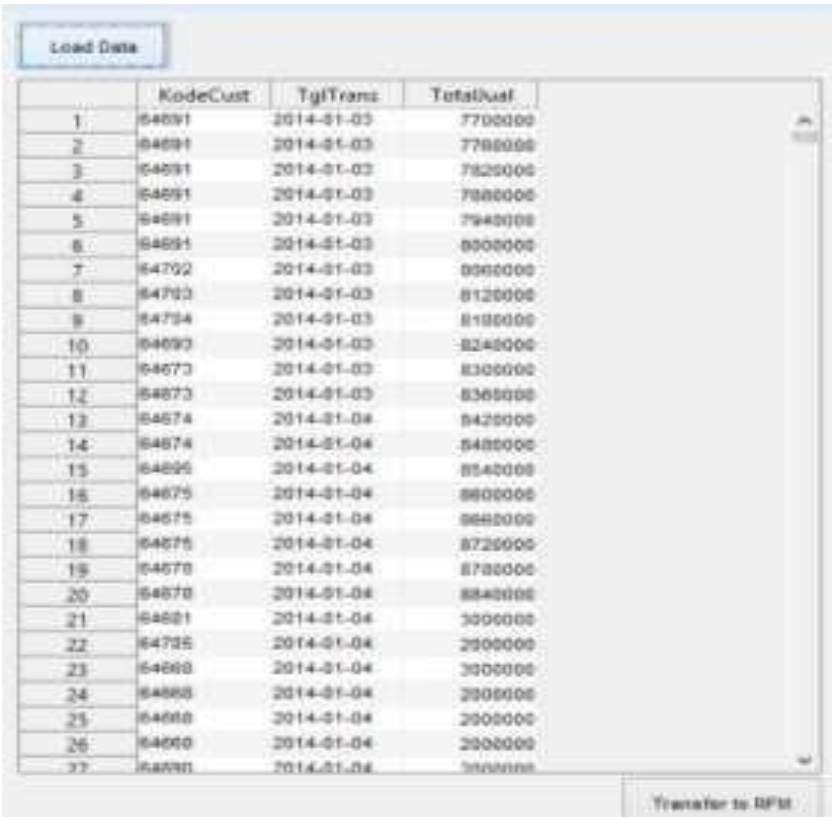


Gambar 2. Gambaran Umum Sistem Implementasi Metode DBSCAN

Proses sistem diawali dengan persiapan data berupa pemilihan data. Data yang digunakan dalam pembuatan sistem ini adalah data yang berkaitan dengan transaksi. Tahap selanjutnya adalah transformasi data untuk merubah data mentah menjadi *field-field* data yang sesuai dengan yang dibutuhkan sebagai *input* dalam proses *clustering* menggunakan Algoritma DBSCAN. Proses *clustering* menghasilkan beberapa *cluster*. Nilai rata-rata dari tiap *cluster* yang didapat inilah yang digunakan untuk menentukan kelas pelanggan pada tahap pemodelan data. Setelah melalui proses *clustering, cluster-cluster* yang terbentuk kemudian diuji tingkat validitasnya untuk mengetahui jumlah *cluster* yang terbaik. Tahap terakhir adalah penentuan pelanggan potensial melalui kelas yang dihasilkan oleh tiap *cluster*.

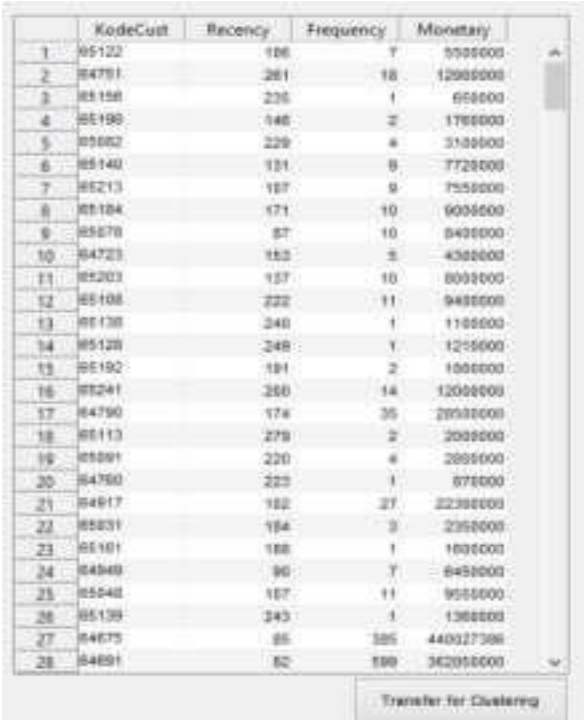
4. Hasil dan Pembahasan

Pengujian sistem implementasi metode DBSCAN ini dilakukan untuk mengetahui bagaimana metode ini mampu menentukan pelanggan potensial untuk membantu proses pengambilan keputusan. Langkah awal yang harus dilakukan dalam implementasi Metode DBSCAN adalah memilih *field-field* yang digunakan dalam proses *clustering*. *Field-field* tersebut harus dapat merepresentasikan nilai-nilai pada proses transformasi data. *Field-field* ini disimpan pada sebuah tabel statis yang disebut tabel standar seperti yang terlihat pada Gambar 3.



