

SEKOLAH TINGGI MANAJEMEN INFORMATIKA dan KOMPUTER WIDYA CIPTA DHARMA

Status Terakreditasi Badan Akreditasi Nasional Perguruan Tinggi
Jl. M. Yamin No. 25 Telp. 0541 - 736071 Fax. 734468 E-mail : wicida@wicida.ac.id

SURAT TUGAS

No : 456 / Ak – SPD / XII / 2019

Yang bertanda tangan dibawah ini Ketua STMIK Widya Cipta Dharma memberi tugas kepada :

NO	NAMA	JABATAN
1.	Dr. H. Nursobah, S.Kom., M.Kom	Dosen Tetap Prodi TI
2.	Drs. Azahari, M.Kom	Dosen Tetap Prodi TI
3.	Eka Arriyanti, S.Pd., M.Kom	Dosen Tetap Prodi TI
4.	Siti Lailiyah, S.Kom., M.Kom	Dosen Tetap Prodi TI
5.	Dr. Heny Pratiwi, S.Kom., M.Pd., M.TI	Dosen Tetap Prodi SI
6.	Kusno Harianto, S,Kom., M.Kom	Dosen Tetap Prodi SI
7.	H. Pajar Pahrudin, S.Kom., MH	Dosen Tetap Prodi SI
8.	Hj. Ekawati Yulsilviana, SP., MM	Dosen Tetap Prodi MI

Acara : Seminar Hasil Luaran PDP
Lokasi : Hotel Menara Bahtera
Jl. Gajah Mada Kota Balikpapan
Tanggal : 16 Desember 2019

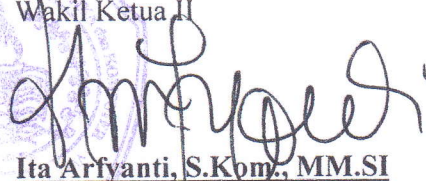
Demikian Surat Tugas ini diberikan untuk dipergunakan sebagaimana mestinya.

Dikeluarkan di : Samarinda

Tanggal : 14 Desember 2019

An. Ketua,

Wakil Ketua II


Ita Arfyanti, S.Kom., MM.SI
NIK. 95.08.1.005

Tembusan :
1. Ka. BAAK
2. Yang Bersangkutan
3. Arsip



Photo setelah Seminar Hasil PDP, 16 Desember 2019



LLDIKTI XI



Pengenalan Pola Data *NOISE* Dengan Pendekatan Chaos Untuk Optimalisasi Penambangan Data¹



Eka Arriyanti
Teknik Informatika



Ita Arfyanti Pitrascha Adytia
Sistem Informasi



STMIK Widya Cipta Dharma, Indonesia

¹Penelitian disponsori oleh Kemenristekdikti-Republik Indonesia, Hibah PDP 2019.



LLDIKTI XI



Pengenalan Pola Data *NOISE* Dengan Pendekatan Chaos Untuk Optimalisasi Penambangan Data¹



Peneliti Utama :

Eka Arriyanti, S.Pd., M.Kom.
Teknik Informatika
STMIK Widya Cipta Dharma
ekaarry@wicida.ac.id

¹Penelitian disponsori oleh Kemenristekdikti-Republik Indonesia, Hibah PDP 2019.

RINGKASAN

Pengenalan pola data *noise* dari suatu hasil penambangan data akan memberikan informasi yang sangat berharga untuk optimalisasi penambangan data. Oleh karena terdapat enam jenis penambangan data, penelitian ini menunjukkan optimalisasi data mining tipe *clustering*-jumlah *clustering* tidak diketahui dengan menemukan "*similarity*" pola data *noise* dalam "*complexity*" data, menggunakan pendekatan Chaos. Eksekusi dengan algoritma asli DBSCAN menggunakan 8.0 Java Language dari 0.7.5 Elki data mining framework. Penelitian menghasilkan temuan baru berupa Spatial Coordinate Way, identifikasi pola *noise*, dan rumus untuk menentukan parameter dari DBSCAN, berupa $\mathcal{E}1, \dots, \mathcal{E}4 = (((\text{minPts} - 1) - 1) - 1) - 1, \dots, (\text{minPts} - 1)$.

