

PERBANDINGAN PERFORMA ALGORITMA K-MEANS, K-MEDOIDS, DAN DBSCAN DALAM PENGGEROMBOLAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT

Ferista Wahyu Saputri ^{a,1*}, Dede Brahma Arianto^{b,2}

^a IPB University, Bogor

^b Magister Informatika, Universitas Islam Indonesia, Yogyakarta

¹ ferista4602wahyu@apps.ipb.ac.id*; ² dede.brahma2@gmail.com;

* corresponding author

ARTICLE INFO

Keywords
DBSCAN
K-Means
K-Medoids
Public welfare
t-SNE

ABSTRACT

One of the development orientations in Indonesia is to improve the welfare of society. Therefore, it is important to identify and understand the characteristics of community welfare in each province in order to determine effective and targeted development strategies. Cluster analysis is one of the analyses that can be used to group provinces in Indonesia that have homogeneous characteristics within a cluster. The partition method is the simplest and fundamental approach to cluster analysis, but it can only find clusters with spherical-shaped forms. On the other hand, DBSCAN is a density-based clustering algorithm that can be used to find clusters with arbitrary shapes. In this study, the performance of the K-Means, K-Medoids, and DBSCAN algorithms was compared using data that had been dimensionally reduced using the t-SNE method. The data used was the indicator data of community welfare in the year 2022. The evaluation results of clustering based on the highest Silhouette coefficient (0.917) and the lowest Davies-Bouldin index (0.089) indicate that the best clustering methods are K-Means and DBSCAN with parameters *perplexity* = 1, *minPts* = 2, and *epsilon* = 9. Both methods produce the same result, which is the formation of eight clusters.

1. Pendahuluan

Memajukan kesejahteraan umum merupakan salah satu tujuan negara Indonesia yang tertuang secara eksplisit dalam alinea keempat pembukaan UUD 1945. Kesejahteraan masyarakat merujuk pada tingkat kesejahteraan ekonomi dan sosial masyarakat yang dikaji dari delapan bidang meliputi kemiskinan, kependudukan, ketenagakerjaan, perumahan dan lingkungan, kesehatan dan gizi, pendidikan, pola konsumsi, serta bidang sosial lainnya yang menjadi acuan dalam upaya peningkatan kualitas hidup [1]. Jumlah penduduk Indonesia mencapai 275.773.800 jiwa tersebar di seluruh wilayah dengan perbedaan kondisi alam, sumber daya manusia, dan infrastruktur, sehingga membuat tingkat kesejahteraan masing-masing wilayah pun berbeda-beda [2].

Mengidentifikasi dan memahami karakteristik kesejahteraan masyarakat di setiap provinsi penting dilakukan untuk menentukan strategi pembangunan. Analisis gerombol dapat digunakan untuk mengelompokkan objek yang relatif homogen berdasarkan karakteristik tertentu ke dalam satu gerombol [3]. Metode partisi adalah pendekatan paling sederhana dan fundamental dari analisis gerombol yang mengorganisir objek-objek dalam sebuah himpunan menjadi beberapa gerombol dengan asumsi jumlah gerombol telah diketahui sebelumnya [4]. Salah satu metode partisi yang umum digunakan adalah algoritma dengan pusat gerombol rata-rata (K-Means). Penelitian sebelumnya oleh [5] menggunakan metode K-Means dengan enam peubah indikator kesejahteraan masyarakat menghasilkan terbentuknya tiga gerombol. Selanjutnya, penelitian [6] menyimpulkan bahwa metode K-Means dianggap lebih baik dibandingkan dengan Average Linkage Clustering karena memiliki nilai

varian yang lebih kecil. Namun, metode K-Means tidak kekar terhadap pencilan [7]. [8] menyebutkan bahwa algoritma K-Medoids dapat digunakan untuk menangani kekurangan K-Means yang mudah dipengaruhi oleh *noise* dan pencilan. Algoritma K-Medoids menggunakan metode *partitioning around medoid* (PAM) untuk memilih objek secara acak sebagai representasi kelompok (*medoid*) hingga semua objek terpilih. Selama proses ini, *medoid* akan diganti dengan titik lain yang memberikan galat yang lebih kecil, hingga galat mencapai nilai minimum [4].

Metode partisi hanya dapat menemukan gerombol dengan bentuk bulat (*spherical-shaped*) dan mengalami kesulitan dalam menemukan gerombol dengan bentuk sembarang (*arbitrary shapes*). DBSCAN merupakan algoritma penggerombolan berdasarkan pada kepadatan (*density*) [4]. Kelebihan algoritma ini adalah tidak harus mengetahui jumlah gerombol sebelumnya, tidak memerlukan asumsi tertentu mengenai kepadatan atau ragam dalam gerombol, dan dapat mendeteksi gerombol dari data berjumlah besar serta mengandung *noise* dan pencilan [9]. Pada penelitian ini, akan membandingkan performa algoritma K-Means, K-Medoids, dan DBSCAN sehingga dapat mengetahui metode terbaik dalam penggerombolan provinsi di Indonesia berdasarkan indikator tingkat kesejahteraan masyarakat tahun 2022. Hasil penelitian ini diharapkan dapat memberikan pemahaman tentang karakteristik kesejahteraan masyarakat di setiap provinsi serta memberikan dasar bagi pengambilan kebijakan pembangunan yang lebih efektif dan tepat sasaran.

2. Metodologi Penelitian

2.1. Bahan dan Data

Data yang digunakan merupakan data sekunder terkait indikator kesejahteraan masyarakat setiap provinsi di Indonesia (34 provinsi) terdiri dari 12 peubah bertipe numerik yang bersumber dari Badan Pusat Statistik dan Ditjen Bina Pembangunan Daerah Kementerian Dalam Negeri pada tahun 2022 [10][11]. Rincian mengenai peubah tersebut dapat dilihat pada Tabel 1. Dalam analisis gerombol, seluruh data sekunder digunakan secara lengkap tanpa melakukan pembagian data sebelumnya.

