

Pengelompokan Kabupaten/Kota di Sumatera Utara Menggunakan Algoritma *Average Linkage* dan *K-Means* Berdasarkan Indikator Pendidikan

Susan Patricia Simanjorang¹, Maulida Yanti^{2,*},

¹Program Studi Statistika, Universitas Sumatera Utara, Medan, Indonesia 20222

²Program Studi Matematika, Universitas Sumatera Utara, Medan, Indonesia 20222

*Corresponding author: maulidayanti@usu.ac.id



P-ISSN: 2986-4178

E-ISSN: 2988-4004

Riwayat Artikel

Dikirim: 04 September 2023

Direvisi: 10 Oktober 2023

Diterima: 7 Desember 2023

ABSTRAK

Pendidikan tidak bisa terlepas dari kehidupan manusia karena merupakan kunci/dasar pengimprovisasian kualitas dalam berbagai sektor/bidang. Gubernur Sumatera Utara menyatakan bahwa pemerataan pendidikan harus terus dioptimalkan. Pengelompokan (*clustering*) kabupaten berdasarkan ciri/ karakteristik yang sama akan membantu dalam penentuan daerah prioritas yang harus ditangani untuk pemerataan pendidikan oleh pemerintah Sumatera Utara. Penelitian ini bertujuan untuk menentukan pengelompokan kabupaten di Sumatera Utara berdasarkan indikator pendidikan menggunakan dua algoritma pengklasteran yang sering digunakan yaitu *K-Means* dan *Average Linkage*. Data pendidikan yang digunakan terdiri dari tiga variabel yaitu jumlah guru bersertifikasi, jumlah sekolah terakreditasi unggul dan jumlah siswa berprestasi tahun 2022. Nilai rasio simpangan baku untuk *K-Means* dan *Average Linkage* masing-masing adalah 0,261 dan 0,196. Berdasarkan nilai rasio simpangan baku, *Average Linkage* lebih baik dalam mengklasterkan data ini dibandingkan *K-Means*.

Kata Kunci: *Clustering, Average Linkage, K-Means, Pendidikan.*

ABSTRACT

Education cannot be separated from human life because it is the key/basis for improving quality in various sectors/fields. The Governor of North Sumatra stated that educational equality must continue to be optimized. Clustering districts based on the same characteristics will help in determining regional priorities that must be addressed for educational equality by the North Sumatra government. This research aims to determine the clustering of districts in North Sumatra based on education using two frequently used clustering algorithms, namely *K-Means* and *Average Linkage*. The educational data used consists of three variables, namely the number of certified teachers, the number of superior accredited schools and the number of outstanding students in 2022. Standard deviation ratio values for *K-Means* and *Average Linkage* are 0.261 and 0.196, respectively. Based

on this value, Average Linkage is better at clustering the data than K-Means.

Keywords: Clustering, Average Linkage; K-Means; Education.

1. Pendahuluan

Pendidikan merupakan aspek umum yang tidak bisa terpisah dari kehidupan manusia [1]. Pendidikan yang berkualitas tentu mempengaruhi kualitas dalam segala bidang. Tanpa dukungan kemajuan pendidikan, tidak mungkin tercapai percepatan pembangunan nasional di masa yang akan datang. Keberhasilan pendidikan dapat dilihat dari indikator pendidikan. Indikator yang mempengaruhi pemerataan pendidikan berkaitan dengan sarana dan prasarana pendidikan, khususnya jumlah sekolah dan tenaga kependidikan [2].

Salah satu rencana Gubernur Sumatera Utara adalah pemerataan pendidikan. Dalam hal ini, pengelompokan daerah-daerah dengan karakter pendidikan mirip sangat diperlukan. Mungkin saja daerah-daerah di kabupaten pinggiran memiliki karakter tingkat pendidikan serupa dan sebaliknya daerah-daerah di perkotaan juga memiliki karakter tingkat pendidikan tersendiri. Dengan mengelompokkan kabupaten/kota berdasarkan indikator pendidikan yaitu, variabel jumlah siswa berprestasi nasional, jumlah guru bersertifikasi dan jumlah sekolah terakreditasi unggul, akan terbentuk kelompok-kelompok daerah yang memiliki keidentikan antar daerah satu dengan yang lain serta akan terlihat pula kelompok daerah yang memiliki perbedaan dengan kelompok daerah lainnya, baik dari segi pendidikan yang terbilang baik maupun kurang. Hal ini tentu akan memberikan pengaruh pada keputusan pemerintah Sumatera Utara dalam memprioritaskan daerah mana yang harus dibenahi terlebih dahulu bidang pendidikannya.