-I-

SUMMARY

Pattern identification on noise data of a data mining result would give very valuable information for data mining optimization. Since there were six types of data mining, this research pointed out the optimizing of clustering type data mining-an unknown number of clustering with finding the "similarity" pattern of noise data in the "complexity" of data, used Chaos approach. The execution by the original algorithm of DBSCAN-8.0 Java Language from 0.7.5 Elki data mining framework. The novel findings of the research are Spatial Coordinate Way, noise pattern identification, and pattern for determining parameters of DBSCAN ; $\mathcal{E}1, \dots, \mathcal{E}4 = (((\text{minPts} - 1) - 1) - 1) - 1, \dots, (\text{minPts} - 1)$.

LATAR BELAKANG

Identifikasi pada pola data *noise* dalam penambangan data sangat penting karena data *noise* mempengaruhi pemrosesan data dalam aplikasi yang menyebabkan hasil penambangan data tidak sesuai harapan, sedangkan implementasi hasil penambangan data melibatkan keputusan manajemen informasi yang tinggi. Keberadaan data *noise* adalah suatu keniscayaan dalam *big data* yang membuatnya hanya mungkin untuk dikurangi, tidak dapat dihilangkan, dan reduksinya akan menghindari kesalahan dalam proses penambangan data [1]. Tetapi, upaya untuk mereduksi data *noise* melibatkan "*data complexity*" [2], karena polanya yang kompleks [3].

Penelitian ini meyakini bahwa reduksi data *noise* melibatkan pemahaman tentang pola data *noise* dari suatu penambangan data. Jika polanya dapat diidentifikasi, maka akan menjadi informasi yang sangat berharga untuk optimalisasi. Karena ada enam jenis penambangan data [4] dan banyaknya algoritma penambangan data, penelitian ini dimulai dengan jenis penambangan data yang populer, *clustering*. Ada dua jenis *clustering*, jumlah *clustering* diketahui dan jumlah *clustering* tidak diketahui. Yang diketahui berarti jumlah *clustering* ditentukan sebelumnya, sedangkan yang tidak diketahui tidak ditentukan. Kemudian, suatu algoritma dipilih dengan memahami akar teorinya, sehingga dapat disimpulkan bahwa suatu jenis penambangan data direpresentasikan paling utama oleh algoritma tersebut.

Setelah mengkaji beberapa algoritma *clustering*, Density-Based Spatial Application with Noise (DBSCAN) mewakili secara terutama untuk jenis jumlah *clustering* tidak diketahui, sedangkan K-Means untuk yang diketahui. Kemudian, diputuskan untuk memulai penelitian optimisasi penambangan data ini dengan jumlah *clustering* tidak diketahui, karena lebih unik. "*Data complexity*" dari data *noise* hasil penambangan data akan diidentifikasi dengan pendekatan Chaos untuk menemukan "*similarity*" dalam "*complexity*" [5], yang tidak digunakan oleh [3]. "*Similarity*" adalah pola yang akan dicantumkan dalam tabulasi.

Masalah dalam penelitian optimalisasi ini adalah : DBSCAN awalnya ditujukan untuk data spasial [6] sementara algoritma harus berkinerja baik pada data apa pun [7], bagaimana dengan data non-spasial, dan ; masalah penentuan parameter DBSCAN. Jawaban dari kedua masalah ini akan menjadi preposisi untuk kesimpulan penelitian. Oleh karena itu, metode untuk menarik kesimpulan adalah Conceptual Analysis-model of Detection Analysis for an undetermined number of times (CA-DAunt) [8]. Tujuan akhir dari penelitian ini adalah terwujudnya optimalisasi pada setiap jenis penambangan data. Dengan demikian, hasil penelitian ini adalah awalan untuk penelitian lanjutan tentang implementasi optimalisasi.

