



InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

Available online at : <http://bit.ly/InfoTekJar>
ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Machine Learning

Identifikasi Predikat Hasil Pengelompokan Data Kualitas Udara dengan Menggunakan *Affinity Propagation* dan *Silhouette Coefficient*

Maulidya Rahmah¹, Ade Candra², Rahmat Widia Sembiring³

^{1,2} Program Studi S2 Teknik Informatika, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Medan, Indonesia

³ Politeknik Negeri Medan, Medan, Indonesia

KEYWORDS

Clustering, Affinity Propagation, Silhouette Coefficient, Kualitas Udara

CORRESPONDENCE

Phone: +6281362299775

E-mail: ade_candra@usu.ac.id

A B S T R A C T

Penelitian ini bertujuan untuk mengidentifikasi pengelompokan data kualitas udara dan mendapatkan predikat hasil pengelompokan kualitas data udara tersebut. Data yang digunakan adalah data kualitas udara Kota Pekanbaru yang diperoleh dari pengolahan data Laboratorium Udara Pemerintah Kota Pekanbaru dengan rentang waktu tahun 2014, 2015, dan 2016. Penulis menerapkan *Affinity Propagation* untuk melakukan klusterisasi pada data tersebut dan menghitung jumlah kluster yang dihasilkan. Berdasarkan penerapan *Affinity Propagation* pada data kualitas udara Kota Pekanbaru dengan pengujian nilai *damping factor* dari rentang 0.5 sampai 0.95 diperoleh 5 kluster saat *damping factor* bernilai 0.95. Sementara jumlah kluster terbanyak adalah 156 saat *damping factor* bernilai 0.55. Nilai rata-rata *Silhouette Coefficient* dari data yang diujikan adalah 0.264, dan jika dikategorikan maka kualitas kluster yang dihasilkan berdasarkan predikat nilai *Silhouette Coefficient* adalah *Weak Structure*.

PENDAHULUAN

Peraturan Pemerintah No. 41 Tahun 1999 secara prinsip menyatakan bahwa udara merupakan sumber daya alam yang mempengaruhi makhluk hidup. Faktanya sektor transportasi dan industri menyumbang sebagian besar polusi udara yang berdampak pada penurunan tingkat kualitas udara sehingga dapat menyebabkan terjadinya pencemaran udara [1]. Pencemaran udara merupakan faktor terbesar yang mempengaruhi penurunan kualitas udara. Proses pencemaran sendiri masuk melalui kumpulan zat atau komponen asing yang tentunya diakibatkan oleh ulah manusia [2].

Penurunan kualitas udara tersebut tentunya berdampak negatif terhadap aktivitas manusia yang erat kaitannya dengan kesehatan dan kerusakan lingkungan. Maka sangat penting melakukan pengukuran tingkat kualitas udara untuk mengetahui baik atau tidaknya kualitas udara tersebut bagi aktivitas kehidupan manusia. Kualitas udara yang terus memburuk akan berdampak negatif dalam jangka panjang. Salah satu dampaknya adalah penipisan lapisan ozon yang memicu pemanasan global. Hal ini tentunya sangat mengganggu aktivitas manusia [3]. Selain itu kebakaran hutan dan peningkatan populasi kendaraan bermotor memicu peningkatan kadar CO₂ yang berkontribusi terhadap penurunan kualitas udara.

Pemantauan kualitas udara di Indonesia yang dilakukan oleh Kementerian Lingkungan Hidup pada 10 kota seperti Jakarta, Bandung, Semarang, Surabaya, Denpasar, Medan, Pekanbaru, Palangkaraya, Jambi, dan Pontianak bertujuan untuk melakukan tindakan pencegahan dari dampak negatif kualitas udara yang termonitoring melalui *Air Quality Management System* (AQMS) [4].