Pengklasteran adalah proses mengelompokkan data ke dalam beberapa kelompok yang disebut klaster (*cluster*) [3], [4]. Terdapat banyak algoritma pengklasteran. Umumnya algoritma pengklasteran dibagi menjadi dua, yaitu pengklasteran secara hirarki dan pengklasteran non-hirarki. Salah satu contoh pengklasteran secara hirarki yang populer adalah *Average Linkage* [5]. Sedangkan contoh pengklasteran secara non-hierarki salah satunya yang sering digunakan adalah *K-Means*. Algoritma *K-Means* sering digunakan karena efisien dan mudah diimplementasikan walaupun perlu penentuan banyak klaster di awal [6]. Sebaliknya, *Average Linkage* cukup populer karena tidak perlu menentukan banyak klaster terlebih dahulu.

Kedua pengklasteran ini sudah banyak digunakan untuk pengelompokan berbagai kasus. *K-Means* pernah digunakan untuk mengelompokkan Kabupaten di Jawa Timur berdasarkan indikator pendidikan [2]. Rifa et al. (2020) [7] menggunakan *K-Means* untuk mengelompokkan resiko gempa di Indonesia. Disisi lain, Ningsih et al. (2016) [8] menggunakan *Average Linkage* untuk mengelompokkan kabupaten di Kalimantan Timur berdasarkan produksi palawija. Paramadina et al. (2019) [9] membandingkan *Average Linkage* dan *Ward* [10] untuk mengelompokkan IPM Provinsi Sulawesi Selatan dan mendapatkan *Average Linkage* memberikan hasil pengelompokan yang lebih baik.

Pada penelitian ini, penulis ingin membandingkan hasil pengelompokan kabupaten di Sumatera Utara berdasarkan indikator pendidikan dengan menggunakan 2 algoritma pengklasteran yaitu *Average Linkage* dan algoritma *K-Means*. Selanjutnya, akan dibandingkan algoritma mana yang menghasilkan klaster yang lebih baik/optimal berdasarkan nilai rasio simpangan baku. Penelitian ini diharapkan mampu menjadi saran kepada pihak pemerintah terkait daerah (kabupaten) mana yang harus diprioritaskan untuk pemerataan pendidikan di Sumatera Utara.

2. Metodologi Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari Kemdikbudristek. Data terdiri dari tiga variabel yaitu jumlah guru bersertifikasi (X_1), jumlah sekolah terakreditasi (X_2), dan jumlah siswa berprestasi (X_3), dari 33 kabupaten/kota di Sumatera Utara pada tahun 2022. Guru bersertifikasi adalah guru dengan kualifikasi minimal sarjana (S1). Sekolah terakreditasi adalah sekolah yang diakui oleh Badan Akreditasi Nasional Sekolah (BANS) dari segi kelayakan dan kinerja suatu Lembaga Pendidikan (sekolah) dengan akreditasi yang diperhitungkan dalam variabel ini hanyalah akreditasi unggul (A) dan baik (B). Siswa berprestasi adalah siswa dengan prestasi tingkat nasional atau internasional baik dalam bidang akademik maupun non akademik.

Teknik analisis data yang dilakukan dalam penelitian ini diuraikan sebagai berikut.

1. Mengambil data sekunder dari Kemdikbudristek.
2. Melakukan analisis deskriptif untuk melihat nilai minimum, maksimum, rata-rata dan standar deviasi dari setiap variabel.
3. Melakukan eksplorasi data untuk melihat persebaran data antar dua variabel menggunakan *scatter plot*.
4. Melakukan uji multikolinearitas menggunakan nilai VIF.
5. Transformasi dan standarisasi data.
6. Melakukan analisis *clustering* menggunakan *K-Means* dan *Average Linkage* menggunakan *software*.
7. Membandingkan hasil pengklasteran antara *K-Means* dan *Average Linkage* berdasarkan nilai rasio simpangan baku.
8. Melakukan interpretasi hasil pengklasteran.
9. Membuat kesimpulan dan saran.

2.1. K-Means

K-Means merupakan pengelompokan yang mempartisi data menjadi dua atau lebih kelompok dengan meminimalkan jumlah kuadrat jarak antara setiap objek kluster dengan pusat kluster terdekat [11]–[13]. Misalkan diberikan n objek dari data yaitu $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n \in \mathbb{R}^d$. Jarak antara objek \mathbf{x}_i dengan pusat kluster $\mathbf{c}_j \in \mathbb{R}^d$ dihitung menggunakan jarak Euclidean sebagai berikut,

$$d_{ij} = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \dots + (x_{id} - c_{jd})^2}, \quad (1)$$

dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, k$.

