

Analisa Mutu Sekolah Pada Provinsi Kalimantan Timur Menggunakan Algoritma K-Means

Mega Yoalifa^{*1}, Haviluddin^{c:2}, Masna Wati³, Novianti Puspitasari⁴, Ummul Hairah⁵

Program Studi Informatika Fakultas Teknik, Universitas Mulawarman, Samarinda 75119, Indonesia
¹yoalifa9@gmail.com; ²haviluddin@unmul.ac.id; ³masnawati@fkti.ac.id; ⁴novia.ftik.unmul@gmail.com;
⁵unmul.hairah@fkti@unmul.ac.id

INFORMASI ARTIKEL

Histori Artikel

Diterima : 21 September 2020

Direvisi : 12 Juli 2021

Diterbitkan : 14 Agustus 2021

Kata Kunci:

Mutu Sekolah

Clustering

K-Means

Sum Square error (SSE)

Euclidean Distance

Manhattan Distance

Minkowski Distance

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan Sekolah Menengah Atas (SMA) berdasarkan Standar Mutu Pendidikan (SNP) sehingga memiliki kategori Standar Tinggi (C1), Standar Sedang (C2), dan Standar Rendah (C3) di Daerah Kutai Barat dan Kutai Kartanegara, Provinsi Kalimantan Timur. Metode analisa telah menggunakan algoritma K-Means dengan tiga metode perhitungan jarak yaitu Euclidean Distance, Manhattan Distance, dan Minkowski Distance. Berdasarkan hasil percobaan dengan Euclidean Distance dan Minkowski Distance terdapat 9 sekolah dengan perhitungan akurasi sum of square error (SSE) sebesar 42.6793 dalam kategori berstandar tinggi (C1), 48 sekolah dengan akurasi perhitungan SSE sebesar 26.6885 berkategori standar sedang (C2), dan 3 sekolah dengan akurasi perhitungan SSE sebesar 52.6727 berkategori standar rendah (C3). Hasil penelitian ini diharapkan menjadi rekomendasi dalam memberikan program kerja peningkatan mutu dan kualitas SMA oleh pihak-pihak terkait seperti Dinas Pendidikan dan Kebudayaan (Dikbud).

2021 SAKTI – Sains, Aplikasi, Komputasi dan Teknologi Informasi.

Hak Cipta.

I. Pendahuluan

Pada awalnya, Lembaga Penjamin Mutu Pendidikan (LPMP) merupakan Balai Penataran Guru (BPG) yang didirikan secara *de jure* pada tahun 1991 melalui SK Mendikbud Tahun 1991 (*SK Mendikbud Nomor 0240a/O/1991 tanggal 2 Mei 1991, n.d.*), dengan fungsi sebagai pelaksana Pendidikan di bawah lingkup Direktorat Jenderal Pendidikan dasar dan Menengah Departemen Pendidikan Nasional. LPMP Provinsi Kalimantan Timur berfungsi untuk melaksanakan penjaminan mutu pendidikan dasar dan menengah di provinsi Kalimantan Timur. Salah satu problema pendidikan di Indonesia yaitu mutu pendidikan. LPMP melakukan berbagai upaya peningkatan mutu pendidikan dengan pelaksanaan proyek seperti Bimbingan Teknis Pembuatan Konten Rumah Belajar, Rapat koordinasi dan Sosialisasi Penjaminan Mutu Pendidikan, hingga Penyelenggaraan Diskusi Kelompok Terumpun (DKT) seluruh Kabupaten Kota di Provinsi Kalimantan Timur. Kemendikbud (2016) menyatakan mutu pendidikan dasar dan menengah adalah tingkat kesesuaian antara penyelenggaraan pendidikan dasar dan menengah dengan Standar Nasional Pendidikan. Mutu diartikan dengan pemenuhan kepuasan pelanggan, sebuah sekolah diartikan bermutu apabila mampu memberikan layanan sesuai kebutuhan dan kepuasan pelanggan sesuai dengan standar yang berlaku. Kondisi pada standar mutu pendidikan pada suatu pendidikan tidak akan meningkat tanpa diiringi dengan penjaminan mutu baik internal maupun eksternal. Peningkatan mutu pendidikan tidak akan banyak berarti jika tidak dikawal oleh Sistem Penjaminan Mutu Internal (SPMI). Sistem penjaminan mutu internal digunakan untuk memantau dan mengevaluasi tingkat ketercapaian Standar Nasional Pendidikan dan acuan pelaksanaan akreditasi satuan pendidikan (MY, 2019).