Penentuan tingkat kualitas udara dapat dilakukan dengan menggunakan teknik *clustering* yang merupakan salah satu dari sekian banyaknya sub bidang dalam kajian *Data Mining*. Analisis terhadap proses *clustering* sering diterapkan terhadap data yang pengelompokannya dapat dijalankan tetapi tersebut belum dilakukan proses pendefinisian [5].

Clustering merupakan bagian dari metode pembelajaran tak terawasi (*unsupervised learning*) karena proses untuk mendapatkan *cluster* tidak memerlukan pendefinisian *cluster* tersebut terlebih dahulu. Pada umumnya pengukuran kemiripan antar objek yang diujikan pada suatu *cluster* dapat dilakukan dengan perhitungan jarak seperti *Euclidean Distance*, *Manhattan Distance* dan *Minkowski Distance* [6].

Clustering memiliki prinsip melakukan maksimalisasi berdasarkan kesamaan dari kumpulan antar anggota dalam suatu kelas dan melakukan proses minimalisasi dari kesamaan antar *cluster* yang berbeda pada suatu data [7]. Maka poin penting dari tujuan *clustering* secara umum adalah minimalisasi keanggotaan *cluster* dalam suatu data berdasarkan variasi antar *cluster* [8].

Salah satu metode untuk proses *clustering* adalah *Affinity Propagation* yang diperkenalkan oleh Brendan J. Frey dan Delbert Duek pada tahun 2007 yang memiliki perbedaan mendasar dengan metode *clustering* lain seperti *K-Means Clustering* yang mengambil titik pusat *cluster* secara acak. *Affinity Propagation* menganggap semua titik data sebagai calon titik pusat (*exemplar*) yang potensial [9].

Beberapa peneliti telah mempublikasikan riset mereka yang menerapkan *Affinity Propagation* untuk *clustering*. Amijaya *et al.* (2018) meneliti penerapan *Affinity Propagation* untuk pengelompokan titik panas bumi. Hasil penelitian tersebut menunjukkan bahwa pengelompokan data titik panas bumi dengan *Affinity Propagation* dapat digunakan untuk melihat potensi kebakaran berdasarkan rata-rata nilai *confidence* dari setiap *cluster* yang dihasilkan. Jumlah *cluster* yang dihasilkan adalah 44 *cluster* dari 125 data yang diujikan dengan nilai *Silhouette Coefficient* sebesar 0.317818 [10].

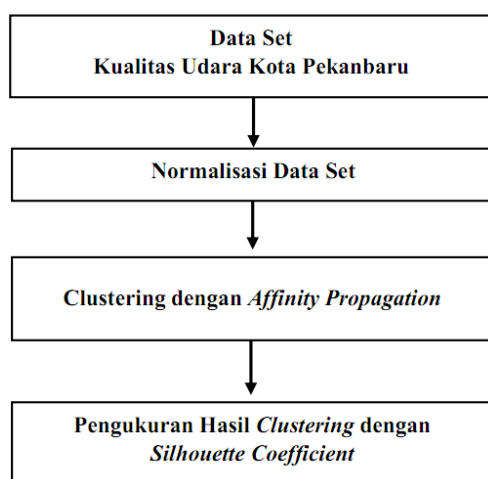
Penelitian Dermawan *et.al* (2019) menerapkan *Affinity Propagation* dalam peringkasan teks untuk deteksi kejadian pada dokumen *twitter* berbahasa Indonesia. Hasil penelitian tersebut menunjukkan bahwa penerapan *Affinity Propagation* dapat mengurangi jumlah *tweet* dari 32.256 menjadi 26.115 untuk data tematik dan dari 102.214 menjadi 34.926 untuk data umum [11].

Berdasarkan penjelasan yang telah dipaparkan maka penulis menerapkan *Affinity Propagation* dalam proses identifikasi pengelompokan data kualitas udara kota Pekanbaru dan melihat predikat dari *cluster* yang dihasilkan berdasarkan *Silhouette Coefficient*.