Pengelompokan dilakukan dengan tujuan untuk mengurangi atau meminimalkan keragaman dalam suatu kelompok dan memaksimalkan keanekaragaman antar kelompok [14]. Misalkan $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\}$ data yang ingin dikelompokkan dan $\mathbf{V} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ adalah pusat kluster, dengan n banyak objek dan k banyak kluster. Algoritma *K-Means* melakukan iterasi dengan tujuan meminimumkan total jarak antara objek dalam kluster dengan pusat klusternya. Secara matematis fungsi tujuan ini dapat ditulis sebagai berikut,

$$J(\mathbf{X}, \mathbf{V}) = \sum_{j=1}^k \sum_{\mathbf{x}_i \in C_j} \|\mathbf{x}_i - \mathbf{c}_j\|^2, \quad (2)$$

dengan

k : banyaknya kluster yang diberikan, \mathbf{x}_i : objek data ke- i , $i = 1, 2, \dots, n$, C_j : kluster ke- j
 \mathbf{c}_j : pusat kluster ke- j , $j = 1, 2, \dots, k$.

Secara lengkap, langkah-langkah pengklasteran dengan algoritma *K-Means* adalah sebagai berikut,

1. Tetapkan banyak kluster yang diinginkan, misalkan ada k kluster.
2. Pilih secara acak pusat awal kluster $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k$.

3. Hitung jarak setiap objek dengan pusat kluster menggunakan persamaan (1) untuk menentukan anggota tiap kluster dengan cara menempatkan objek pada kluster yang memiliki jarak objek dengan pusat kluster tersebut minimum.
4. Untuk setiap k kluster, hitung rata-rata dari anggota kluster dan tetapkan sebagai pusat baru c_1, c_2, \dots, c_k .
5. Ulangi langkah 3 dan 4 hingga pusat kluster konvergen.

2.2. Average Linkage

Average Linkage merupakan variasi dari algoritma *single linkage* dan *complete linkage* [8]. Jarak antar dua kluster dihitung sebagai jarak rata-rata antara semua pasangan objek data dalam kluster yang satu dengan objek pada kluster lain [8]. Jarak kluster P dan kluster Q dihitung menggunakan formula berikut [15], [16],

$$d_{PQ} = \frac{\sum_{p \in P} \sum_{q \in Q} d_{pq}}{N_P N_Q} \quad (3)$$

dengan,

d_{pq} : jarak antar objek p anggota kluster P dan objek q anggota kluster Q , N_P : banyak objek kluster P , N_Q : banyak objek kluster Q .

Misalkan p dan q di \mathbb{R}^d adalah sebarang dua objek data, jarak antara p dengan q dihitung menggunakan jarak Euclidean sebagai berikut,

$$d_{pq} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_d - q_d)^2} \quad (4)$$

Pada langkah awal dari algoritma *Average Linkage* setiap kluster beranggotakan masing-masing satu objek berbeda dari data. Secara lengkap algoritma pengklusteran dengan menggunakan *Average Linkage* adalah sebagai berikut:

1. Tentukan jarak antar kluster menggunakan Persamaan (3) untuk memperoleh matrik jarak *Euclidean* antar kluster.
2. Tentukan jarak antar kluster yang paling minimum.
3. Gabung dua kluster dengan jarak minimum menjadi 1 kluster.
4. Ulangi langkah 1 hingga 3 sampai terbentuk k kluster (k banyak kluster yang diinginkan).

2.3. Nilai VIF

Nilai VIF dapat digunakan untuk melihat apakah ada multikolinearitas antar variabel [17]. Nilai VIF > 10 menandakan adanya multikolinearitas [18]. Nilai VIF dapat dihitung menggunakan formula berikut [18],

$$VIF = \frac{1}{1 - r_{xy}^2} \quad (5)$$

dengan r_{xy} koefisien korelasi Pearson antara variabel x dan variabel y . Koefisien korelasi Pearson antara variabel x dan variabel y dapat dihitung menggunakan persamaan (6),

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

2.4. Rasio Simpangan Baku

Untuk melihat hasil pengelompokan terbaik dari dua algoritma yang digunakan, dapat dilakukan dengan cara membandingkan nilai rasio antara rata-rata simpangan baku dalam kluster (S_w) dan simpangan baku antar kluster (S_b) menggunakan Persamaan (10) [19]. Pengelompokan yang baik ditandai dengan nilai simpangan baku dalam kluster minimum dan nilai simpangan baku antar kluster maksimum [20]. Dengan demikian semakin kecil nilai rasio simpangan baku semakin baik hasil pengelompokan.