Tabel 1 Peubah-peubah yang digunakan

Peubah	Keterangan	Satuan
X1	Kepadatan penduduk	Ribu jiwa
X2	Distribusi persentase PDRB atas dasar harga berlaku	-
X3	Tingkat penyelesaian pendidikan SMA	Persen
X4	Pengeluaran per kapita disesuaikan	Ribu rupiah/orang/tahun
X5	Umur harapan hidup	Tahun
X6	Persentase rumah tangga yang memiliki akses terhadap hunian yang layak dan terjangkau	Persen
X7	Tingkat pengangguran terbuka	Persen
X8	Tingkat partisipasi angkatan kerja	Persen
X9	Persentase penduduk miskin	Persen
X10	Indeks pembangunan gender	-
X11	Proporsi rumah tangga dengan akses terhadap pelayanan dasar	Persen
X12	Prevalensi stunting	Persen

2.2. Metode Penelitian

Prosedur analisis data dalam penelitian ini adalah sebagai berikut:

1. Mengeksplorasi data pada setiap peubah untuk melihat gambaran umum data, memeriksa keberadaan pencilan, dan melihat korelasi antar peubah.
2. Melakukan standarisasi peubah menggunakan *Z-Score* yaitu dengan menghitung perbedaan nilai

- suatu observasi pada peubah tertentu dengan nilai rata-rata peubah tersebut, lalu menskalakan dengan standar deviasi peubah tersebut [12].
3. Melakukan reduksi dimensi dengan metode *t-Distributed Stochastic Neighbor Embedding* (t-SNE) menjadi dua dimensi. t-SNE merupakan teknik reduksi dimensi nonlinier yang cocok untuk menyematkan data berdimensi tinggi dan memvisualisasikannya dalam ruang berdimensi rendah [13]. Parameter *perplexity* mengontrol seberapa banyak tetangga yang diperhitungkan saat memetakan data ke dalam ruang baru. Oleh karena amatan pada penelitian ini hanya sedikit maka parameter *perplexity* yang dicoba adalah 1, 2, 3, dan 5.
 4. Menggerombolkan data hasil reduksi dimensi dengan metode K-Means. Tahapan metode K-Means [4]:
 - a. Menetapkan banyaknya gerombol yang akan dibuat (k).
 - b. Menetapkan titik pusat gerombol awal dengan acak.
 - c. Melakukan penghitungan jarak *euclid* pada tiap objek dengan masing-masing titik pusat gerombol.
 - d. Mengalokasikan setiap objek dalam gerombol yang memiliki jarak terdekat.
 - e. Memperbarui posisi titik pusat gerombol dengan melakukan penghitungan rata-rata dari semua objek yang termasuk dalam gerombol tersebut.
 - f. Ulangi dari langkah c hingga tidak ada perubahan objek dalam gerombol.
 5. Menggerombolkan data hasil reduksi dimensi dengan metode K-Medoids. Tahapan metode K-Medoids [4]:
 - a. Menentukan k objek sebagai *medoid* awal.
 - b. Menempatkan setiap objek pada gerombol yang berjarak dekat dengan *medoid* gerombol.
 - c. Melakukan pencarian objek yang memiliki rata-rata jarak terkecil ke semua objek pada gerombol dengan membandingkan semua pasangan objek *medoid* dan *non medoid* sebagai *medoid* baru.
 - d. Mengulangi langkah b sampai c hingga tidak ada perubahan pada *medoid*.
 6. Menggerombolkan data hasil reduksi dimensi dengan metode DBSCAN. Tahapan metode DBSCAN [14]:
 - a. Masukkan dataset (D), parameter radius ketetanggaan (ϵ), dan minimum banyak tetangga (*minPts*) yang diperlukan untuk membentuk sebuah gerombol.
 - b. Tandai semua titik data sebagai belum dikunjungi.
 - c. Ulangi langkah-langkah berikut sampai semua titik data telah dikunjungi.
 - Pilih secara acak sebuah titik data P yang belum dikunjungi dan tandai sebagai telah dikunjungi.
 - Hitung jarak antara titik P dan titik lain menggunakan metode jarak *euclid*.
 - Jika terdapat setidaknya *minPts* titik dalam radius ketetanggaan P maka:
 - Jika P belum termasuk dalam gerombol manapun, buat gerombol baru dan tambahkan P beserta titik-titik dalam radius ketetanggaan P ke dalam gerombol tersebut.
 - Jika P termasuk dalam suatu gerombol, tambahkan titik-titik dalam radius tetanggaan P ke dalam gerombol tersebut. Jika terdapat kurang dari *minPts* titik dalam radius tetanggaan P dan P belum termasuk dalam gerombol manapun, tandai P sebagai *noise*.
 - d. Ulangi langkah b dan c untuk setiap titik dalam radius ketetanggaan sebagai titik P sampai tidak ada lagi titik yang belum dikunjungi dalam radius ketetanggaan.
 7. Mengevaluasi hasil penggerombolan menggunakan koefisien Silhouette dan indeks Davies-Bouldin untuk mengetahui banyaknya gerombol optimal pada metode K-Means dan K-Medoids; serta mengetahui parameter *perplexity*, *minPts*, dan *epsilon* optimal pada metode DBSCAN. Koefisien

Silhouette yang paling mendekati 1 dan indeks Davies-Bouldin yang paling mendekati 0 merupakan penggerombolan yang terbaik. Nilai koefisien Silhouette dapat dihitung menggunakan rumus berikut [15]:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

$$\text{Silhouette} = \frac{1}{N} \sum_{i=1}^N S_i$$

Silhouette = nilai koefisien Silhouette untuk seluruh provinsi

S_i = nilai koefisien Silhouette untuk provinsi ke- i

a_i = rata-rata jarak antara provinsi ke- i dengan objek lain dalam gerombol provinsi ke- i berada

b_i = rata-rata jarak antara provinsi ke- i dengan seluruh provinsi yang berada pada gerombol terdekat

N = jumlah provinsi

- Memilih hasil penggerombolan terbaik dan menginterpretasikan karakteristik setiap gerombol yang dihasilkan.

3. Hasil dan Pembahasan

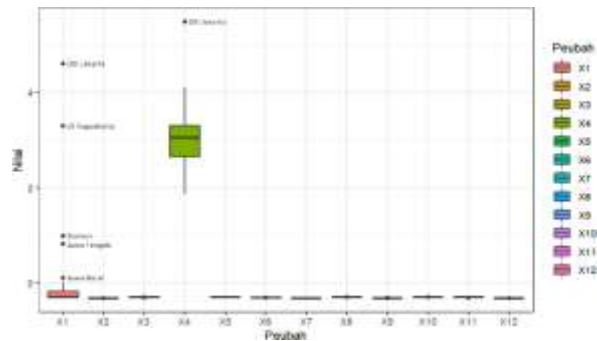
3.1. Eksplorasi Data

Analisis statistik deskriptif pada Tabel 2 dilakukan untuk memperoleh gambaran umum dari setiap indikator yang terkait dengan kesejahteraan rakyat di Indonesia.