Dalam rangka membantu menganalisa mutu dan kualitas yang dimiliki oleh setiap sekolah, dilakukan pengolahan data arsip rapor penilaian mutu pendidikan yang berfokus pada daerah Kutai Barat dan Kutai Kartanegara. Agar pihak LPMP dapat memantau pencapaian dan mengevaluasi perkembangan yang terjadi pada sekolah hingga dapat mengoptimalkan bantuan dan pembuatan program kerja yang tepat untuk peningkatan mutu sekolah berdasarkan 8 kriteria yang telah ditetapkan oleh pemerintah. Dari permasalahan tersebut, diperlukan adanya pengelompokan sekolah kedalam beberapa kategori dengan mengimplementasikan data mining yaitu *clustering* data menggunakan algoritma K-Means. Algoritma K-Means merupakan algoritma pengelompokan *iterative* yang sederhana diimplementasikan, relatif cepat, dan mudah beradaptasi (Ghosh & Dubey, 2012). Data di kelompokkan kedalam kategori Standar Tinggi dengan anggota Sekolah yang telah mencapai nilai standar nasional, Standar Sedang dengan anggota sekolah yang

menuju standar nasional, dan Standar Rendah dengan anggota sekolah yang belum mencapai nilai standar nasional.

Berbagai penelitian yang menggunakan algoritma K-Means telah dilakukan oleh peneliti. Peneliti (Nasari & Darma, 2015), Nasari, F. dkk telah menggunakan algoritma K-Means untuk *clustering* pada data penerimaan mahasiswa baru, dengan jumlah 279 data. Hasil penelitian yaitu mahasiswa yang asal sekolahnya SMA rata-rata mengambil jurusan Sistem Informasi, sedangkan mahasiswa yang asal sekolahnya SMK rata-rata mengambil jurusan Teknik Informatika. Peneliti (Priyatman et al., 2019), Priyatman, H. dkk telah menggunakan algoritma K-Means untuk memprediksi waktu kelulusan mahasiswa menggunakan data akademik sebanyak 100 data. Implementasi algoritma K-Means pada data telah berhasil, namun untuk melihat tingkat kemampuan real-k-means clustering dalam memprediksi waktu kelulusan tergantung mahasiswa itu sendiri. Peneliti (Talaku et al., 2017), M.W. Talakuma, dkk telah menggunakan algoritma K-Means untuk mengelompokkan kota berdasarkan indikator Indeks Pembangunan Manusia (IPM). Data IPM diperoleh dari Badan Pusat Statistik (BPS) dengan jumlah data sebanyak 11 kota. Berdasarkan hasil penelitian, dari 11 kota hanya terdapat 1 kota saja yang memiliki nilai IPM tertinggi, sedangkan kelompok ketiga yaitu nilai IPM terendah memiliki anggota yang paling banyak.

Peneliti (Taslim & Fajrizal, 2016), Taslim dkk telah menggunakan algoritma K-Means untuk *clustering* data obat pada Puskesmas Rumbai. Data yang digunakan berasal dari LPLPO Puskesmas Rumbai pada tahun 2014 sebanyak 133 data. Disimpulkan bahwa banyak diantara obat yang masuk dalam kelompok cluster kurang dikarenakan tidak adanya permintaan selama beberapa bulan terhadap obat tersebut. Peneliti (Nurzahputra et al., 2017), Aldi Nurzahputra, dan kawan-kawan telah menggunakan algoritma K-Means untuk *clustering* penilaian dosen berdasarkan Indeks Kepuasan Mahasiswa. Data diperoleh dari kusioner terhadap mahasiswa dengan jumlah 146 responden. Diperoleh hasil penilaian terhadap 12 dosen dengan dibagi menjadi dua kategori yaitu 5 dosen cluster_baik dan 7 dosen cluster_kurang. Peneliti (Sibuea & Safta, 2017), Mustika Larasati S. dan Andy telah menggunakan algoritma K-Means untuk *clustering* siswa berprestasi. Data penelitian diperoleh dari rapor peserta didik dengan jumlah 10 peserta dan dikelompokkan menjadi 3 *cluster*. Keakuratan yang didapat cukup baik terhadap permasalahan yang terjadi terkait prestasi siswa dengan hasil 70%. Artikel ini menggunakan algoritma K-Means untuk menganalisa mutu pendidikan SMA berdasarkan data mutu pendidikan tahun 2018. Tujuan penelitian ini adalah untuk membantu LPMP, Dinas Pendidikan dan Kebudayaan Provinsi, Kabupaten/Kota dalam mendata dan mengolah mutu pendidikan sesuai Kategori Standar Nasional Pendidikan (SNP) yang ditetapkan oleh Kemendikbud sehingga layanan pendidikan kepada siswa dapat terlayani dengan baik, tepat, dan bersesuaian dengan kategori yang dibutuhkan oleh SMA. Artikel ini terdiri dari motivasi penulisan artikel pada bagian pertama. Kedua, menerangkan model kerja algoritma K-Means. Ketiga, analisa hasil percobaan. Kesimpulan penelitian pada bagian akhir.

