



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. 001/Penelitian/P3M-WCD/2020

Yang bertanda tangan di bawah ini, Kepala Pusat Penelitian dan Pengabdian Masyarakat STMIK Widya Cipta Dharma memberikan tugas kepada :

No.	NIDN	Nama	Jabatan
1	1101127504	Eka Arriyanti, S.Pd, M.Kom.	Dosen/Ketua Peneliti
2	1126038303	Pitrasacha Adytia, M.T.	Dosen/Anggota Peneliti

Untuk melaksanakan penelitian hibah Penelitian Dosen Pemula (PDP) tahun 2019-2020 dengan judul kegiatan “Prototipe Perangkat Lunak Optimalisasi Penambangan Data Clustering-Jumlah *Clustering* Tidak Diketahui”.

Demikian surat tugas ini diberikan untuk dipergunakan sebagaimana mestinya.

Samarinda, 2 Januari 2020

Ka. P3M,

Siti Qomariah, M.Kom.

NIK : 11.12.1.057



SERTIFIKAT

Nomor : SN-WCD/OP/IV/2020/056

Diberikan kepada

EKA ARRIYANTI

Sebagai Pemakalah (Oral Presenter) yang telah mempresentasikan hasil Penelitian/Pengabdian yang berjudul

PROTOTYPE PERANGKAT LUNAK OPTIMALISASI PENAMBAHAN DATA CLUSTERING-JUMLAH CLUSTERING TIDAK DIKETAHUI

Dalam temu ilmiah **Seminar Nasional Multidisiplin Sebatik (SNSEBATIK) 2020** dengan tema **"Unggul dan Berdaya di Era Revolusi Industri 4.0"** yang diselenggarakan di **STMIK Widya Cipta Dharma Samarinda**, 18 April 2020. *Temu ilmiah ini adalah hasil kerjasama STMIK Widya Cipta Dharma dan para Editorial Jurnal Nasional dan Internasional.*

Ketua
STMIK Widya Cipta Dharma

Komite Pelaksana
SNSEBATIK 2020



Dr. H. Nursobah, M.Kom
Scopus ID: 57203969399



Reza Andrea, M.Kom
Scopus ID: 57195557287





LLDIKTI XI



PROTOTIPE PERANGKAT LUNAK OPTIMALISASI PENAMBANGAN DATA CLUSTERING-JUMLAH CLUSTERING TIDAK DIKETAHUI¹



Eka Arriyanti
Teknik Informatika



Pitrascha Adytia
Sistem Informasi

STMIK Widya Cipta Dharma, Indonesia

¹Penelitian disponsori oleh Kemenristekdikti-Republik Indonesia, Hibah PDP 2020.



LLDIKTI XI



PROTOTIPE PERANGKAT LUNAK OPTIMALISASI PENAMBANGAN DATA CLUSTERING-JUMLAH CLUSTERING TIDAK DIKETAHUI¹



Peneliti Utama :