METODOLOGI PENELITIAN

Tahapan Penelitian

Tahapan dari metode yang digunakan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Semua tahapan penelitian pada Gambar 1 akan dijelaskan sebagai berikut:

Dataset

Data Kualitas Udara Kota Pekanbaru digunakan sebagai data pengujian pada penelitian ini. Adapun data set tersebut diperoleh dari Pengolahan Data Laboratorium Udara Pemerintah Kota

Pekan Baru dengan rentang waktu Tahun 2014, 2015 dan 2016, dan merupakan hasil pengumpulan data oleh Laboratorium Udara Pemerintah Kota Pekanbaru tentang Ketentuan Kualitas Udara Kota Pekanbaru [12].

Data tersebut digolongkan dalam enam kategori. Data tersebut berjumlah sebanyak 1.080 data dengan 5 atribut. Adapun rincian informasi atribut pada Data Kualitas Udara Kota Pekanbaru yaitu pada Tabel 1.

Tabel 1. Atribut Data Kualitas Udara Kota Pekanbaru

No.	Atribut	Penanda
1.	PM10	X1
2.	SO2	X2
3.	CO	X3
4.	O3	X4
5.	NO2	X5
6.	Kategori	-

Kemudian rincian dari Data Kualitas Udara Kota Pekanbaru yaitu pada Tabel 2 berikut:

Tabel 2. Data Kualitas Udara Kota Pekanbaru

No.	PM10	SO2	CO	O3	NO2	Kategori
1	47	51	8	67	2	SEDANG
2	48	51	9	37	2	SEDANG
3	37	51	9	26	2	SEDANG
4	24	50	2	51	1	SEDANG
5	25	50	3	36	1	BAIK
6	18	50	3	53	2	SEDANG
7	15	50	6	61	1	SEDANG
8	20	50	4	36	1	BAIK
9	27	50	7	36	2	BAIK
10	34	50	5	36	2	BAIK
11	31	50	4	40	1	BAIK
12	11	50	3	21	2	BAIK
13	11	50	4	36	1	BAIK
14	20	50	3	36	1	BAIK
15	35	50	2	76	2	SEDANG
16	36	50	6	63	2	SEDANG
17	37	50	4	66	3	SEDANG
18	48	50	5	66	3	SEDANG
19	38	50	6	65	3	SEDANG
20	60	50	13	49	3	SEDANG
21	33	49	6	74	5	SEDANG
22	41	50	8	79	7	SEDANG
23	58	49	8	77	8	SEDANG
...	...	...	...	...	...	...
1070	27	7	6	15	5	BAIK
1071	26	7	6	14	5	BAIK
1072	28	8	6	16	6	BAIK
1073	34	11	8	22	6	BAIK
1074	31	10	8	21	6	BAIK
1075	30	9	7	18	5	BAIK
1076	33	11	7	24	6	BAIK
1077	32	12	6	21	6	BAIK
1078	40	13	9	32	8	BAIK
1079	33	10	7	21	7	BAIK
1080	37	11	7	23	7	BAIK

Kemudian untuk jumlah data pada masing-masing kategori dapat dilihat pada Tabel 3.

Tabel 3. Jumlah Data Berdasarkan Kategori

No.	Kategori	Jumlah Data
1.	SEDANG	237
2.	BAIK	740
3.	TIDAK SEHAT	61
4.	SANGAT TIDAK SEHAT	8
5.	BERBAHAYA	33
6.	TIDAK ADA DATA	1
Jumlah		1080

Normalisasi Data

Normalisasi data merupakan bagian dari *Preprocessing* yang bertujuan untuk mengolah data agar diperoleh data yang relevan serta mengurangi *noise* pada data. Tahap normalisasi data dilakukan dengan menggunakan Normalisasi Z-Score dengan persamaan berikut [13]:

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan:

- z : Nilai *Standard Score*
- x : Data Observasi
- μ : *Mean* Per Variabel
- σ : Standar Deviasi Per Variabel.