II. Material dan Metode

A. Algoritma K-Means

Pengelompokkan data atau *clustering* merupakan proses mempartisi atau membagi data dalam suatu himpunan yang memiliki kemiripan karakteristik satu sama lain. Salah satu algoritma mempartisi data adalah K-Means yang bersifat *iterative*. Secara prinsip, metode K-Means memiliki K sebagai jumlah *cluster* yang diinginkan. Sedangkan, Means yaitu rata-rata dari suatu kelompok data yang didefinisikan sebagai *cluster* (Nasari & Darma, 2015). Lebih lanjut, metode mempartisi data dengan melihat jarak terpendek data dengan titik pusat *cluster* (*centroid*) didapatkan dengan rumus perhitungan jarak yaitu *Euclidean Distance*, *Manhattan Distance*, *Minkowski Distance* (Puspitasari & Haviluddin, 2016). Berikut *flowchart* metode K-Means pada Gambar 1 (Nur et al., 2017).

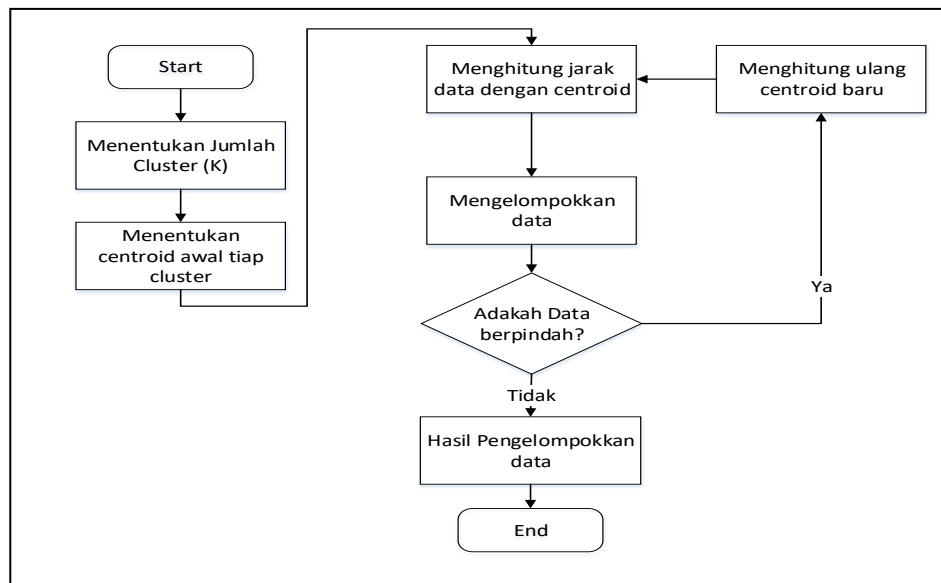
Dalam penelitian ini, alur algoritma K-Means yang digunakan terdiri dari lima tahapan. Pertama, menentukan jumlah cluster (k). Dalam penelitian ini, sebanyak 3 *cluster* yaitu Standar Tinggi (C1), Standar Sedang (C2), dan Standar Rendah (C3) telah ditetapkan. Kedua, menentukan titik pusat (*centroid*) awal dari masing-masing cluster berdasarkan nilai maksimal, minimal, dan nilai tengah pada *cluster*. Ketiga, menghitung jarak antara data dengan *centroid*. Dalam penelitian ini, rumus perhitungan jarak telah menggunakan *Euclidean Distance*, *Manhattan Distance*, dan *Minkowski Distance*. Rumus metode *Euclidean Distance* ditunjukkan pada Persamaan 1, *Manhattan Distance* pada Persamaan 2, dan *Minkowski Distance* pada Persamaan 3 (Nishom, 2019).