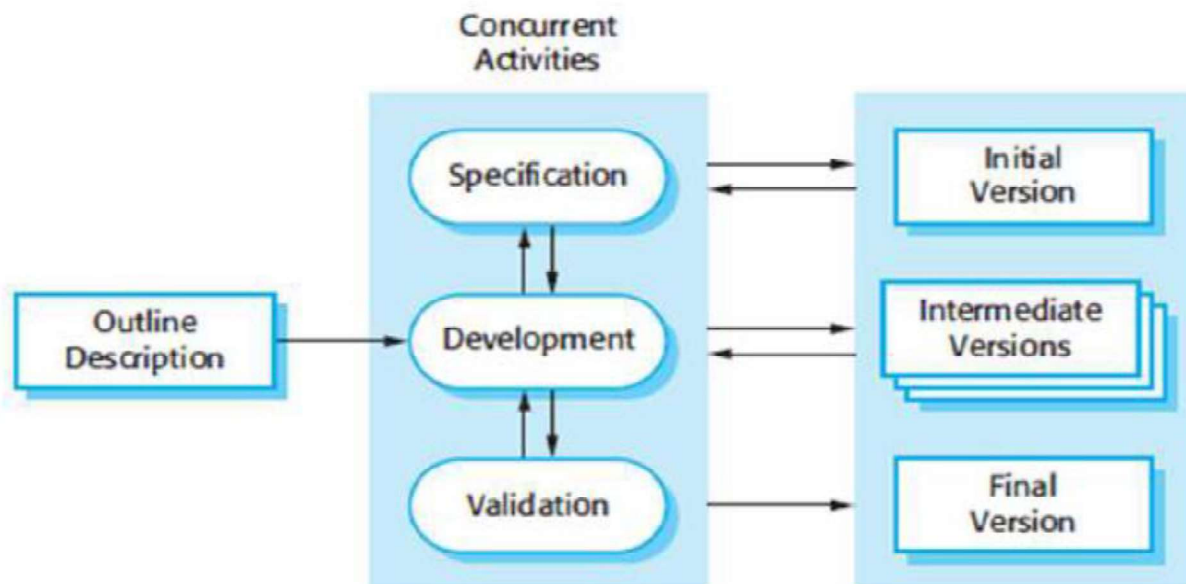
Eka Arriyanti, S.Pd., M.Kom.
Teknik Informatika
STMIK Widya Cipta Dharma
ekaarry@wicida.ac.id

¹Penelitian disponsori oleh Kemenristekdikti-Republik Indonesia, Hibah PDP 2020.

RINGKASAN

Penambangan terhadap data harus dapat memberikan hasil yang optimal agar informasi yang didapatkan dari tumpukan data (*big data*) selama kurun waktu kegiatan (pekerjaan) hanya menyisakan sedikit ‘sampah’ data. Uji coba *preprocessing* data non spasial dengan Spasial Koordinat, *clustering* hasilnya, dan identifikasi pola data *noise* yang ditunjukkan oleh penelitian terdahulu, menunjukkan bahwa DBSCAN bekerja dengan baik untuk data non spasial, dan penelitian tersebut memberikan suatu langkah agar penambangan data menghasilkan *clustering* yang optimal, yaitu jumlah *noise* terkecil, jumlah *cluster* terbanyak, dan SC (*Silhouette Coeficient*) positif. Prototipe awal yang direncanakan, diarahkan dapat mengeksekusi data spasial maupun non spasial, dan mendukung *web* dan *cloud computing*. Bagaimana pun, penelitian ini baru dapat mendefinisikan persyaratan untuk sistem optimalisasi penambangan data *clustering*, sehingga merupakan suatu sintesis.

Model Penelitian



Hasil Penelitian

1. DBSCAN untuk Data Spasial :

A. Converting dimensions ;

- 1) Define non-spatial datasets and the number of Dimensions; D
- 2) Analyze the relationship between variables.
- 3) Sort the coordinates formed based on the tendency of variables based on clustering objectives.
- 4) Normalize the data in the previous step variable (3) ; the normalization scale must be Integer ; I .
- 5) Repeat steps 2-4 so that there are only 3 variables.
- 6) Determine the representation of the variables in the coordinates formed to be the coordinates (X, Y) .
- 7) Normalize the data in the preceding step variable (5) ; the normalization scale must be integer.

B. Determining the radius parameter range, \mathcal{E} range ;

$\mathcal{E}1, \dots, \mathcal{E}4 = (((\text{minPts} - 1) - 1) - 1) - 1, \dots, (\text{minPts} - 1)$ where $\text{minPts} \geq D + 1$

D : number of Dimensions ; number of tabular data attributes or variables of non-spatial data.

minPts : number of neighborhood density thresholds.