Sitasi :

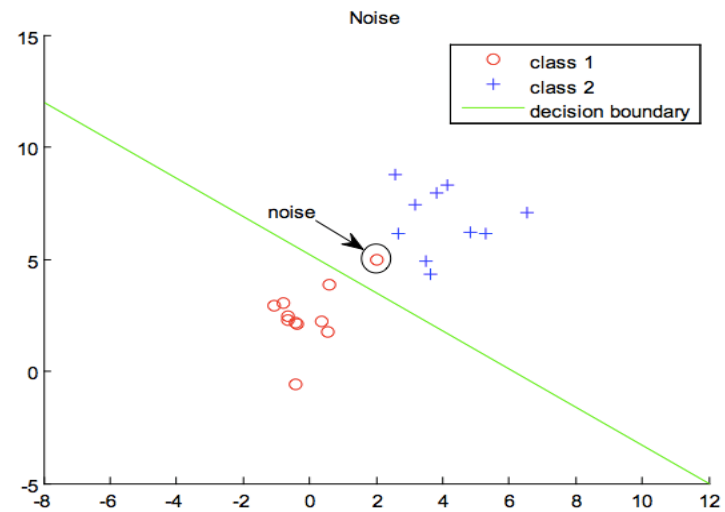
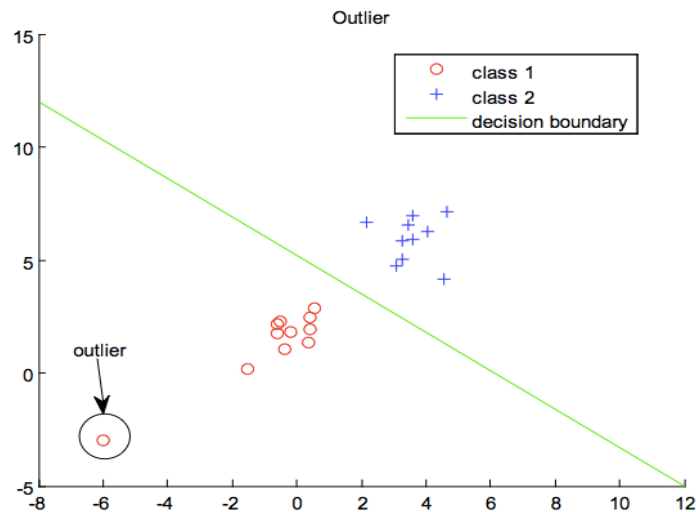
1. Seiffert, C. At all (2011) An Empirical Study Of The Classification Performance Of Learners On Imbalanced And Noisy Software Quality Data. Elsevier Journal On Information Sciences, vol. 259, p. 571-595, ISSN : 0020-0255, DOI : 10.1016/j.ins.2010.12.016.
2. Saez, J.A., Luengo, J., and Herera, F. (2013) Predicting Noise Filtering Efficacy With Data Complexity Measures For Nearest Neighbor Classification. Elsevier Journal On Pattern Recognition, vol. 46, p. 355-364, ISSN : 0031-3203, DOI : 10.1016/j.patcog.2012.07.009.
3. Saez, J.A. at all. (2014) Analyzing the Presence of Noise in Multi-class Problems : Alleviating Its Influence With The One-Vs-One Decomposition. Knowledge and Information Systems, Springer Regular Paper, vol. 38(1), p. 179-206, DOI :10.1007/s10115-012-0570-1.
4. Agustin, S. dkk. (2016) Perbandingan Metode K-Means Dan Metode DBSCAN Pada Pengelompokan Rumah Kos Mahasiswa Di Kelurahan Tembalang, Semarang. Jurnal Gaussian, vol. 5(4), hal. 757-762, ISSN : 2339-2541.
5. Kusmarni, Y.(2008) Teori Chaos : Sebuah Keteraturan Dalam Keacakan. Universitas Pendidikan Indonesia, Bandung.
6. Ester, M., Kriegel, H.P., Sander, J., and Xu, X. (1996) A Density-Based Algorithm For Discovering Clusters In Large Spatial Databases With Noise, Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD), pp. 226-231, aaai.org.
7. Schubert, E., Sander, J., Ester, M., Kriegel, H.P., and Xu, X. (2017) DBSCAN Revisited, Revisited : Why And How You Should (still) Use DBSCAN, ACM Transactions on Database Systems, vol. 42, no. 3, article 19, 21 pages, July 2017, DOI : 10.1145/3068335.
8. Kosterec, M. (2016) Methods Of Conceptual Analysis. Journal Of Filozovia, Komenskeho University, vol. 71(3), p. 220-230, ISSN : 0046-385X, Bratislava, Republic Of Slovakia.