Misalkan k banyak kluster, kluster ke- i adalah C_i , $i = 1, 2, \dots, k$, dengan anggota kluster ini adalah x_j , $j = 1, 2, \dots, n_i$ dan rata-rata dari kluster ke- i adalah \bar{x}_i . Simpangan baku kluster ke- i dapat dihitung menggunakan Persamaan (7),

$$s_i = \sqrt{\frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_j - \bar{x}_i)^2} \quad (7)$$

Lebih lanjut, rata-rata simpangan baku dalam klaster (S_w) dapat dituliskan secara matematis seperti pada Persamaan (8),

$$S_w = \frac{1}{k} \sum_{i=1}^k s_i \quad (8)$$

Misalkan \bar{x} adalah rata-rata keseluruhan klaster. Simpangan baku antar klaster dapat dihitung menggunakan Persamaan (9),

$$S_b = \sqrt{\frac{1}{k - 1} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2} \quad (9)$$

Sehingga untuk mencari nilai rasio simpangan baku dapat menggunakan Persamaan (10) berikut,

$$S = \frac{S_w}{S_b} \quad (10)$$

dengan S_w simpangan baku dalam klaster dan S_b adalah simpangan baku antar klaster.

3. Hasil dan Pembahasan

3.1. Analisis Deskriptif

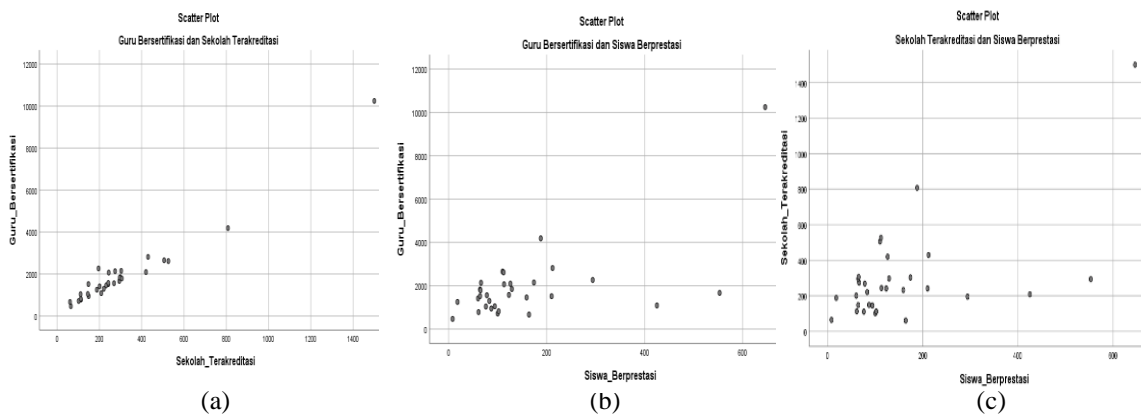
Tabel 3.1 merangkum statistika deskriptif dari data pendidikan Sumatera Utara tahun 2022 yang digunakan. Jumlah guru bersertifikasi di Sumatera Utara paling banyak terdapat di Kota Medan dengan jumlah 10.245 orang dan paling rendah berada di Kabupaten Pakpak Bharat dengan jumlah 448 orang. Rata-rata dan standar deviasi jumlah guru bersertifikat adalah masing-masing 2.191,21 dan 2.026,88. Jumlah sekolah yang telah terakreditasi unggul (A) dan baik (B) di Sumatera Utara juga terdapat paling banyak di Kota Medan dengan jumlah 1.500 bangunan dan paling rendah berada di Kota Sibolga yaitu sebanyak 61 bangunan. Siswa berprestasi tingkat nasional atau internasional, dalam bidang akademik maupun non akademik berasal paling banyak dari kota Medan dengan angka 646 orang dan paling rendah yaitu 8 orang berasal dari Kabupaten Pakpak Bharat. Kota Medan memiliki nilai paling tinggi di ke-tiga variabel.

Tabel 3.1 Statistika Deskriptif

Variabel	Minimum	Maksimum	Mean	Std. Deviation
Jumlah guru bersertifikat (X_1)	468	10.245	2.191,21	2.026,88
Jumlah sekolah terakreditasi (X_2)	61	1.500	339,36	318,04
Jumlah siswa berprestasi (X_3)	8	646	157,45	142,56

3.2. Eksplorasi Data

Scatter plot antara dua variabel yang digunakan, masing-masing dapat dilihat pada Gambar 3.1 berikut.