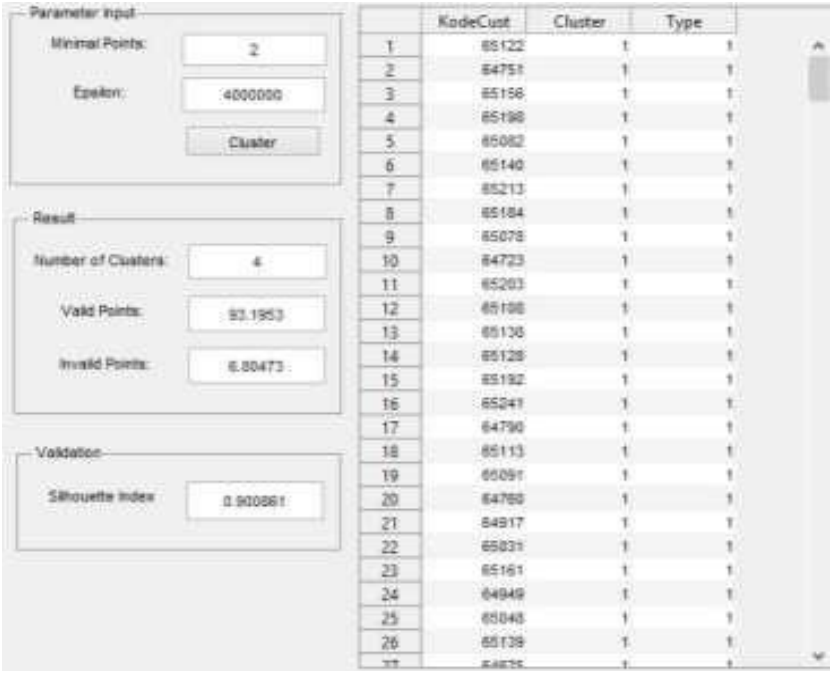
Gambar 3. Pemilihan Data

Proses yang dilakukan setelah *field* yang digunakan telah ditentukan adalah proses transformasi menjadi nilai yang dibutuhkan untuk *input* dalam proses *clustering* seperti yang terlihat pada Gambar 4.



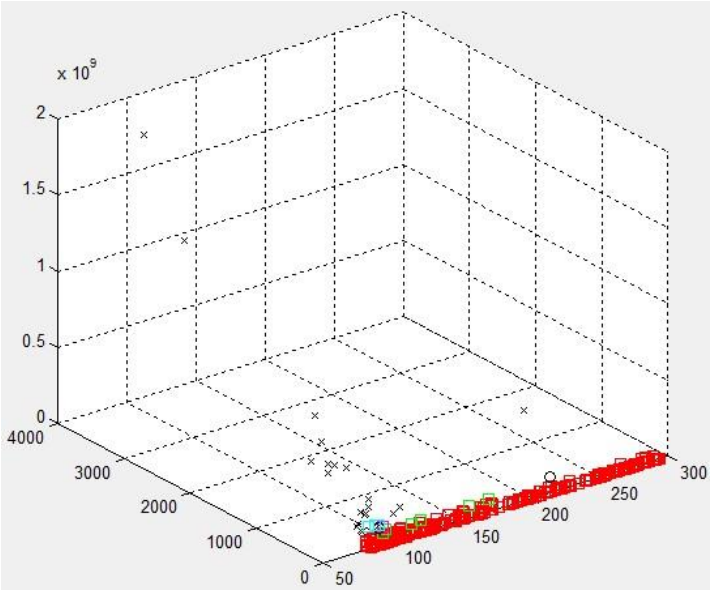
Gambar 4. Transformasi Data

Proses uji coba *clustering* Metode DBSCAN ini menggunakan data selama satu tahun yang dibentuk menjadi beberapa *cluster*. Jumlah *cluster* pada Metode DBSCAN tidak ditentukan oleh pengguna. Jumlah *cluster* bergantung pada nilai *Minimal Points* (*minpts*) dan *Epsilon* (*eps*) yang digunakan. Hasil proses *clustering* seperti yang terlihat pada Gambar 5 menunjukkan bahwa penggunaan nilai parameter *minpts* = 2 dan *eps* = 4000000 menghasilkan 4 *cluster* dengan *silhouette index* sebesar 0,900861. Nilai *silhouette index* yang dihasilkan lebih besar dari 0 dan mendekati 1 yang mana berarti bahwa jumlah *cluster* yang dihasilkan pada proses *clustering* sudah optimal.