Tahapan Penelitian



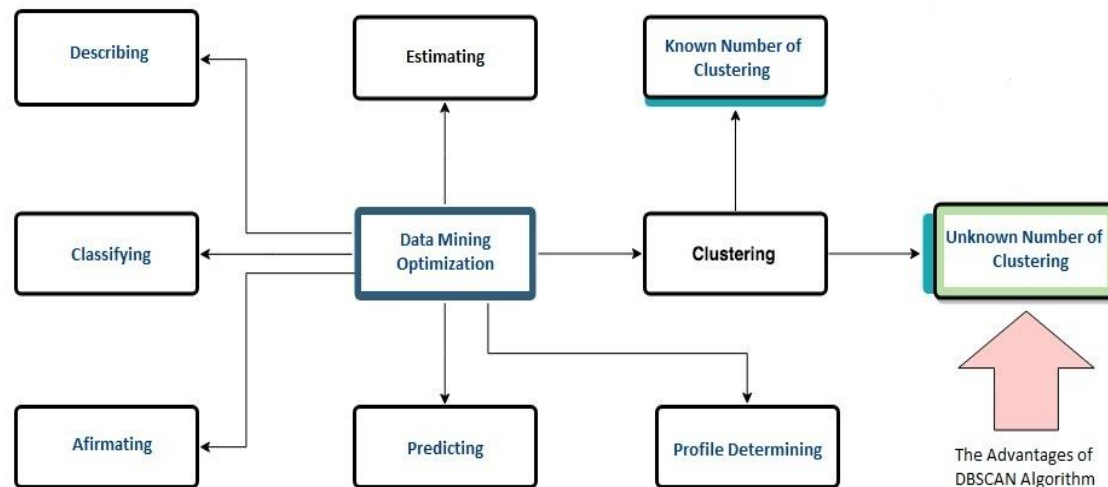
Noise ?

An outlier can be a valid data point, or it can be noise, but noise is unwanted and wrong data, and should be removed (A. Al Shaebi, 2016). Noise before preprocessing is easy to remove. Noise when data mining is how it be resulted much lower. Noise after data mining is what to be removed. So that, efforts to reduce noise mean how the data mining resulted much lower noise, or much lower data should be removed after data mining



Jenis-jenis Penambangan Data

Penambangan data terdiri dari klasifikasi, estimasi, prediksi, afirmasi (*afinitation*), *clustering*, deskripsi dan penentuan profil (M.J.A Berry dan G.S. Linoff dalam S. Agustin, 2016).

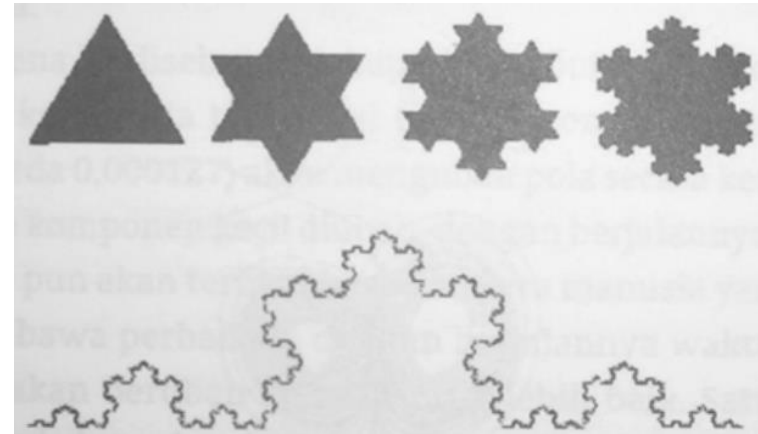


Pendekatan Chaos

Pendekatan Chaos adalah pendekatan terhadap penyelesaian masalah yang didasari oleh teori Chaos, yaitu “terdapat suatu keteraturan dalam keacakan” atau “teori tentang gerakan atau dinamika yang kompleks dan tidak terduga (*chaotic*) dari suatu sistem yang tergantung dari kondisi awalnya ; sistem *chaotic* tetap dapat digeneralisasikan secara matematis, karena sistem *chaotic* mengikuti hukum-hukum yang berlaku di alam ; sifatnya yang tidak teratur mengakibatkan penampakan keadaan yang acak”. Untuk mencari dan menemukan pola dalam beragam proses acak alamiah, gunakan berbagai teknik matematis sehingga kemiripan muncul”. Dengan demikian, dalam *chaos*, pengulangan teratur atau kemiripan pada skala yang berbeda, selanjutnya dapat mengakibatkan chaos lagi, jika chaos tidak dikontrol (Mandelbrot dalam Y. Kusmarni, 2008).