2. DBSCAN untuk Data Non-Spasial :

DBSCAN-rpr (radius parameter range) is :

$\mathcal{E} = \text{minPts} - 1 \quad \mathcal{E} \leq \text{minPts} - 1 ; \mathcal{E}1, \dots, \mathcal{E}4$

$$\mathcal{E}1, \dots, \mathcal{E}4 \begin{cases} \mathcal{E}_1 = (((\text{minPts} - 1) - 1) - 1) - 1 = \text{minPts} - 4 \\ \mathcal{E}_2 = (((\text{minPts} - 1) - 1) - 1) = \text{minPts} - 3 \\ \mathcal{E}_3 = ((\text{minPts} - 1) - 1) = \text{minPts} - 2 \\ \mathcal{E}_4 = (\text{minPts} - 1) = \text{minPts} - 1 \end{cases}$$

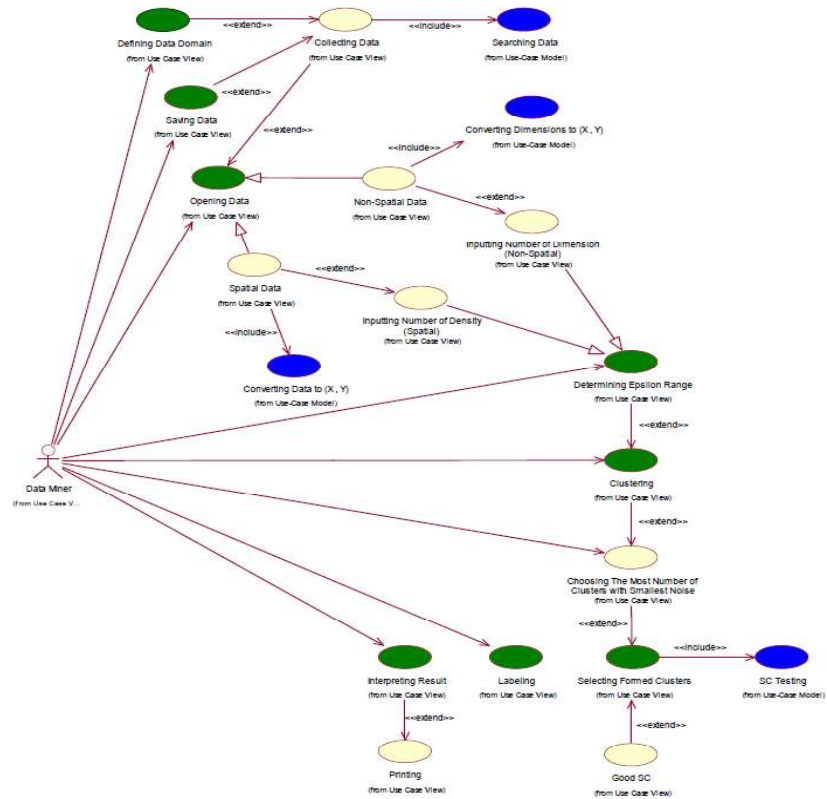
where $\text{minPts} \geq D + 1$; $D = \begin{cases} \text{Density}; & \text{for spatial data} \\ \text{Dimension}; & \text{for non - spatial data} \end{cases}$

Density : amount of data per 1 unit square coordinates; taken the minimum for this algorithm.

Dimension : number of the tabular data attributes or variables of non-spatial data.

Hasil Penelitian

3. Sintesis Perangkat Lunak :



Hasil Penelitian

4. Sintesis Clustering Data Kategorik :

For instance, $X(X1, X2, X3, X4) = \text{Green}$, $Y(Y1, Y2, Y3, Y4) = \text{Blue}$, and $Z(Z1, Z2, Z3, Z4) = \text{Purple}$;
Then :

Table 1 : $Z(Y(X1)) = (Y1, Y2, Y3, Y4, (X1, Y1))$

Z	D					(x, y)
	Y				X	
	Y1	Y2	Y3	Y4	(X1,Y1)	
Z1	Blue	Blue	Blue	Blue	Green	5,1
Z2	White	Blue	Blue	Blue	Green	5,3
Z3	Blue	Blue	Blue	Blue	Green	5,4
Z4	Blue	Blue	White	Blue	Green	5,5
Number of Variables	4				1	