Tabel 2 Statistik deskriptif tiap peubah yang digunakan

Peubah	Nilai minimum	Rata-rata	Nilai maksimum	Simpangan baku
X1	4,74	1259,13	16084,1	3401,59
X2	0,25	2,941	16,64	4,08
X3	38,47	64,68	87,92	10,86
X4	7146	11080	18927	2246,73
X5	65,63	70,42	75,08	2,45
X6	27,28	60,06	84,94	12,59
X7	2,34	4,966	8,31	1,6
X8	63,08	68,64	77,75	3,64
X9	4,53	10,299	26,8	5,28
X10	81,04	90,79	94,99	3,14
X11	40,91	78,83	90,82	9,64
X12	1,2	9,312	23,1	5,38

Tabel 2 menunjukkan penyebaran data yang beragam yaitu selisih nilai minimum dan maksimum yang

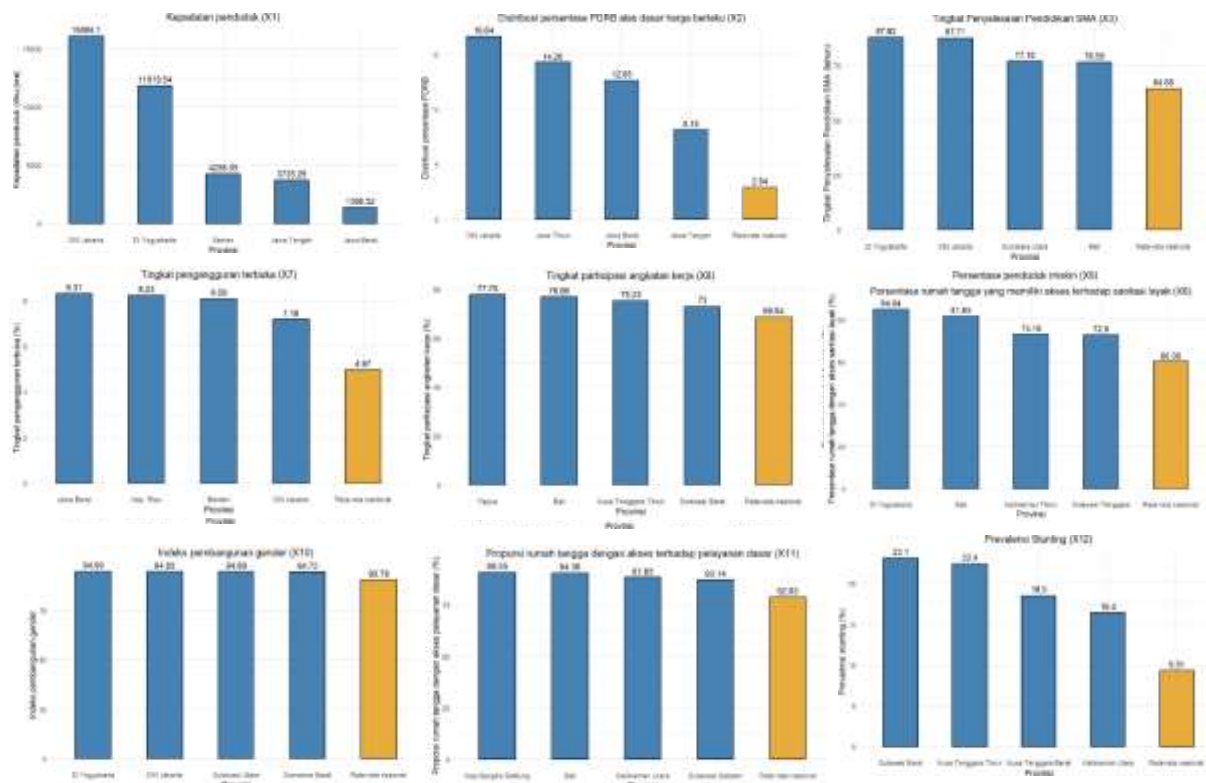


Gambar 1 Boxplot setiap peubah pada data yang sudah distandarisasi

cukup besar pada hampir semua peubah. Selain itu, terdapat perbedaan nilai yang besar antar peubah disebabkan perbedaan satuan dari setiap peubah. Hal tersebut dapat menyebabkan perhitungan jarak pada analisis gerombol menjadi tidak valid karena peubah dengan nilai yang lebih besar akan memiliki pengaruh yang lebih dominan dalam perhitungan jarak, sehingga data perlu distandarisasi sebelum melakukan analisis gerombol. Setelah melakukan standarisasi peubah, dilakukan pendeteksian pencilan secara eksploratif dengan menggunakan boxplot yang ditunjukkan pada Gambar 1.

Pada Gambar 1, dapat diamati bahwa mayoritas peubah memiliki nilai median yang hampir sama dan rentang interkuartil yang sempit, menunjukkan bahwa provinsi di Indonesia memiliki nilai yang relatif sama, kecuali untuk peubah X4. Selain itu, terdapat pencilan atas pada peubah X1 dan X4 yang menunjukkan provinsi-provinsi tersebut memiliki nilai yang jauh lebih besar.

Gambar 2 menampilkan provinsi yang memiliki nilai tertinggi pada setiap indikator kesejahteraan rakyat yang digunakan.



Gambar 2 Provinsi dengan nilai tertinggi pada tiap peubah

Provinsi dengan kepadatan penduduk, distribusi persentase PDRB, umur harapan hidup, dan tingkat pengangguran terbuka tertinggi berada pada wilayah barat Indonesia. Sementara itu, provinsi dengan tingkat penyelesaian pendidikan SMA, pengeluaran per kapita, persentase rumah tangga yang memiliki akses hunian layak dan terjangkau, indeks pembangunan gender, proporsi rumah tangga dengan akses pelayanan dasar, dan prevalensi stunting tertinggi berada di wilayah timur dan barat Indonesia. Sedangkan provinsi dengan tingkat partisipasi angkatan kerja dan persentase penduduk miskin tertinggi berada di wilayah timur. Secara keseluruhan, provinsi dengan peringkat tertinggi yang memiliki masalah ekonomi dan sosial menyebar di wilayah timur, sehingga mengindikasikan tingkat kesejahteraan di wilayah timur Indonesia masih tergolong rendah.