Clustering dengan Affinity Propagation

Tahapan perhitungan *Affinity Propagation* dimulai dengan mengukur kesamaan antara pasangan titik data dan membuat *passing messages* bernilai *real* antara titik-titik data sampai *exemplar* berkualitas tinggi ditemukan dan *cluster* yang sesuai ditemukan.

Affinity Propagation mengansumsikan semua titik data sebagai pusat *cluster* dan tidak tergantung pada kualitas set awal pusat *cluster*. Langkah-langkah dalam perhitungan *Affinity Propagation* adalah sebagai berikut [14]:

Asumsi $x_1 \dots x_n$ menjadi satu set titik data dan S menjadi fungsi yang mengkuantifikasikan kesamaan antara dua titik, sehingga:

$$s(X_i, X_j) > s(X_i, X_k) \quad (2)$$

Pada persamaan (2) berlaku ketentuan jika X_i lebih mirip dengan X_j daripada X_k .

Kemudian perhitungan *Affinity Propagation* dimulai dengan langkah *message passing*. Untuk *update* dua matriks yaitu:

1. *Responsibility* matriks R memiliki nilai $r(i, k)$ yang mengukur seberapa cocok X_k untuk dijadikan eksemplar pada X_i .
2. *Availability* matriks A berisi nilai $a(i, k)$ menunjukan seberapa layak untuk X_i untuk menjadikan X_k sebagai eksemplarnya.

Kemudian, kedua matriks tersebut diinisialisasi ke nol. Setelah itu dilakukan pembaruan *iteration* dengan persamaan (3) sebagai berikut:

1. Update *responsibility* seperti dibawah ini:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\} \quad (3)$$

$$r(j, j) \leftarrow s(k, k) - \max\{s(i, k')\} \quad (4)$$

2. Kemudian update *Availability* dari *Responsibility Matrix* dengan persamaan berikut:

$$a(i, k) \leftarrow \min\{0, r(k, k) + \sum_{i' \in \{i, k\}} \max\{0, r(i', k)\}\} \quad (5)$$

$$a(k, k) = \sum_{i' \in \{i, k\}} \max\{0, r(i', k)\} \quad (6)$$

3. Setelah itu *Criterion Matrix* dihitung dengan menjumlahkan nilai *Availability Matrix* dan *Responsibility Matrix* dengan persamaan (7) berikut:

$$c(i, k) = a(i, k) + r(i, k) \quad (7)$$

4. Setelah kondisi *convergence* tercapai dimana tidak terjadi lagi perubahan nilai pada *Responsibility* dan *Availability*, maka proses penentuan *Exemplar* dari setiap titik data i dapat dilakukan dengan persamaan (8). Titik data i akan dijadikan sebagai *Exemplar* jika nilai $k = i$, jika tidak maka titik data i akan dijadikan anggota dari *exemplar* k .

$$k \leftarrow \arg \max\{r(i, k) + a(i, k)\} \quad (8)$$

Pengukuran Hasil Clustering dengan Silhouette Coefficient

Pada penelitian ini pengukuran hasil *clustering* dilakukan menggunakan perhitungan *Silhouette Coefficient*. *Silhouette Coefficient* merupakan hasil kombinasi dari perhitungan *Cohesion* dan *Separation*. *Cohesion* mengukur sedekat apa relasi antar satu objek dengan objek lainnya pada sebuah *cluster*, sedangkan *Separation* menghitung seberapa jauh sebuah *cluster* berpisah dari *cluster* yang lain atau sejauh apa sebuah *cluster* dengan *cluster* yang lainnya [15].

Rentang hasil perhitungan *Silhouette Coefficient* berada pada nilai -1 sampai 1. Nilai *Silhouette Coefficient* memperoleh kategori kualitas semakin baik apabila nilai yang dihasilkan mendekati nilai 1. Sebaliknya jika mendekati angka -1, maka kualitas *cluster* yang dihasilkan akan semakin buruk.