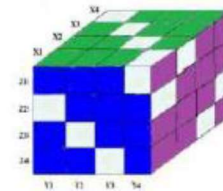


Table 2 : $Z(Y(X4)) = (Y1, Y2, Y3, Y4, (X4, Y1))$

Z	D					(x, y)
	Y				X	
	Y1	Y2	Y3	Y4	(X4,Y1)	
Z1	Blue	Blue	Blue	Blue		5,3
Z2	White	Blue	Blue	Blue		5,3
Z3	Blue	Blue	Blue	Blue		5,3
Z4	Blue	Blue	White	Blue		5,3
Number of Variables	4				1	

Table 4 : $(X1, Y1)$ is uncolored

Z	D					(x, y)
	Y				X	
	Y1	Y2	Y3	Y4	(X1,Y1)	
Z1	Blue	Blue	Blue	Blue		5,1
Z2	White	Blue	Blue	Blue		5,2
Z3	Blue	Blue	Blue	Blue		5,3
Z4	Blue	Blue	White	Blue		5,4
Number of Variables	4				1	

Table 3 : $Z(Y(X4)) = (Y1, Y2, Y3, Y4, (X4, Y4))$

Z	D					(x, y)
	Y				X	
	Y1	Y2	Y3	Y4	(X4,Y4)	
Z1	Blue	Blue	Blue	Blue	Green	5,1
Z2	White	Blue	Blue	Blue	Green	5,3
Z3	Blue	Blue	Blue	Blue	Green	5,4
Z4	Blue	Blue	White	Blue	Green	5,5
Number of Variables	4				1	

Table 5 : $Z1(Y) = Y1, Y2, Y3,$ and $Y4$ are uncolored

Z	D					(x, y)
	Y				X	
	Y1	Y2	Y3	Y4	(X1,Y1)	
Z1	White	White	White	White	Green	5,1
Z2	Blue	Blue	Blue	Blue	Green	5,3
Z3	Blue	Blue	Blue	Blue	Green	5,4
Z4	Blue	Blue	White	Blue	Green	5,5
Number of Variables	4				1	

Referensi :

1. S.A.D. Budiman, D. Safitri, and D. Ispriyanti, "Perbandingan metode K-Means dan metode DBSCAN pada pengelompokan rumah kos mahasiswa di kelurahan Tembalang, Semarang", *Gaussian Journal*, vol. 5(4), pp. 757-762, 2016, ISSN : 2339-2541.
2. E. Arriyanti, I. Arfyanti, dan P. Adytia, "Pengenalan Pola Data *Noise* dengan Pendekatan *Chaos* untuk Optimalisasi Penambangan Data," Hibah PDP 2019.
3. C. Seiffert, T.M. Khoshgoftaar, J.V. Hulse, and A. Folleco, "An empirical study of the classification performance of learners on imbalanced and noisy software quality data", *Elsevier Journal On Information Sciences*, vol. 259, pp. 571-595, 2014, ISSN : 0020-0255, doi : 10.1016/j.ins.2010.12.016.
4. E. Arriyanti, I. Arfyanti and P. Adytia, "Spatial Coordinate Trial : Converting Non-Spatial Data Dimension for DBSCAN," 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Bandung, Indonesia, 2019, pp. 223-228, doi : 10.23919/EECSI48112.2019.8977130.
5. P. Perner, "Data mining-concepts and techniques", *researchgate.net*, article 220633275, 2015.
6. A.D. Ciaccio and G.M. Giorgi, "Statistics in the big data era", *Rivista Italiana di Economia Demografia e Statistica*, vol. LXX (4), Oct.-Dec. 2016.
7. A. Shlapentokh, "Defining integers", *Bulletin of Symbolic Logic* 17(2), Aug. 2010.
8. "Data mining", Article ID 047122854.pdf, accessed Sep., 2020.
9. J. Han, M. Kamber, and J. Pei, "Data mining : concepts and techniques", 3rd ed., pp. 445-447, the Morgan Kaufmann series, Elsevier, 2012.
10. J. Han, M.Kamber, and J. Pei, "Data mining : concepts and techniques", 2nd ed., pp. 110-112, the Morgan Kaufmann series, Elsevier, 2006.