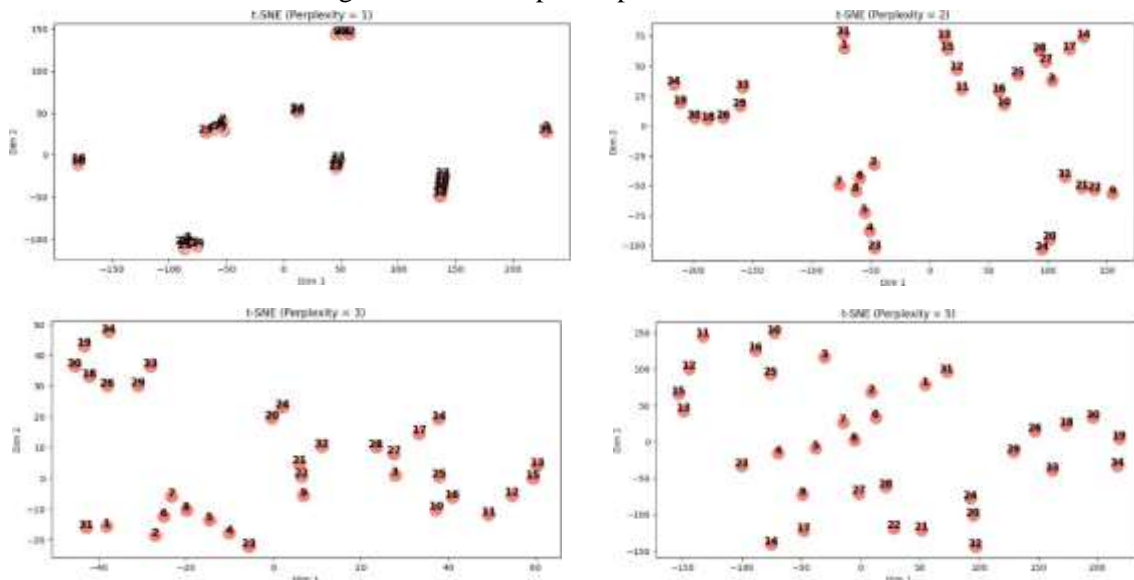
Korelasi Pearson yang lebih besar dari mutlak 0,7 menunjukkan keterkaitan antar peubah yang tinggi [16]. Berdasarkan Gambar 3 terlihat ada beberapa pasangan peubah yang menunjukkan korelasi tinggi.



Gambar 3 Matriks korelasi Pearson antar pasangan peubah

3.2. Reduksi Dimensi

Pada penelitian ini, t-SNE digunakan untuk mereduksi dimensi pada data yang telah distandarisasi dengan tujuan mengurangi pengaruh keterkaitan antar peubah serta mendapatkan gambaran pola atau kelompok dalam data. Parameter *perplexity* pada t-SNE dicobakan dengan beberapa nilai yaitu 1, 2, 3, dan 5. Hasil reduksi dimensi dengan t-SNE ditampilkan pada Gambar 4.



Gambar 4 Scatter plot hasil reduksi dimensi t-SNE

Hasil reduksi dimensi menggunakan t-SNE dengan *perplexity* 1 menunjukkan adanya potensi terbentuk delapan kelompok. Namun, ketika nilai *perplexity* ditingkatkan menjadi 2, 3, dan 5 titik-titik data menjadi cukup menyebar, sehingga sulit untuk menentukan jumlah kelompok yang terbentuk hanya berdasarkan visualisasi tersebut, maka selanjutnya dilakukan analisis gerombol.

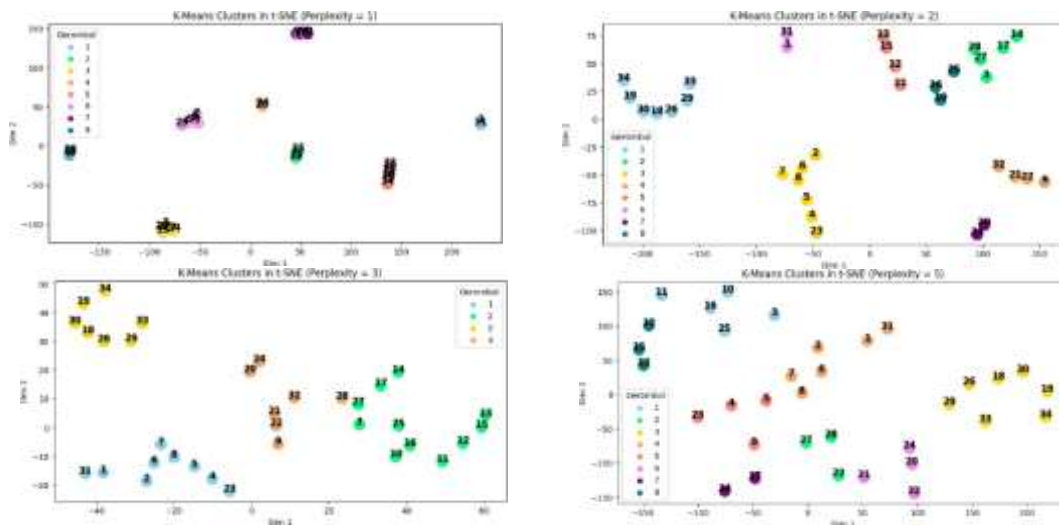
KL Divergence pada t-SNE adalah metrik yang mengukur perbedaan antara distribusi probabilitas data asli dan distribusi probabilitas pada peta dua atau tiga dimensi yang dihasilkan t-SNE [17]. Semakin rendah nilai *KL Divergence*, semakin akurat t-SNE merepresentasikan struktur data dalam ruang dimensi yang lebih rendah. Hasil reduksi dimensi t-SNE dengan parameter *perplexity* 1, 2, 3, dan 5 memiliki nilai *KL Divergence* berturut-turut sebesar 0,29; 0,43; 0,37; dan 0,49.