Gambar 3.1 Scatter Plot antar Variabel

Gambar 3.1(a) memperlihatkan *scatter plot* antara jumlah sekolah terakreditasi dan jumlah guru bersertifikat. Terlihat bahwa titik cenderung mengumpul di tiga kluster, dimana satu titik (paling atas) terpisah dari dua kluster lainnya. *Scatter plot* antara jumlah siswa berprestasi dan jumlah guru bersertifikat terdapat pada Gambar 3.1(b) sedangkan Gambar 3.1(c) merupakan *scatter plot* antara jumlah siswa berprestasi dan jumlah sekolah terakreditasi. Kedua *scatter plot* ini memperlihatkan pola yang cenderung sama, yaitu satu titik terpisah sendiri dan titik lainnya tersebar di lokasi yang berdekatan.

3.3. Uji Multikolinieritas

Tabel 3.2 dan Tabel 3.3 masing-masing menunjukkan nilai korelasi dan nilai VIF antar variabel. Karena terdapat nilai VIF > 10 antara variabel X_1 dan X_2 , maka terdapat multikolinieritas. Penulis melakukan transformasi logaritma natural untuk variabel X_1 sebagai upaya untuk menghilangkan multikolinieritas dan memisalkan X_1T menyatakan transformasi logaritma natural dari variabel ini.

Tabel 3.2 Nilai Korelasi Variabel

	X_1	X_2	X_3
X_1	1		
X_2	0.982	1	
X_3	0.595	0.568	1

Tabel 3.3 Nilai VIF Variabel

Variabel	Korelasi Pearson (r)	r^2	Toleransi ($1 - r^2$)	VIF
X_1 dan X_2	0.982	0.964	0.036	28.030
X_1 dan X_3	0.595	0.354	0.646	1.548
X_2 dan X_3	0.568	0.323	0.677	1.476

Selanjutnya diperiksa kembali multikolinieritas dari variabel X_1T, X_2 dan X_3 . Tabel 3.4 dan Tabel 3.5 masing-masing menunjukkan nilai korelasi dan Nilai VIF antar variabel. Terlihat bahwa nilai VIF antar variabel sudah kurang dari (<) 10 yang artinya tidak terjadi multikolinieritas antar variabel.

Tabel 3.4 Nilai Korelasi Variabel Baru

	X_1T	X_2	X_3
X_1T	1		
X_2	0.908	1	
X_3	0.497	0.568	1

Tabel 3.5 Nilai VIF Baru

Variabel	Korelasi Pearson (r)	r^2	Toleransi ($1 - r^2$)	VIF
X_1T dan X_2	0.908	0.826	0.174	5.756
X_1T dan X_3	0.497	0.247	0.753	1.328
X_2 dan X_3	0.568	0.323	0.677	1.476

3.4. Pengelompokan dengan K-Means dan Average Linkage

Sebelum dilakukan pengelompokan dengan *Average Linkage* ataupun dengan *K-Means*, data terlebih dahulu distandarisasi. Standarisasi dilakukan terhadap variabel X_1T, X_2 dan X_3 dan misalkan kembali hasil standarisasi dari masing-masing variabel ini adalah ZX_1T, ZX_2 dan ZX_3 . Standarisasi dilakukan menggunakan bentuk z skor dengan formula sebagai berikut:

$$z = \frac{x_i - \bar{x}}{s}, \quad (11)$$

dengan x_i nilai variabel ke-i, s adalah standar deviasi, yaitu $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$, dari variabel yang ingin distandarisasi. Masing-masing hasil pengelompokan menggunakan *K-Means* dan *Average Linkage* diperoleh seperti pada Tabel 3.6 dan Tabel 3.7.