$$\text{Euclidean Distance} \quad d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$\text{Manhattan Distance} \quad d(x,y) = \left(\sum_{i=1}^n |x_i - y_i| \right) \quad (2)$$

$$\text{Minkowski Distance} \quad d(x,y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (3)$$

Dimana, d adalah jarak antara x dan y; x adalah data pusat cluster; y adalah data pada atribut; i adalah setiap data; n adalah jumlah data; x_i adalah data pada pusat cluster ke-i; y_i adalah data pada setiap data ke-i; p adalah power.



Gambar. 1. Flowchart Algoritma K-Means

Tahap keempat, mengelompokkan data dengan hasil perhitungan jarak setiap data ke *centroid* dengan mencari nilai terpendek atau terkecil. Kelima, melakukan iterasi lalu mencari titik *centroid* baru dengan persamaan 4. Ulangi langkah 3 sampai 5 hingga tidak ada lagi data yang berpindah *cluster*. Rumus mencari *centroid* baru ditunjukkan pada Persamaan 4.

$$C(i) = \frac{x_1 + x_2 + x_{...} + x_{...}}{\sum x} \quad (4)$$

Dengan x_1 adalah nilai data ke-1, x_2 adalah nilai data ke-2, dan $\sum x$ adalah jumlah data.

B. Akurasi Algoritma K-Means

Dalam penelitian ini, pengukuran akurasi jarak metode Euclidean, Manhattan, dan Minkowski telah menggunakan perhitungan nilai error pada jarak data yaitu Sum Square Error (SSE). Secara prinsip, metode ini akan memberikan informasi error jarak data ke centroid. Semakin kecil nilai SSE maka semakin baik hasil clustering yang didapat. Rumus SSE ditampikan pada Persamaan 5 (James, 2019).

$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5)$$

Dengan x_i adalah nilai data tiap cluster, dan \bar{x} adalah nilai rata-rata data cluster.

C. Dataset

Dalam penelitian ini, berfokus pada data rapor mutu SMA di daerah Kutai Barat dan Kutai Kartanegara tahun 2018 yang telah diperoleh dari LPMP Provinsi Kalimantan Timur dengan jumlah 60 data. Kemudian, terdapat 8 kriteria terdiri dari Standar Kompetensi Kelulusan (SKK), Standar Isi (SI), Standar Proses (SP), Standar Penilaian Pendidikan (SPP), Standar Pendidik dan Tenaga Kerja (SPTK), Standar Sarana dan Prasarana Pendidikan (SSPP), Standar Pengelolaan Pendidikan (SPLP) dan, Standar Pembiayaan (SPN) dimana telah ditetapkan oleh Kemendikbud yang menjadi dasar penilaian sekolah-sekolah oleh pihak LPMP. Adapun, data mutu SMA ditampilkan pada Tabel 1.

Tabel 1. Model Tabel Yang Diijinkan Jurnal SAKTI

No	Nama Sekolah	Variabel	Indikator		
			Indikator1	Indikator2	Indikator3
1	Advent Sendawar	S1	6	0	6.37
2	N 1 Bentian Besar	S2	6.91	3.79	6.66
3	N 1 Bongan	S3	6.92	4.6	6.46
4	N 1 Damai	S4	6.97	3.79	6.66
5	N 1 Jempang	S5	6.44	4.67	6.28
6	N 1 Linggang Bingung	S6	6.84	4.32	6.58
7	N 1 Long Iram	S7	6.95	6.3	6.65
8	N 1 Mook Manaar Bulatn	S8	6.97	3.2	6.24
9	N 1 Muara Lawa	S9	7	7	7
10	N 1 Muara Pahu	S10	6.85	3.79	6.35
...
60	YPM Diponegoro Tgr	S60	6.92	3.2	6.49

Pra-pengolahan data dengan proses normalisasi telah dilakukan bertujuan untuk mengubah nilai-nilai dalam range [0..1]. Dalam penelitian ini, normalisasi data dengan metode *Z-Score* telah digunakan (Murtioso, 2019), rumus *z-score* ditampilkan pada Persamaan 6.

$$Z_i = \frac{X_i - \bar{X}}{S} \quad (6)$$

Dimana, Z_i adalah data hasil normalisasi; x_i adalah data asli; \bar{X} adalah nilai rata-rata; S adalah simpangan baku.