3.3. Analisis Gerombol dan Evaluasi Hasil Gerombol

Penelitian ini menggunakan metode non-hirarki berbasis partisi yaitu K-Means dan K-Medoids serta metode berbasis kepadatan data yaitu DBSCAN untuk menggerombolkan data indikator kesejahteraan masyarakat hasil reduksi dimensi.

3.3.1 Penggerombolan dengan K-Means

Dalam metode K-Means, penentuan jumlah gerombol (k) merupakan faktor kritis. Untuk menentukan k yang optimal dilakukan iterasi dengan mencoba berbagai nilai k ($k = 1, 2, 3, 4, 5, 6, 7, 8, 9$), kemudian dievaluasi menggunakan koefisien Silhouette serta indeks Davies-Bouldin. Pengelompokan yang optimal akan memiliki nilai koefisien Silhouette tertinggi dan nilai indeks Davies-Bouldin terendah. Hasil penggerombolan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat menggunakan metode K-Means ditunjukkan pada Gambar 5.



Gambar 5 Scatter plot hasil penggerombolan K-Means dengan k optimal yang dievaluasi berdasarkan koefisien Silhouette dan indeks Davies-Bouldin

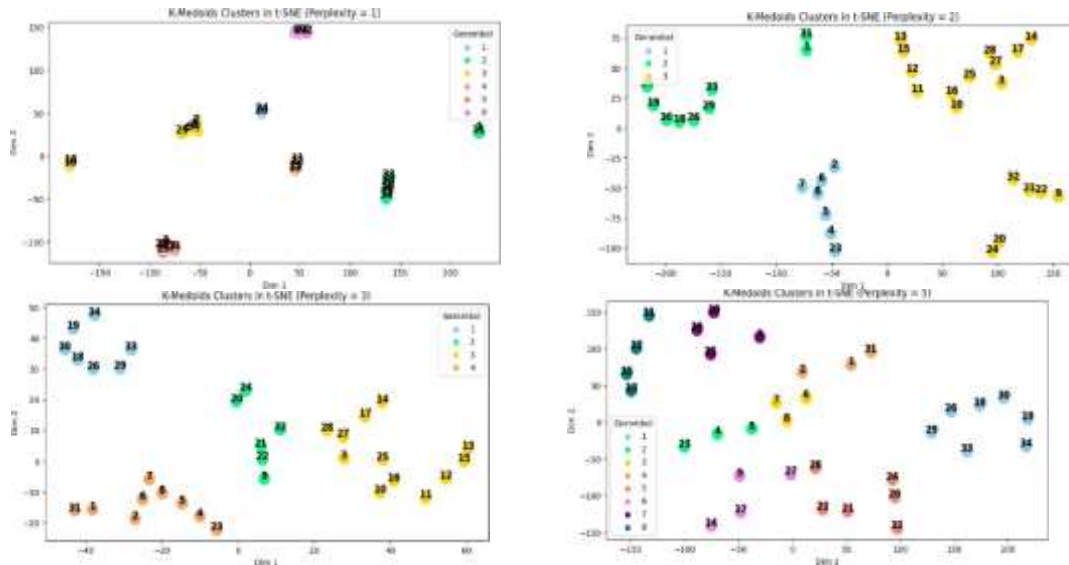
Hasil penggerombolan optimal untuk *perplexity* 1 menunjukkan terbentuknya delapan gerombol yang terpisah dengan baik serta objek-objek dalam satu gerombol saling berdekatan. Hasil penggerombolan untuk *perplexity* 2 dan *perplexity* 5 menunjukkan terbentuk delapan gerombol dan terlihat antar gerombol cukup terpisah dengan baik. Hasil penggerombolan untuk *perplexity* 3 menunjukkan terbentuk empat gerombol, tetapi terdapat beberapa objek yang terletak sedikit lebih jauh dibandingkan objek lainnya dalam satu gerombol. Perbandingan hasil penggerombolan berdasarkan k optimal yang dievaluasi menggunakan koefisien Silhouette dan indeks Davies-Bouldin ditampilkan pada Tabel 3. Terlihat bahwa hasil penggerombolan K-Means terbaik diperoleh saat parameter *perplexity* 1.

Tabel 3 Evaluasi hasil penggerombolan K-Means

t-SNE <i>perplexity</i>	Jumlah gerombol	Koefisien Silhouette	Indeks Davies-Bouldin
1	8	0,917	0,089
2	8	0,647	0,391
3	4	0,551	0,581
5	8	0,439	0,633

3.3.2 Penggerombolan dengan K-Medoids

Seperti pada metode K-Means, untuk menentukan jumlah gerombol optimal (k) dilakukan iterasi kemudian hasil penggerombolan dievaluasi menggunakan koefisien Silhouette dan indeks Davies-Bouldin. Hasil penggerombolan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat dengan metode K-Medoids ditunjukkan pada Gambar 6.



Gambar 6 Scatter plot hasil penggerombolan K-Medoids dengan k optimal yang dievaluasi berdasarkan koefisien Silhouette dan indeks Davies-Bouldin

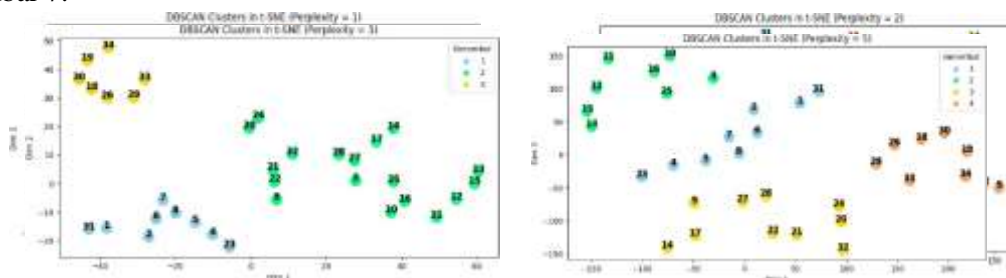
Hasil penggerombolan K-Medoids untuk *perplexity* 1, 2, 3, dan 5 berturut-turut menunjukkan terbentuk enam, tiga, empat, dan delapan gerombol. Terdapat beberapa objek yang terletak lebih jauh dibandingkan objek lainnya dalam satu gerombol, serta terlihat objek-objek menyebar sehingga antar gerombol terlihat berdekatan. Hal tersebut menunjukkan hasil penggerombolan dengan K-Medoids kurang optimal. Perbandingan hasil penggerombolan dapat dilihat pada Tabel 4. Terlihat bahwa hasil penggerombolan K-Medoids terbaik diperoleh pada parameter *perplexity* 1.