Secara umum, penghitungan untuk menghitung nilai *Silhouette Coefficient* adalah sebagai berikut:

$$s(i) = \frac{(b(x_i) - a(x_i))}{\max(a(x_i), b(x_i))} \quad (9)$$

Pada persamaan (9), nilai rata-rata *Silhouette Coefficient* dari setiap data atau objek pada sebuah *cluster* adalah merupakan suatu takaran yang menunjukkan seketat apa objek-objek dimasukkan pada *cluster* tertentu. Berikut rincian predikat kategori nilai *Silhouette Coefficient* menurut Kaufman dan Rousseeuw [15].

Tabel 4. Predikat Kategori Nilai *Silhouette Coefficient*

No.	Rentang Nilai SC	Kategori
1.	$0.7 < SC \leq 1$	<i>Strong Structure</i>
2.	$0.5 < SC \leq 0.7$	<i>Medium Structure</i>
3.	$0.25 < SC \leq 0.5$	<i>Weak Structure</i>
4.	$SC \leq 0.25$	<i>No Structure</i>

HASIL DAN PEMBAHASAN

Hasil Penelitian

Pada penelitian ini, penerapan *Affinity Propagation* untuk proses identifikasi pengelompokan data kualitas udara dan juga untuk mengukur kualitas *cluster* yang dihasilkan dengan *Silhouette Coefficient*, penulis menggunakan bantuan *Tools Jupyter Notebook* dengan bahasa pemrograman *Python* agar mempermudah dan mempersingkat waktu pengujian,

Langkah pertama yaitu melakukan *preprocessing* untuk menormalisasi data set yang digunakan agar diperoleh data yang valid sebelum dilakukan proses *Clustering*. Normalisasi data dilakukan menggunakan perhitungan normalisasi *Z-Score* sesuai dengan persamaan (1) Adapun hasil normalisasi data pada Data Kualitas Udara Kota Pekanbaru yaitu pada Tabel 5 berikut.