III. Hasil dan Pembahasan

Pada sub bagian ini, menguraikan proses kerja algoritma K-Means yang diterapkan untuk menganalisa mutu dan kualitas SMA berdasarkan Kategori Standar Nasional Pendidikan (SNP). Dalam percobaan ini, pengelompokan data mutu dan kualitas SMA Kutai Barat dan Kutai Kartanegara telah dilakukan dengan menggunakan algoritma K-Means. Tahap pertama adalah melakukan proses normalisasi terhadap 60 data. Hasil Normalisasi ditampilkan pada Tabel 2.

Tabel 2. Hasil Normalisasi Data

No	Variabel	Indikator		
		Indikator1	Indikator2	Indikator3
1	S1	0.7409	-3.8476	0.3279
2	S2	0.6819	-1.3633	0.5180
3	S3	0.6884	-0.8323	0.3869
4	S4	0.7212	-1.3633	0.5180
5	S5	0.3738	-0.7864	0.2689
6	S6	0.6360	-1.0159	0.4655
7	S7	0.7081	0.2820	0.5114
8	S8	0.7212	-1.7500	0.2427
9	S9	0.7409	0.7409	0.7409
10	S10	0.6425	-1.3633	0.3148
...
60	S60	0.6884	-1.7500	0.4066

A. Perhitungan K-Means Metode Euclidean Distance

Menentukan titik pusat pada data dengan nilai *max*, *min*, dan nilai tengah, dari data hasil normalisasi. Titik pusat data (centroid) ditampilkan pada Tabel 3.

Tabel 3. Hasil Normalisasi Data

Cluster	Indikator 1	Indikator 2	Indikator 3
C1	0.7409	0.7409	0.7409
C2	0.7212	-1.0159	0.5114
C3	0.3738	-3.8476	0.0395

Setelah mendapatkan nilai centroid, kemudian jarak data dengan centroid dihitung. Dalam percobaan ini, data variabel S1 - S60 dihitung dengan Persamaan (1). Hasil perhitungan jarak *Euclidean Distance* pada iterasi pertama ditampilkan pada Tabel 4.

Tabel 4. Iterasi 1 Euclidean Distance

Variabel Sekolah	Jarak Data Ke Centroid			Jarak data terdekat
	C1	C2	C3	
S1	4.6070	2.8377	0.4668	C3
S2	2.1167	0.3497	2.5487	C2
S3	1.6134	0.2242	3.0515	C2
S4	2.1160	0.3475	2.5537	C2
S5	1.6402	0.4818	3.0697	C2
S6	1.7812	0.0968	2.8756	C2
S7	0.5140	1.2979	4.1699	C1
S8	2.5403	0.7818	2.1358	C2
S9	0.0000	1.7717	4.6562	C1
S10	2.1491	0.4069	2.5139	C2
...
S60	2.5137	0.7423	2.1526	C2
Jumlah Anggota	7	50	3	

Setelah seluruh data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali titik pusat cluster yang baru berdasarkan rata-rata anggota pada cluster tersebut. Selanjutnya, lakukan perhitungan kembali jarak data ke centroid. Langkah perhitungan ulang centroid baru terus dilakukan hingga tidak ada lagi data yang berpindah cluster. Pada metode *Euclidean Distance*, iterasi berhenti pada iterasi ke-3. Penghentian iterasi dilakukan karena kelompok data sama dengan iterasi sebelumnya. Hasil iterasi terakhir ditampilkan pada Tabel 5.