Tabel 4 Evaluasi hasil penggerombolan K-Medoids

t-SNE <i>Perplexity</i>	Jumlah gerombol	Koefisien Silhouette	Indeks Davies-Bouldin
1	6	0,702	0,386
2	3	0,538	0,545
3	4	0,556	0,569
5	8	0,422	0,670

3.3.3 Penggerombolan dengan DBSCAN

Parameter yang digunakan dalam metode DBSCAN adalah *minPts* = 2 dan epsilon yang diiterasi dari 1 hingga 100. Penentuan parameter *perplexity* dan *epsilon* yang optimal didasarkan pada nilai koefisien Silhouette tertinggi dan indeks Davies-Bouldin terendah. Hasil penggerombolan provinsi berdasarkan indikator kesejahteraan masyarakat menggunakan DBSCAN dapat dilihat pada Gambar 7.



Gambar 7 Scatter plot hasil penggerombolan DBSCAN berdasarkan epsilon optimal yang dievaluasi dengan koefisien Silhouette dan indeks Davies-Bouldin

Hasil penggerombolan untuk *perplexity* 1, 2, 3, dan 5 sama-sama menunjukkan bahwa tidak ada provinsi yang menjadi *outlier*, jumlah gerombol yang terbentuk berturut-turut adalah delapan, tujuh, tiga, dan empat. Secara umum, terlihat bahwa jarak antar objek dalam satu gerombol sudah cukup dekat. Perbandingan hasil penggerombolan yang disajikan pada Tabel 5. Hasil penggerombolan DBSCAN terbaik berdasarkan *perplexity* dan *epsilon* optimal diperoleh pada parameter *perplexity* = 1, *epsilon* = 9, dan *minPts* = 2.

Tabel 5 Evaluasi hasil penggerombolan DBSCAN untuk epsilon optimal

t-SNE <i>Perplexity</i>	DBSCAN		Jumlah gerombol	Koefisien Silhouette	Indeks Davies-Bouldin
	<i>epsilon</i>	<i>minPts</i>			
1	9	2	8	0,917	0,089
2	26	2	7	0,652	0,385
3	15	2	3	0,579	0,497
5	57	2	4	0,446	0,803

3.3.4 Pemilihan metode penggerombolan terbaik

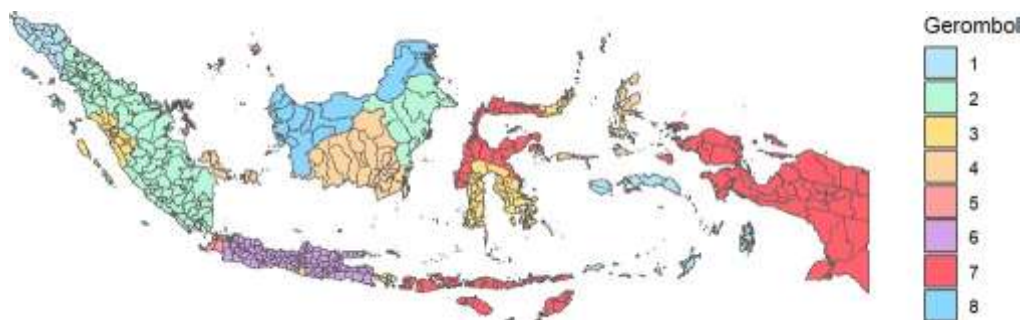
Hasil penggerombolan terbaik dilihat dari hasil di Tabel 3, Tabel 4, dan Tabel 5 adalah penggerombolan dengan metode K-Means dan DBSCAN. Kedua metode tersebut memberikan hasil yang sama yaitu terbentuknya delapan gerombol dengan tidak ada *outlier*. Distribusi tiap anggota gerombol ditampilkan pada Tabel 6.

Tabel 6 Hasil penggerombolan terbaik

Gerombol	Anggota	Anggota gerombol
1	2	Aceh, Maluku
2	7	Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kalimantan Timur
3	6	Sumatera Barat, DI Yogyakarta, Bali, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara
4	4	Kepulauan Bangka Belitung, Kalimantan Tengah, Kalimantan Selatan, Maluku Utara
5	2	Kepulauan Riau, Banten
6	4	DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur
7	7	Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Papua Barat, Papua
8	2	Kalimantan Barat, Kalimantan Utara

Gambar 8 Peta penggerombolan provinsi berdasarkan indikator kesejahteraan masyarakat

3.3.5 Interpretasi Hasil Gerombol

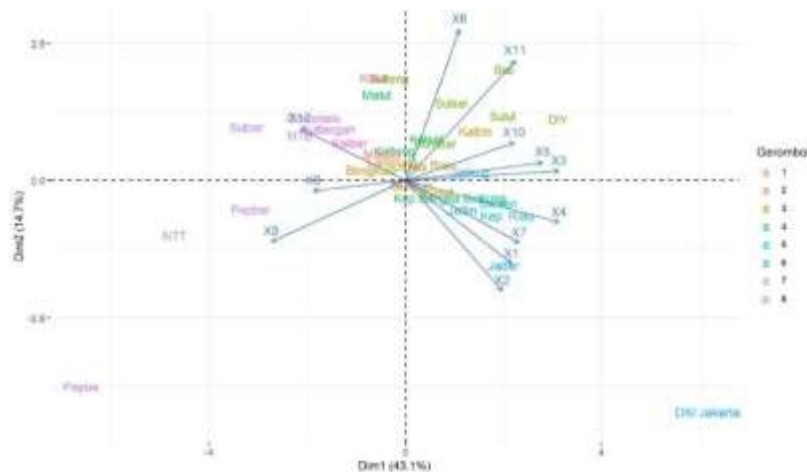


Karakteristik tiap gerombol dapat direpresentasikan berdasarkan nilai rata-rata peubah pada

tiap-tiap gerombol yang dapat dilihat pada Tabel 7.