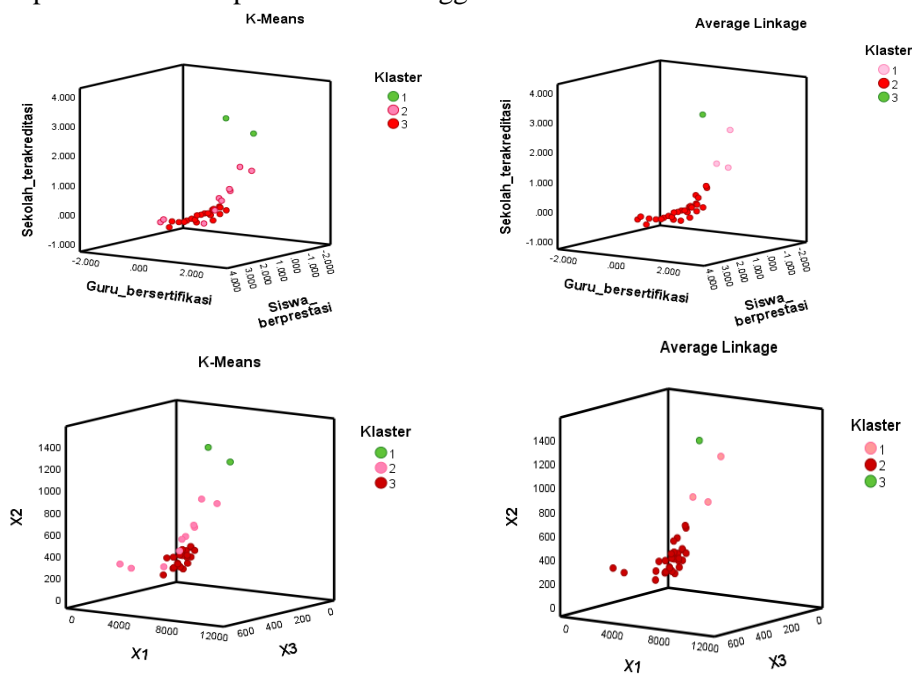
Tabel 3.6 Hasil Pengklasteran *K-Means*

Klaster	Anggota
Klaster 1	Kota Medan, Kab. Deli Serdang
Klaster 2	Kab. Langkat, Kab. Simalungun, Kab. Asahan, Kab. Labuhanbatu, Kab. Tapanuli Utara, Kab. Tapanuli Tengah, Kab. Samosir, Kab. Mandailing Natal, Kab. Serdang Bedagai, Kota Pematangsiantar
Klaster 3	Kab. Karo, Kab. Dairi, Kab. Tapanuli Selatan, Kab. Tobasa, Kab. Nias Selatan, Kab. Humbang Hasundutan, Kab. Batu Bara, Kab. Padang Lawas Utara, Kab. Padang Lawas, Kab. Labuhanbatu Utara, Kab. Labuhanbatu Selatan, Kab. Nias Utara, Kota Binjai, Kota Tebing Tinggi, Kota Padang Sidempuan, Kota Gunungsitoli, Kab. Nias, Kab. Pakpak Bharat, Kab. Nias Barat, Kota Tanjungbalai, Kota Sibolga.

Tabel 3.7 Hasil Pengklasteran *Average Linkage*

Klaster	Anggota
Klaster 1	Kab. Deli Serdang, Kab. Langkat, Kab. Simalungun.
Klaster 2	Kab. Karo, Kab. Dairi, Kab. Asahan, Kab. Labuhanbatu, Kab. Tapanuli Utara, Kab. Tapanuli Selatan, Kab. Mandailing Natal, Kab. Tobasa, Kab. Nias Selatan, Kab. Humbang Hasundutan, Kab. Serdang bedagai, Kab. Batu Bara, Kab. Tapanuli Tengah, Kab. Samosir, Kab. Padang Lawas, Kab. Labuhanbatu Utara, Kab. Labuhanbatu Selatan, Kota Binjai, Kota Pematangsiantar, Kota Padang Sidempuan, Kab. Nias, Kab. Pakpak Bharat, Kab. Padang Lawas Utara, Kab. Nias Barat, Kab. Nias Utara, Kota Tanjungbalai, Kota Sibolga, Kota Tebing Tinggi, Kota Gunungsitoli.
Klaster 3	Kota Medan

Gambar 3.2 memperlihatkan *scatter plot* hasil pengelompokan menggunakan *K-Means* (kiri) dan *Average Linkage* (kanan) dengan sumbu masing-masing menggunakan nilai variabel setelah tranformasi dan standarisasi (atas) dan nilai asli (bawah). Secara visual, terlihat *Average Linkage* menghasilkan pengelompokan yang lebih tepat dibandingkan dengan *K-Means*. Anggota klaster 1, 2 dan 3 pada *Average Linkage* terpisah sempurna sementara pada *K-Means* anggota klaster 1 dan 2 masih terlihat berdekatan.