Tabel 5. Iterasi Akhir Euclidean Distance

Variabel Sekolah	Jarak Data Ke Centroid			Jarak data terdekat
	C1	C2	C3	
S1	4.2419	2.5640	0.0831	C3
S2	1.7506	0.1045	2.4873	C2
S3	1.2405	0.4585	3.0158	C2
S4	1.7503	0.1107	2.4867	C2
S5	1.2714	0.6153	3.0861	C2
S6	1.4113	0.2760	2.8341	C2
S7	0.1584	1.5705	4.1309	C1
S8	2.1687	0.5074	2.1044	C2
S9	0.3745	2.0498	4.6003	C1
S10	1.7767	0.1581	2.4880	C2
...
S60	2.1454	0.4648	2.0982	C2
Jumlah Anggota	9	48	3	

B. Perhitungan K-Means Metode Manhattan Distance

Menentukan titik pusat pada data dengan nilai max, min, dan nilai tengah, dari data hasil normalisasi. Titik pusat data ditampilkan pada Tabel 2. Setelah mendapatkan nilai centroid, kemudian jarak data dengan centroid dihitung. Dalam percobaan ini, data variabel S1 – S60 dihitung dengan Persamaan (2). Hasil perhitungan jarak *Manhattan Distance* pada iterasi pertama ditampilkan pada Tabel 6.

Seluruh data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali titik pusat cluster yang baru berdasarkan rata-rata anggota pada cluster tersebut. Selanjutnya, lakukan perhitungan kembali jarak data ke centroid. Langkah perhitungan ulang centroid baru terus dilakukan hingga tidak ada lagi data yang berpindah cluster. Pada metode *Manhattan Distance*, iterasi berhenti pada iterasi ke-4. Penghentian iterasi dilakukan karena kelompok data tidak berubah. Hasil iterasi terakhir ditampilkan pada Tabel 7.

Tabel 6. Iterasi 1 Manhattan Distance

Variabel Sekolah	Jarak Data Ke Centroid			Jarak data terdekat
	C1	C2	C3	
S1	5.0014	2.9956	0.6555	C3
S2	2.3860	0.3802	3.2709	C2
S3	1.9796	0.0262	3.6773	C2
S4	2.3467	0.3409	3.3102	C2
S5	2.3663	0.3605	3.2906	C2
S6	2.1369	0.1311	3.5200	C2
S7	0.7210	1.2848	4.9359	C1
S8	3.0087	1.0029	2.6482	C2
S9	0.0000	2.0058	5.6569	C1
S10	2.6285	0.6227	3.0284	C2
...
S60	2.8776	0.8718	2.7793	C2
Jumlah Anggota	6	51	3	

Tabel 7. Iterasi Akhir Manhattan Distance

Variabel Sekolah	Jarak Data Ke Centroid			Jarak data terdekat
	C1	C2	C3	
S1	4.4239	2.6010	0.0808	C3
S2	1.8085	0.0144	2.5346	C2
S3	1.4021	0.4208	2.9410	C2
S4	1.7692	0.0537	2.5739	C2
S5	1.7888	0.0340	2.5542	C2
S6	1.5594	0.2635	2.7837	C2
S7	0.1436	1.6793	4.1995	C1
S8	2.4312	0.6084	1.9118	C2
S9	0.5775	2.4004	4.9206	C1
S10	2.0510	0.2282	2.2920	C2
...
S60	2.3001	0.4773	2.0429	C2
Jumlah Anggota	10	47	3	

C. Perhitungan K-Means Metode Minkowski Distance

Menentukan titik pusat pada data dengan nilai max, min, dan nilai tengah, dari data hasil normalisasi. Titik pusat data ditampilkan pada Tabel 2. Setelah mendapatkan nilai centroid, kemudian jarak data dengan centroid dihitung. Dalam percobaan ini, data variabel S1 – S60 dihitung dengan Persamaan (3). Hasil perhitungan jarak *Minkowski Distance* pada iterasi pertama ditampilkan pada Tabel 8.

Tabel 8. Iterasi 1 Minkowski Distance

Variabel Sekolah	Jarak Data Ke Centroid			Jarak data terdekat
	C1	C2	C3	
S1	4.6070	2.8377	0.4668	C3
S2	2.1167	0.3497	2.5487	C2
S3	1.6134	0.2242	3.0515	C2
S4	2.1160	0.3475	2.5537	C2
S5	1.6402	0.4818	3.0697	C2
S6	1.7812	0.0968	2.8756	C2
S7	0.5140	1.2979	4.1699	C1
S8	2.5403	0.7818	2.1358	C2
S9	0.0000	1.7717	4.6562	C1
S10	2.1491	0.4069	2.5139	C2
...
S60	2.5137	0.7423	2.1526	C2
Jumlah Anggota	7	50	3	

Setelah seluruh data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali titik pusat cluster yang baru berdasarkan rata-rata anggota pada cluster tersebut. Selanjutnya, lakukan perhitungan kembali jarak data ke centroid. Langkah perhitungan ulang centroid baru terus dilakukan hingga tidak ada lagi data yang berpindah cluster. Pada metode *Minkowski Distanc*, iterasi berhenti pada iterasi ke-3. Penghentian iterasi dilakukan karna kelompok data sama dengan iterasi sebelumnya. Hasil iterasi terakhir ditampilkan pada Tabel 9.