Tabel 7 Rataan setiap peubah pada masing-masing gerombol

Peubah	Nilai rata-rata peubah tiap gerombol							
	1	2	3	4	5	6	7	8
X1	66,710	121,903	2291,735	60,015	2688,705	5323,653	156,319	33,415
X2	0,695	3,163	1,408	0,805	2,755	12,925	0,787	1,030
X3	71,375	68,327	71,903	65,915	69,975	70,095	49,940	56,600
X4	9419,500	11114,714	11978,500	11420,750	13342,500	13393,250	9078,000	9352,500
X5	68,315	71,240	72,000	69,735	70,445	73,2875	67,186	71,845
X6	62,420	63,843	72,840	51,560	53,835	55,725	52,664	63,975
X7	6,525	4,796	4,937	4,437	8,160	6,637	3,221	4,720
X8	64,480	67,877	69,473	66,760	66,830	67,825	72,051	68,295
X9	15,490	9,577	8,222	5,203	6,135	8,515	17,430	6,835
X10	92,720	90,031	93,857	89,623	92,840	92,410	88,406	87,730
X11	75,310	75,686	86,180	84,668	81,460	81,4525	70,313	81,550
X12	8,950	5,671	7,717	8,9	5,850	6,750	15,100	16,350



Gambar 9 Biplot dengan hasil penggerombolan DBSCAN

Berdasarkan Tabel 6, Tabel 7, dan Gambar 9 diperoleh karakteristik gerombol sebagai berikut:

- Gerombol 1 merupakan provinsi dengan tingkat kesejahteraan rendah dan memerlukan upaya intensif dalam mengatasi masalah ekonomi untuk meningkatkan kualitas hidup penduduknya. Meskipun memiliki kepadatan penduduk dan prevalensi stunting yang rendah serta tingkat penyelesaian pendidikan SMA, persentase rumah tangga dengan hunian layak, dan indeks pembangunan gender yang tinggi, tetapi terdapat beberapa masalah ekonomi. Masalah tersebut meliputi distribusi persentase PDRB, pengeluaran per kapita, umur harapan hidup, tingkat partisipasi angkatan kerja, dan proporsi rumah tangga dengan akses pelayanan dasar yang rendah, serta persentase penduduk miskin dan tingkat pengangguran terbuka masih tinggi.
- Gerombol 2 merupakan provinsi dengan tingkat kesejahteraan menengah yang memerlukan upaya perbaikan dalam meningkatkan kualitas dan ketersediaan layanan dasar. Terdapat beberapa aspek positif seperti distribusi persentase PDRB dan tingkat penyelesaian pendidikan SMA yang tinggi. Selain itu, pengeluaran per kapita, umur harapan hidup, tingkat partisipasi angkatan kerja berada

- pada tingkat menengah relatif sama dengan rata-rata nasional. Persentase penduduk miskin, kepadatan penduduk, tingkat pengangguran terbuka, dan prevalensi stunting juga menunjukkan nilai rendah, menandakan indikasi kesejahteraan yang baik. Namun, proporsi rumah tangga dengan akses terhadap pelayanan dasar masih rendah.
- c. Gerombol 3 merupakan provinsi dengan tingkat kesejahteraan yang tinggi. Terlihat dari beberapa indikator yaitu kepadatan penduduk cukup tinggi menunjukkan tingginya aktivitas ekonomi dan kehidupan sosial di daerah tersebut. Meskipun distribusi PDRB relatif rendah, tetapi faktor-faktor lain seperti tingkat penyelesaian pendidikan SMA, pengeluaran per kapita, umur harapan hidup, persentase rumah tangga dengan akses terhadap hunian layak, tingkat partisipasi angkatan kerja, indeks pembangunan gender, serta proporsi rumah tangga dengan akses pelayanan dasar termasuk tinggi, hal tersebut mengindikasikan kesejahteraan yang baik. Selain itu, tingkat pengangguran terbuka, persentase penduduk miskin, dan prevalensi stunting yang rendah, menunjukkan adanya ketersediaan sumber daya dan pelayanan yang memadai.
 - d. Gerombol 4 merupakan provinsi dengan tingkat kesejahteraan menengah yang masih memerlukan perbaikan dalam infrastruktur dan pelayanan dasar. Provinsi di gerombol 4 memiliki pengeluaran per kapita dan indeks pembangunan gender yang relatif tinggi, tingkat penyelesaian pendidikan SMA relatif sama dengan rata-rata nasional, serta persentase penduduk miskin, tingkat pengangguran terbuka, dan prevalensi stunting termasuk rendah. Namun, disisi lain tingkat partisipasi angkatan kerja, umur harapan hidup, persentase rumah tangga dengan akses hunian layak, kepadatan penduduk, distribusi persentase PDRB, serta proporsi rumah tangga dengan akses pelayanan dasar masih rendah.
 - e. Gerombol 5 merupakan provinsi dengan tingkat kesejahteraan menengah yang memerlukan upaya untuk meningkatkan lapangan kerja dan peningkatan akses hunian layak guna meningkatkan kesejahteraan penduduknya. Hal tersebut disebabkan karena provinsi pada gerombol 5 memiliki tingkat pengangguran yang tinggi, partisipasi angkatan kerja dan distribusi PDRB serta akses terhadap hunian layak yang masih rendah. Walaupun, di sisi lain tingkat pendidikan SMA, pengeluaran per kapita, umur harapan hidup, indeks pembangunan gender, dan akses terhadap pelayanan dasar tinggi, serta prevalensi stunting dan persentase penduduk miskin rendah.
 - f. Gerombol 6 merupakan provinsi dengan tingkat kesejahteraan yang relatif tinggi dengan indikasi positif yaitu kepadatan penduduk, distribusi persentase PDRB, pengeluaran per kapita, dan umur harapan hidup tertinggi dibandingkan gerombol lainnya. Selain itu, tingkat penyelesaian pendidikan SMA dan proporsi rumah tangga dengan akses terhadap pelayanan dasar juga termasuk tinggi, indeks pembangunan gender dan tingkat partisipasi angkatan kerja relatif sama dengan rata-rata nasional, serta memiliki persentase penduduk miskin dan prevalensi stunting yang rendah. Meskipun demikian, terdapat tantangan yang perlu diatasi antara lain tingkat pengangguran terbuka yang tinggi dan persentase rumah tangga dengan akses terhadap hunian layak yang rendah. Upaya perbaikan diperlukan untuk terus meningkatkan kesejahteraan.
 - g. Gerombol 7 merupakan provinsi dengan tingkat kesejahteraan rendah yang memerlukan peningkatan pendidikan, aksesibilitas ekonomi, kesehatan masyarakat, dan pengurangan kesenjangan sosial. Tingkat penyelesaian pendidikan SMA, pengeluaran per kapita, umur harapan hidup, tingkat pengangguran terbuka, dan proporsi rumah tangga dengan akses terhadap pelayanan dasar memiliki nilai terendah dibandingkan dengan gerombol lainnya. Sedangkan tingkat partisipasi angkatan kerja dan persentase penduduk miskin pada provinsi di gerombol ini memiliki nilai tertinggi dibandingkan gerombol lainnya. Selain itu, gerombol ini memiliki kepadatan penduduk yang cukup tinggi, prevalensi stunting yang tinggi, serta distribusi persentase PDRB, persentase rumah tangga dengan akses terhadap hunian yang layak, dan indeks pembangunan