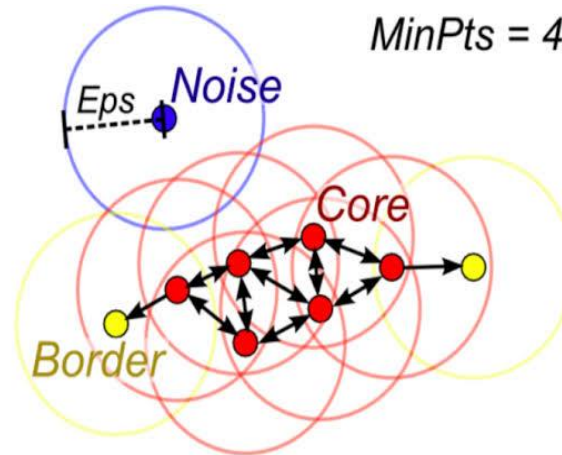
Kurva Lorenz ; Butterfly Effect



Kurva Koch

DBSCAN

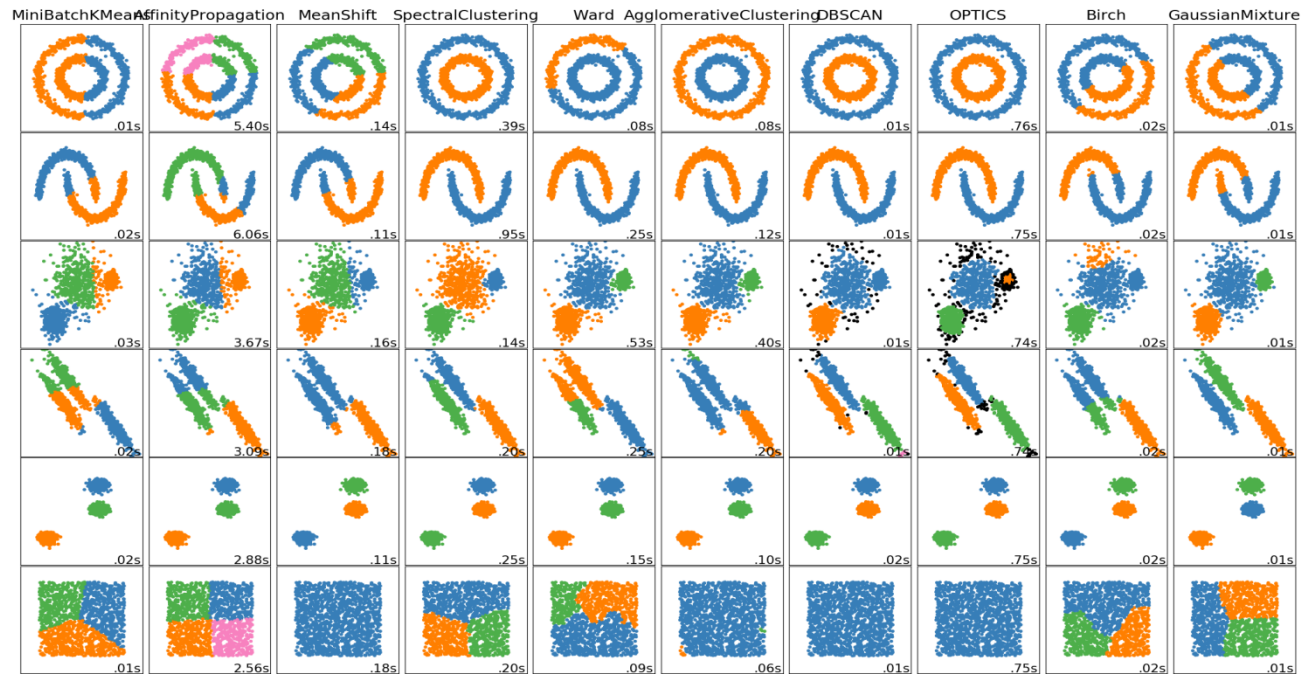
MinPts = 4



Red: Core Points

Yellow: Border points. Still part of the cluster because it's within epsilon of a core point, but does not meet the min_points criteria