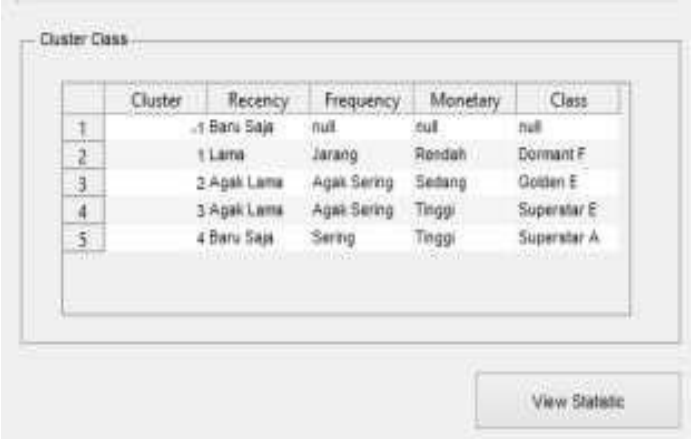
Gambar 5. *Clustering* dengan DBSCAN

Hasil *clustering* digambarkan dalam sebuah grafik tiga dimensi. Titik berbentuk kotak adalah *core point*, titik berbentuk lingkaran adalah *border point* sedangkan titik berbentuk silang adalah *noise* seperti yang terlihat pada Gambar 6.



Gambar 6. Grafik Proses *Clustering*

Langkah selanjutnya setelah melakukan proses *clustering* adalah mencari kelas bagi tiap *cluster*. Hasil proses perbandingan untuk mencari kelas bagi tiap *cluster* dapat dilihat pada Gambar 7.



Gambar 7. Penentuan Pelanggan Potensial

Tabel kelas *cluster* menunjukkan kelas yang dimiliki oleh tiap *cluster*. Kelas inilah yang digunakan untuk membantu proses pengambilan keputusan dalam menentukan pelanggan potensial. Hasil *clustering* menunjukkan bahwa Metode DBSCAN telah berhasil membentuk beberapa *cluster* dengan jenis kelas yang berbeda. *Cluster* 4 adalah *cluster* dengan pelanggan yang paling potensial karena masuk ke dalam kelas *superstar A*.

5. Kesimpulan

Pemilihan parameter merupakan proses yang sangat penting dalam Algoritma DBSCAN karena mempengaruhi kinerja algoritma dalam pembentukan *cluster* dan jumlah *noise* yang dihasilkan. Proses *clustering* menggunakan Algoritma DBSCAN menghasilkan sejumlah *cluster* dengan rata-rata nilai indeks validitas menggunakan Algoritma Silhouette lebih besar dari 0 dan mendekati 1, yang mana interval nilai indeks validitas algoritma ini yaitu -1 sampai dengan 1. Hal ini menandakan bahwa proses *clustering* menggunakan algoritma DBSCAN telah dapat dikategorikan baik.

Daftar Pustaka

[1] Fayyad U, Piatetsky-Shapiro G, and Smyth P, “Knowledge Discovery and Data Mining: Towards a Unifying Framework,” in *Proceedings of the 2nd Int. Conference on Knowledge Discovery and Data Mining. Portland*, 1996, pp. 82–88.

[2] Shekhar S, Zhang P, Huang Y, and Vatsavai RR, “Trends in spatial data mining. Data mining: Next generation challenges and future directions,” pp. 357–380, 2003.

[3] Matheus CJ, Chan PK, and Piatetsky-Shapiro G, “Systems for Knowledge Discovery in Databases,” *IEEE Trans. Knowl. Data Eng.*, vol. 5, no. 6, pp. 903–913, 1993.

[4] Mumtaz K, “An Analysis on Density Based Clustering of Multi Dimensional Spatial Data,” *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 1, no. 1, pp. 8–12, 2010.

[5] Zakrzewska D and Murlewski J, “Clustering Algorithms for Bank Customer Segmentation,” in *Proceedings of 5th Int. Conference on Intelligent Systems Design and Applications, Poland*, 2005, pp. 197–202.

[6] H. Xiaohui, “A New Customer Segmentation Framewok Based on Biclustering Analysis,” *J. Softw.*, vol. 9, no. 6, pp. 1359–1366, 2014.

[7] P. Tan, *Introduction to Data Mining*. Boston: Pearson Education, 2006.