gender yang rendah.

- h. Gerombol 8 merupakan provinsi dengan pencapaian kesejahteraan menengah yang memerlukan upaya dalam pengelolaan populasi, peningkatan kesetaraan gender, dan implementasi program gizi untuk meningkatkan kesejahteraan. Meskipun umur harapan hidup, persentase rumah tangga dengan akses terhadap hunian yang layak, tingkat partisipasi angkatan kerja, dan proporsi rumah tangga dengan akses terhadap pelayanan dasar tinggi, serta tingkat pengangguran terbuka dan persentase penduduk miskin rendah. Namun, memiliki masalah kepadatan penduduk tinggi, prevalensi stunting tertinggi, serta indeks pembangunan gender terendah dibandingkan dengan gerombol lain.

Kesimpulan

Penelitian ini membandingkan metode penggerombolan K-Means, K-Medoids, dan DBSCAN pada penggerombolan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat menggunakan data hasil reduksi dimensi t-SNE. Hasil evaluasi penggerombolan berdasarkan koefisien Silhouette tertinggi dan indeks Davies-Bouldin terendah diperoleh metode penggerombolan terbaik adalah K-Means dan DBSCAN dengan parameter $perplexity = 1$, $minPts = 2$ dan $epsilon = 9$. Keduanya memperoleh hasil sama yaitu terbentuk delapan gerombol. Gerombol 1 dan gerombol 7 merupakan provinsi yang memiliki tingkat kesejahteraan terendah dibandingkan dengan gerombol lain, sehingga memerlukan upaya perbaikan yang lebih serius dan intensif. Provinsi pada gerombol tersebut diantaranya Provinsi Aceh, Maluku, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Papua Barat, dan Papua. Sedangkan provinsi pada gerombol lainnya hanya memerlukan upaya perbaikan dalam aspek tertentu.

Daftar Pustaka

- [1] BPS. (2022). *Jumlah penduduk pertengahan tahun (ribu jiwa) 2020-2022*. Available: <https://www.bps.go.id/indicator/12/1975/1/jumlah-penduduk-pertengahan-tahun.html>.
- [2] Budiati I., et al., *Indikator Kesejahteraan Rakyat*. Jakarta: BPS, 2022.
- [3] Hair J.F., et al., *Multivariate Data Analysis*, 7th ed., New Jersey: Pearson, 2010.
- [4] Hennig Christian, et al., *Handbook of cluster analysis*. New York: CRC Press, 2015.
- [5] Ramdhani F., Hoyyi A., and Mukid M.A., “Pengelompokan provinsi di Indonesia berdasarkan karakteristik kesejahteraan rakyat menggunakan metode K-Means cluster”, *Jurnal Gaussian*, vol.4, pp.875-884, Oct. 2015.
- [6] Dini S.K. and Fauzan A., “Clustering Provinces in Indonesia based on Community Welfare Indicators”, *EKSAKTA: Journal of Sciences and Data Analysis*, Yogyakarta, vol.1, pp.56-63, Feb. 2020. DOI: 10.20885/EKSAKTA.vol1.iss1.art9
- [7] Han J., Kamber M., and Pei J., *Data Mining Concepts and Techniques*, 3th ed., Waltham: Morgan Kaufmann, 2012.
- [8] Riyanto B. “Penerapan algoritma K-Medoids clustering untuk pengelompokan penyebaran diare di Kota Medan (Studi kasus: Kantor Dinas Kesehatan Kota Medan)”, *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol.3, pp.562–568, Oct. 2019. DOI: 10.30865/komik.v3i1.1659.
- [9] Shah G.H., Bhensdadia C.K., and Ganatra A.P., “An empirical evaluation of density based clustering techniques”. *International Journal of Soft Computing and Engineering (IJSCE)*. vol.2, pp.216-223, 2012.
- [10] BPS. (2022). Available: <https://www.bps.go.id/>.
- [11] Ditjen Bina Pembangunan Daerah Kementerian dalam Negeri. (2022). *Monitoring pelaksanaan 8 aksi Konvergensi Intervensi Penurunan Stunting Terintegrasi*. Available:

<https://aksi.bangda.kemendagri.go.id/emonev/DashPrev/index/4>.

- [12] Larose D.T. and Larose C.D., *Discovering Knowledge in Data An Introduction to Data Mining*. London: John Wiley & Sons, 2014.
- [13] Sehgal N.K. and Gupta P., *Introduction to Machine Learning in the Cloud with Python: Concepts and Practices*. Germany: Springer International Publishing, 2021.
- [14] Wang T., Yuan P., Ji C., Liu S., "Cooperative Anomaly Detection Model and Real-Time Update Strategy for Industrial Stream Data". Di dalam: Sun Y., Liu D., Liao H., Fan H., Gao L., editor. *Computer Supported Cooperative Work and Social Computing*. ChineseCSCW 2020; 2020 Nov 7-9; Shenzhen, China. Singapore: Springer. hlm 307-319. 2021.
- [15] Bali R., Sharma T., Sarkar D., *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*. United States: Apress, 2017.
- [16] Everitt B.S., et al., *Cluster Analysis*, 5th ed., London: Wiley, 2011.
- [17] Otterbach, Johannes. (2016, May.23). *Curse of dimensionality, t-SNE and Kullback-Leibler Divergence* [online]. Available: <https://jotterbach.github.io/content/posts/tsne/2016-05-23-TSNE/>.