Blue: Noise point. Not assigned to a cluster



Hasil Penelitian

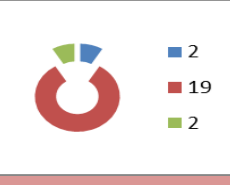
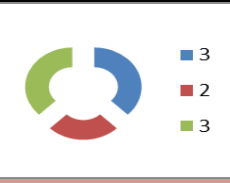
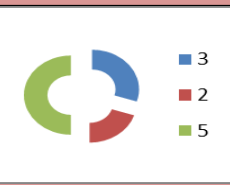
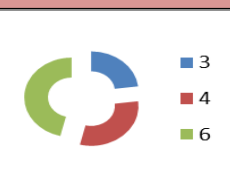
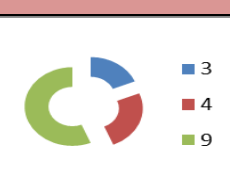
Spatial Coordinate way :

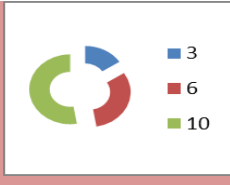
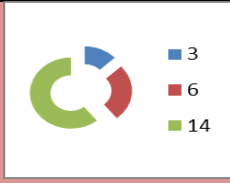
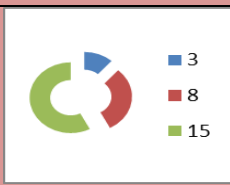
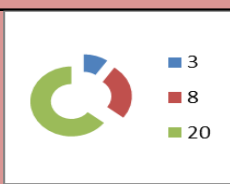
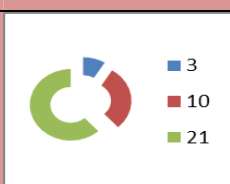
Convert **dimensions**, the steps are :

- 1) Define non-spatial datasets and their number of dimensions; D.
- 2) Analyzing the relationship between variables.
- 3) Sorting the coordinates formed based on the tendency of variables based on the purpose of clustering.
- 4) Normalize the data in the previous step variable (3; 4-1); the normalization scale must be Integer.
- 5) Determine the representation of the variables in the coordinates formed to become coordinates (X, Y).
- 6) Normalize data in the previous step variable (5; 6-1); the normalization scale must be Integer.

Hasil Penelitian

Identifikasi Pola Noise :

n	Index	ϵ	minPts	Clustered (Cd)	Noise	Pattern
30	1	2	5	2	19	
60	2	3	5	3	2	
90	3	3	5	5	2	
120	4	3	5	6	4	
150	5	3	5	9	4	

180	6	3	5	10	6	
210	7	3	5	14	6	
240	8	3	5	15	8	
270	9	3	5	20	8	
300	10	3	5	21	10	
...	

Hasil Penelitian

Rumus Parameter DBSCAN :

$$\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4 = (((\mathit{minPts} - 1) - 1) - 1) - 1, (((\mathit{minPts} - 1) - 1) - 1), ((\mathit{minPts} - 1) - 1), (\mathit{minPts} - 1)$$

atau

$$\text{Pilih } \mathcal{E}_1, \dots, \mathcal{E}_4 = (((\mathit{minPts} - 1) - 1) - 1) - 1, \dots, (\mathit{minPts} - 1)$$

Penelitian Lanjutan :

- 1) Aplikasi implementasi algoritma DBSCAN untuk sembarang data ; PROTOTIPE PERANGKAT LUNAK OPTIMALISASI PENAMBANGAN DATA CLUSTERING-JUMLAH CLUSTERING TIDAK DIKETAHUI ; judul proposal Hibah PDP 2019-2020
- 2) *Deep Analysis* Koordinat Cartesian untuk Ruang Lanjut

Referensi :

B. Ren, L. Pueyo, G.B. Zhu, J. Debes, and G. Duchene, “Non-negative matrix factorization : robust extraction of extended structures,” *The Astrophysical Journal*, vol. 852, no. 104, 16 pages, January 2018.