Tabel 5. Hasil Normalisasi Data Kualitas Udara Kota Pekanbaru

No.	PM10	SO2	CO	O3	NO2	Kategori
1	-0.139	2.804	-0.197	1.043	-0.864	SEDANG
2	-0.129	2.804	-0.121	-0.06	-0.864	SEDANG
3	-0.234	2.804	-0.121	-0.46	-0.864	SEDANG
4	-0.357	2.722	-0.654	0.457	-0.919	SEDANG
5	-0.347	2.722	-0.578	-0.09	-0.919	BAIK
6	-0.414	2.722	-0.578	0.53	-0.864	SEDANG
7	-0.442	2.722	-0.349	0.823	-0.919	SEDANG
8	-0.395	2.722	-0.502	-0.09	-0.919	BAIK
9	-0.328	2.722	-0.273	-0.09	-0.864	BAIK
10	-0.262	2.722	-0.425	-0.09	-0.864	BAIK
11	-0.29	2.722	-0.502	0.054	-0.919	BAIK
12	-0.48	2.722	-0.578	-0.64	-0.864	BAIK
13	-0.48	2.722	-0.502	-0.09	-0.919	BAIK
14	-0.395	2.722	-0.578	-0.09	-0.919	BAIK
15	-0.253	2.722	-0.654	1.373	-0.864	SEDANG
16	-0.243	2.722	-0.349	0.896	-0.864	SEDANG
17	-0.234	2.722	-0.502	1.006	-0.81	SEDANG
18	-0.129	2.722	-0.425	1.006	-0.81	SEDANG
19	-0.224	2.722	-0.349	0.97	-0.81	SEDANG
20	-0.016	2.722	0.1844	0.384	-0.81	SEDANG
21	-0.272	2.64	-0.349	1.299	-0.701	SEDANG
22	-0.196	2.722	-0.197	1.482	-0.593	SEDANG
23	-0.035	2.64	-0.197	1.409	-0.539	SEDANG
24	-0.054	2.64	-0.121	2.618	-0.376	TIDAK SEHAT
25	-0.092	2.64	0.0319	1.849	-0.05	SEDANG
26	-0.073	2.64	0.3368	1.885	0.1672	SEDANG
27	-0.129	2.64	0.1844	2.398	1.253	TIDAK SEHAT
28	-0.073	2.722	0.2606	1.006	0.2215	SEDANG
29	-0.243	2.722	-0.349	0.713	0.0043	SEDANG
30	-0.092	2.722	0.0319	1.336	1.1444	SEDANG
...	...	...	...	...	...	...
1068	-0.253	-0.57	-0.273	-0.57	-0.593	BAIK
1069	-0.309	-0.81	-0.349	-0.79	-0.701	BAIK
1070	-0.328	-0.81	-0.349	-0.86	-0.701	BAIK
1071	-0.338	-0.81	-0.349	-0.9	-0.701	BAIK
1072	-0.319	-0.73	-0.349	-0.83	-0.647	BAIK
1073	-0.262	-0.48	-0.197	-0.61	-0.647	BAIK
1074	-0.29	-0.57	-0.197	-0.64	-0.647	BAIK
1075	-0.3	-0.65	-0.273	-0.75	-0.701	BAIK
1076	-0.272	-0.48	-0.273	-0.53	-0.647	BAIK
1077	-0.281	-0.4	-0.349	-0.64	-0.647	BAIK
1078	-0.205	-0.32	-0.121	-0.24	-0.539	BAIK
1079	-0.272	-0.57	-0.273	-0.64	-0.593	BAIK
1080	-0.234	-0.48	-0.273	-0.57	-0.593	BAIK

Setelah diperoleh hasil normalisasi dari Data Kualitas Udara Kota Pekanbaru, kemudian langkah selanjutnya yaitu melakukan identifikasi pengelompokan data dengan *Affinity Propagation*.

Pada pengujian kualitas terhadap kinerja dari *Affinity Propagation* pada Data Kualitas Udara Kota Pekanbaru, pengukuran kualitas dilakukan dengan perhitungan nilai *Silhouette Coefficient* yang dihasilkan dengan pengujian menggunakan berbagai nilai faktor *damping* (λ). Kemudian nilai faktor *damping* (λ) yang akan digunakan bervariasi sesuai dengan ketentuan baku dari penentuan nilai faktor *damping* (λ) pada perhitungan *Silhouette Coefficient* yaitu:

$$0.5 \leq \text{damping } (\lambda) < 1$$

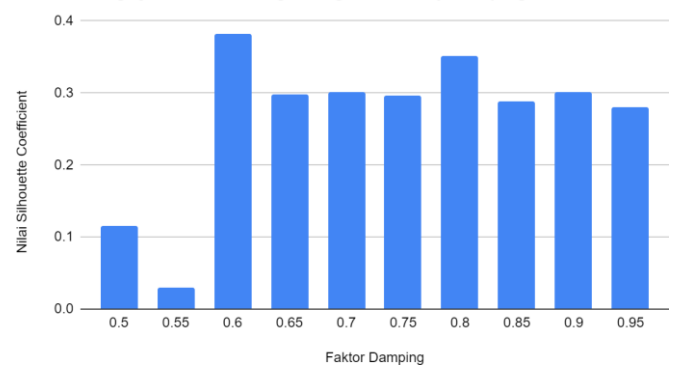
Adapun hasil pengujian *clustering* terhadap Data Kualitas Udara Kota Pekanbaru yaitu pada Tabel 4.6 berikut.