Gambar 3.2. Hasil Pengelompokan Menggunakan *K-Means* dan *Average Linkage*

3.5. Penentuan Hasil Pengklasteran Terbaik

Nilai rasio antara rata-rata simpangan baku dalam klaster dan simpangan baku antar klaster dapat digunakan untuk melihat mana algoritma yang menghasilkan pengklasteran terbaik [19]. Nilai ini disebut nilai rasio simpangan baku. Semakin kecil nilai

rasio ini maka semakin baik algoritma tersebut dalam mengelompokkan karena memiliki nilai homogenitas yang tinggi di dalam kluster [19]. Nilai rasio simpangan baku dihitung menggunakan Persamaan (10). Dari perhitungan diperoleh nilai rata-rata simpangan baku untuk *K-Means* dan *Average Linkage* masing-masing adalah 0,261 dan 0,196. Berdasarkan nilai rasio simpangan baku diperoleh bahwa *Average Linkage* memiliki kinerja lebih baik dalam mengemlopokkan data pendidikan Sumatera Utara tahun 2022 dibandingkan *K-Means*.

3.6. Interpretasi Hasil Pengelompokan Terbaik

Dari perhitungan diperoleh nilai rata-rata simpangan baku untuk *K-Means* dan *Average Linkage* masing-masing adalah 0,261 dan 0,196. Dari nilai ini metode *Average Linkage* lebih akurat dalam mengklusterkan dibandingkan metode *K-means*. Lebih lanjut, interpretasi hasil pengelompokan yang digunakan adalah hasil pengelompokan *Average Linkage*, yaitu sebagai berikut:

1. Kluster 1 beranggotakan 3 kabupaten, yaitu Deli Serdang, Langkat dan Simalungun. Dari kluster ini diperoleh, rata-rata guru bersertifikasi sebanyak 5.654 orang, sekolah terakreditasi unggul 932 unit dan 207 siswa berprestasi di tahun 2022. Sehingga, kluster ini dapat dikatakan sebagai kluster kabupaten dengan kualitas pendidikan yang cukup baik di Sumatera Utara.
2. Kluster 2 beranggotakan 29 kabupaten (dapat dilihat pada Tabel 3.7). Rata-rata guru berkualifikasi sebanyak 1.555 orang, sekolah terakreditasi unggul sebanyak 238 unit dan 136 siswa berprestasi di kluster ini. Dengan demikian dapat dikatakan kabupaten di kluster ini adalah kabupaten yang memiliki tingkat kualitas pendidikan secara keseluruhan menengah ke bawah.
3. Kluster 3 hanya terdiri dari 1 kota yaitu kota Medan dengan banyak guru bersertifikasi sebanyak 10.245, sekolah terakreditasi unggul sebanyak 1.500 unit dan 646 siswa berprestasi di tahun 2022. Dengan demikian, kluster ini adalah kluster yang memiliki kualitas pendidikan paling baik di Sumatera Utara.

4. Kesimpulan

Berdasarkan pembahasan di atas, kesimpulan yang dapat diambil adalah sebagai berikut:

1. Hasil kluster yang terbentuk dengan menggunakan *Average Linkage* dan *K-Means* adalah sebagai berikut:
 - a. Untuk *Average Linkage* diperoleh hasil kluster pertama terdiri dari 3 kabupaten/kota, kluster kedua terdiri dari 29 kabupaten/kota, kluster ketiga yaitu 1 kabupaten/kota. Dengan urutan kluster 3, 1, 2 berdasarkan tingkat kualitas pendidikan dari yang terbaik yang diukur melalui banyaknya guru bersertifikasi, sekolah terakreditasi dan siswa berprestasi.
 - b. Sedangkan dengan menggunakan *K-Means*, menghasilkan kluster pertama terdiri dari 2 kabupaten/kota, kluster kedua terdiri dari 10 kabupaten/kota, kluster 3 terdiri dari 21 kabupaten/kota.
2. Nilai rasio simpangan baku *Average Linkage* yaitu 0,196 lebih kecil dari nilai rasio simpangan baku *K-Means* yaitu 0,261. Oleh karena itu, *Average Linkage* lebih baik dalam mengelompokkan kabupaten/kota di Sumatera Utara berdasarkan indikator pendidikan.
3. Berdasarkan interpretasi hasil pengklusteran metode *Average Linkage*, kluster 2 beranggotakan 29 kabupaten merupakan kluster dengan pendidikan rendah dibandingkan kluster yang lain. Sehingga untuk pemerataan pendidikan, pemerintah Sumatera Utara dapat memprioritaskan daerah ini (Kab. Karo, Kab. Dairi, Kab. Asahan, Kab. Labuhanbatu, Kab. Tapanuli Utara, Kab. Tapanuli Selatan, Kab. Mandailing Natal, Kab. Tobasa, Kab. Nias Selatan, Kab. Humbang Hasundutan, Kab. Serdang bedagai, Kab. Batu Bara, Kab. Tapanuli Tengah, Kab. Samosir, Kab. Padang Lawas, Kab. Labuhanbatu Utara, Kab. Labuhanbatu Selatan, Kota Binjai, Kota Pematangsiantar,