E. Schubert, J. Sander, M. Ester, H.P. Kriegel, and X. Xu, “DBSCAN revisited, revisited : why and how you should (still) use DBSCAN,” *ACM Transactions on Database Systems*, vol. 42, no. 3, article 19, 21 pages, July 2017.

I.D. Id, Astrid, E. Mahdiyah, “DBSCAN modification (Density-Based Spatial Clustering With Noise) on 3-dimensional objects ; modifikasi DBSCAN (Density-Based Spatial Clustering With Noise) pada Objek 3 Dimensi,” *Caltex Polytechnic Journal on Applied Computers*, vol. 3, no. 1, pp. 41-52, May 2017.

Encyclopedia of Mathematics, “Cartesian Orthogonal Coordinate System, October 2017.

V. Sharma, “Spatial and non-spatial data,” *GeolTHub*, June 2016.

PENGENALAN POLA DATA *NOISE* DENGAN PENDEKATAN CHAOS UNTUK OPTIMALISASI PENAMBANGAN DATA¹



Eka Arriyanti
Teknik Informatika
STMIC Widya Cipta Dharma, Indonesia



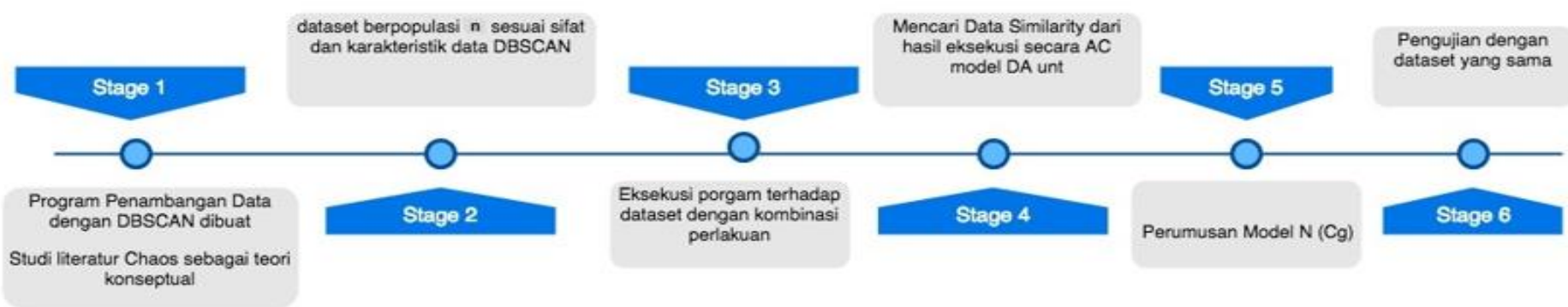
Ita Arfyanti Pitrascha Adytia
Sistem Informasi



STMIC Widya Cipta Dharma, Indonesia

¹Penelitian disponsori oleh Kemenristekdikti-Republik Indonesia, Hibah PDP 2019.

Tahapan Penelitian



Hasil Penelitian

Identifikasi Pola Noise :

n	Index	E	minPts	Clustered (Cd)	Noise	Pattern
30	1	2	5	2	19	
60	2	3	5	3	2	
90	3	3	5	5	2	
120	4	3	5	6	4	
150	5	3	5	9	4	

180	6	3	5	10	6	
210	7	3	5	14	6	
240	8	3	5	15	8	
270	9	3	5	20	8	
300	10	3	5	21	10	

30	2	5	2	19
60	3	5	3	2
90	3	5	5	2
120	3	5	6	4
150	3	5	9	4

Referensi :

E. Schubert, J. Sander, M. Ester, H.P. Kriegel, and X. Xu, "DBSCAN revisited, revisited : why and how you should (still) use DBSCAN," ACM Transactions on Database Systems, vol. 42, no. 3, article 19, 21 pages, July 2017.