Tabel 6. Hasil Pengujian *Clustering*

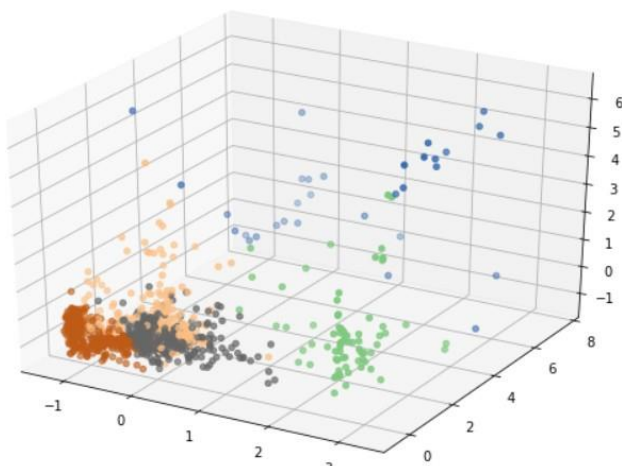
No.	Faktor <i>Damping</i> (λ)	Jumlah <i>Cluster</i>	Nilai <i>Silhouette Coefficient</i>
1	0.50	76	0.115
2	0.55	156	0.030
3	0.60	7	0.382
4	0.65	7	0.298
5	0.70	7	0.301
6	0.75	8	0.297
7	0.80	7	0.352
8	0.85	7	0.289
9	0.90	7	0.302
10	0.95	5	0.280
Rata-Rata <i>Silhouette Coefficient</i>			0.264

Kemudian dari hasil pengujian pada Tabel 6 juga disajikan dalam bentuk grafik seperti pada Gambar 2.

Hasil Pengujian *Clustering* dengan *Affinity Propagation*



Gambar 2. Grafik Hasil Pengujian



Gambar 3. Grafik Hasil Pada Nilai Faktor Damping 0.95

Pembahasan

Dari Tabel 6 dan Gambar 2 dapat dilihat bahwa hasil pengujian yang telah dilakukan untuk proses *clustering* Data Kualitas Udara Kota Pekanbaru dengan pengujian berdasarkan nilai faktor *damping* dari rentang nilai 0.5 sampai dengan 0.95. Nilai *Silhouette Coefficient* yang dihasilkan nilai yang beragam dari setiap pengujian faktor *damping*. Jumlah *cluster* yang dihasilkan paling optimal yaitu pada saat nilai faktor *damping* bernilai 0.95 dapat dilihat hasil grafik pada Gambar 3 dan jumlah *cluster* terbanyak pada saat bernilai 0.55 yaitu 156 *cluster*.

Adapun nilai *Silhouette Coefficient* tertinggi yaitu pada saat nilai faktor *damping* bernilai 0.60 yaitu sebesar 0.382, dan nilai *Silhouette Coefficient* terendah yaitu pada saat nilai faktor *damping* bernilai 0.50 yaitu sebesar 0.115. Kemudian rata-rata dari seluruh nilai *Silhouette Coefficient* yang dihasilkan yaitu 0.264.

Untuk kategori predikat kualitas dari hasil pengujian yang diperoleh jika diukur berdasarkan predikat kategori nilai *Silhouette Coefficient* pada Tabel 4 sebelumnya, maka dapat disimpulkan bahwa predikat kualitas dari hasil pengujian Data Kualitas Udara Kota Pekanbaru dengan *Affinity Propagation* dengan nilai rata-rata *Silhouette Coefficient* sebesar 0.264 yaitu dengan kategori *Weak Structure*.