Kota Padang Sidempuan, Kab. Nias, Kab. Pakpak Bharat, Kab. Padang Lawas Utara, Kab. Nias Barat, Kab. Nias Utara, Kota Tanjungbalai, Kota Sibolga, Kota Tebing Tinggi, Kota Gunungsitoli) terlebih dahulu.

5. Daftar Pustaka

- [1] R. Ariana, "Model Pembelajaran Kooperatif Tipe Make a Match Terhadap Hasil Belajar Siswa," pp. 1–23, 2016.
- [2] H. S. Karti, "Pengelompokan Kabupaten / Kota di Provinsi SMA / SMK / MA dengan Metode C-Means dan Fuzzy C-Means," vol. 2, no. 2, 2013.
- [3] P. Rai and S. Singh, "A Survey of Clustering Techniques," *Int. J. Comput. Appl.*, vol. 7, no. 12, 2010, doi: 10.5120/1326-1808.
- [4] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. 2005. doi: 10.1002/0471687545.
- [5] H. K. Seifoddini, "Single linkage versus average linkage clustering in machine cells formation applications," *Comput. Ind. Eng.*, vol. 16, no. 3, 1989, doi: 10.1016/0360-8352(89)90160-5.
- [6] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*, vol. 9, no. 8. 2020. doi: 10.3390/electronics9081295.
- [7] I. H. Rifa, H. Pratiwi, and R. Respatiwan, "Clustering of Earthquake Risk in Indonesia Using K-Medoids and K-Means Algorithms," *Media Stat.*, vol. 13, no. 2, 2020, doi: 10.14710/medstat.13.2.194-205.
- [8] S. Ningsih, S. Wahyuningsih, and Y. N. Nasution, "Perbandingan Kinerja Metode Complete Linkage dan Average Linkage dalam Menentukan Hasil Analisis Cluster," *Pros. Semin. Sains dan Teknol. FMIPA Unmul*, vol. 1, no. 1, 2016.
- [9] M. Paramadina, S. Sudarmin, and M. K. Aidid, "Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan)," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 1, no. 2, p. 22, Jul. 2019, doi: 10.35580/variansiunm9357.
- [10] F. Murtagh and P. Legendre, "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?," *J. Classif.*, vol. 31, no. 3, 2014, doi: 10.1007/s00357-014-9161-z.
- [11] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
- [12] E. Forgy, "Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications," *Biometrics*, vol. 21, no. 3, 1965.
- [13] P. Fránti and S. Sieranoja, "K-means properties on six clustering benchmark datasets," *Appl. Intell.*, vol. 48, no. 12, 2018, doi: 10.1007/s10489-018-1238-7.
- [14] Z. Cebeci and F. Yildiz, "Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures," *J. Agric. Informatics*, vol. 6, no. 3, Oct. 2015, doi: 10.17700/jai.2015.6.3.196.
- [15] R. R. Sokal, "A statistical method for evaluating systematic relationships," *Univ Kans Sci Bull*, vol. 38, 1958.
- [16] L. Ramos Emmendorfer and A. M. de Paula Canuto, "A generalized average linkage criterion for Hierarchical Agglomerative Clustering," in *Applied Soft Computing*, 2021. doi: 10.1016/j.asoc.2020.106990.
- [17] R. K. Paul, "Multicollinearity: causes, effects and remedies," *IASRI, New Delhi*, vol. 1, no. 1, pp. 58–65, 2006.
- [18] A. Alin, "Multicollinearity," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 3, 2010, doi: 10.1002/wics.84.
- [19] S. B. Purnamasari, H. Yasin, and T. Wuryandari, "Pemilihan Cluster Optimum Pada Fuzzy C-means (Studi Kasus: Pengelompokan Kabupaten/kota Di Provinsi Jawa Tengah Berdasarkan Indikator Indeks)," *J. Gaussian*, vol. 3, no. 3, pp. 491–498, 2014.
- [20] A. R. Barakbah and K. Arai, "Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering," *Ind. Electron. Semin. 2004*, no. October, pp. 409–413, 2004.