KESIMPULAN

Berdasarkan hasil akhir dari pengujian yang diperoleh dalam penerapan *Affinity Propagation* pada Data Kualitas Udara Kota Pekanbaru maka beberapa hal yang dapat disimpulkan adalah sebagai berikut:

1. *Affinity Propagation* mampu dalam melakukan pengelompokan pada Data Kualitas Udara Kota Pekanbaru dengan memperoleh jumlah *cluster* yang bervariasi berdasarkan pengujian dari berbagai nilai faktor *damping*.
2. Dari hasil pengujian penelitian yang telah dilakukan, kategori predikat kualitas dari pengujian pada Data Kualitas Udara Kota Pekanbaru berdasarkan nilai rata-rata *Silhouette Coefficient* yang dihasilkan yaitu 0.264 dan nilai tersebut jika dikelompokkan berdasarkan predikat nilai *Silhouette Coefficient* yaitu dalam predikat *Weak Structure*.
3. Dari pengujian yang telah dilakukan, dapat disimpulkan bahwa pengujian dengan nilai faktor *damping* yang bervariasi berpengaruh terhadap perolehan jumlah *cluster* serta perolehan nilai *Silhouette Coefficient* yang akan dihasilkan.

REFERENSI

- [1] Peraturan Pemerintah No. 41 Tahun 1999.
- [2] A. T. Tugaswati, "Review of Air Pollution and Its Health Impact in Indonesia," *Environmental Research*, pp. 95-100, 1991.
- [3] A. Kurniawan, "Pengukuran Parameter Kualitas Udara (CO, NO₂, O₃, dan PM₁₀) di Bukit Kototabang Berbasis ISPU," *Teknosains*, vol. 7, no. 1, pp. 1-13, 2017.
- [4] E. R. Zulfima. "Visualisasi Dari Klasterisasi dan Peramalan Kualitas Udara Kota Surabaya Menggunakan Metode Klasterisasi K-Means dan Peramalan Artificial Neural Network". *Skripsi, Institut Teknologi Sepuluh Nopember*, 2017.
- [5] P. N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining," vol. 1, Pearson Addison Wesley: Boston, 2006.
- [6] Nisha and P. J. Kaur, "Cluster Quality Based Performance Evaluation of Hierarchical Clustering Method," *Proceedings of 2015 1st Internatinal Conference on Next Computing Technologies*, pp. 649-653, 2015.
- [7] D. Xia, F. Wu, X. Zhang, and Y. Zhuang, "Local and Global Approaches of Affinity Propagation Clustering for Large Scale Data," *Journal of Zhejiang University Science A*, vol. 9, no. 10, pp. 1373-1381, 2008.
- [8] Y. Agusta, "K-Means - Penerapan, Permasalahan dan Metode Terkait," *Jurnal Sistem dan Informatika*, vol. 3, no. 1, pp. 47-60, 2007.
- [9] B. J. Frey, D. Dueck, "Clustering by Passing Messages Between Data Points," *Science*, vol. 315, no. 5814, pp. 972-976, 2007.
- [10] B. K. Amijaya, M. T. Furqon, C. Dewi, "Clustering Titik Panas Bumi Menggunakan Algoritme Affinity Propagation," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 1, pp. 47-60, 2018.
- [11] R. Dermawan, F. A. Bachtiar, and P. P. Adikara, "Peringkasan Teks Untuk Deteksi Kejadian Pada Dokumen Twitter Berbahasa Indonesia Dengan Metode Affinity Propagation," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2018.
- [12] K. L. Sitompul, M. Zarlis, and P. Sihombing, "Increased accuracy in the classification method of backpropagation neural network using principal component analysis," *In IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, pp. 1-5, 2020.
- [13] S. Patro, and K. Sahu, "Normalization: A Preprocessing Stage," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 5, no. 1, pp. 20-22, 2015.
- [14] D. Delbert, "Affinity Propagation: Clustering Data by Passing Messages," *Thesis Projects, Department of Electrical & Computer Engineering University of Toronto*, 2009.
- [15] D. F. Azuri, Zulhanif, and R. S. Pontoh, "Pengelompokkan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means," *Prosiding Seminar Nasional MIPA 2016*.
- [16] M. Anggara, H. Sujiani, and H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member di Alvaro Fitness," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 1, no. 1, 2016.