Tabel 9. Iterasi akhir Minkowski Distance

Variabel Sekolah	Jarak Data Ke Centroid			Jarak data terdekat
	C1	C2	C3	
S1	4.2419	2.5640	0.0831	C3
S2	1.7506	0.1045	2.4873	C2
S3	1.2405	0.4585	3.0158	C2
S4	1.7503	0.1107	2.4867	C2
S5	1.2714	0.6153	3.0861	C2
S6	1.4113	0.2760	2.8341	C2
S7	0.1584	1.5705	4.1309	C1
S8	2.1687	0.5074	2.1044	C2
S9	0.3745	2.0498	4.6003	C1
S10	1.7767	0.1581	2.4880	C2
...
S60	2.1454	0.4648	2.0982	C2
Jumlah Anggota	9	48	3	

D. Perhitungan Akurasi Algoritma K-Means

Tahap pengujian nilai error menggunakan metode Sum Square Error (SSE) yang memberikan informasi error jarak data ke centroid (Mohamad, 2016). Dalam penelitian, SSE digunakan untuk menguji hasil clustering yang dibuat sebanyak tiga cluster, empat cluster, dan lima cluster berdasarkan tiap metode penentuan jarak. Dengan menggunakan tabel, akan dibandingkan hasil nilai error yang didapatkan pada tiap metode perhitungan jarak dan akan dicari nilai error terkecil dari ketiga metode. Hasil nilai SSE ditampilkan pada Tabel 10.

Tabel 10. Iterasi akhir Minkowski Distance

Distance	Cluster	SSE
Euclidean	K=3	122.0584
	K=4	159.0304
	K=5	189.6692
Manhattan	K=3	139.6123
	K=4	159.0304
	K=5	189.6692
Minkowski	K=3	122.0584
	K=4	159.0304
	K=5	189.6692

Tabel 10 menampilkan hasil dari perhitungan Nilai SSE dari ketiga jarak. Dari ketiga jarak tersebut, jarak yang memiliki nilai terkecil dimiliki oleh Euclidean Distance dan Minkowski Distance dengan nilai yang sama yaitu 122.0584. Manhattan Distance memiliki nilai error terbesar dengan nilai 139.6123.

IV. Kesimpulan

Penerapan algoritma K-Means dalam menganalisa standar mutu Sekolah Menengah Atas (SMA) di Daerah Kutai Barat dan Kutai Kartanegara, Provinsi Kalimantan Timur telah dilakukan. Berdasarkan hasil percobaan telah diperoleh model tiga kelompok standar sekolah dengan Standar Tinggi (C1), Standar Sedang (C2), dan Standar Rendah (C3) yang mengacu pada Kriteria 1 (K1) hingga Kriteria 8 (K8). Percobaan ini juga telah menerapkan tiga metode perhitungan jarak yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Minkowski Distance* dimana akurasi perhitungan jarak telah digunakan metode *Sum Square Error* (SSE). Berdasarkan hasil pengelompokkan *Euclidean Distance* dan *Minkowski Distance*, telah diperoleh sekolah dengan Standar Tinggi (C1) sebanyak 9 sekolah dengan akurasi perhitungan SSE sebesar 42.6973, kemudian sekolah dengan Standar Sedang (C2) sebanyak 48 sekolah dengan akurasi perhitungan SSE sebesar 26.6885, dan sekolah dengan Standar Rendah (C3) sebanyak 3 sekolah dengan akurasi perhitungan SSE sebesar 52.6727. Dengan metode *Manhattan Distance*, kelompok Standar Tinggi (C1) sebanyak 10 sekolah dengan akurasi perhitungan SSE sebesar 44.8970, kemudian sekolah dengan Standar Sedang (C2) sebanyak 47 sekolah dengan akurasi perhitungan SSE sebesar 33.5125, dan sekolah dengan Standar Rendah (C3) sebanyak 3 sekolah dengan akurasi perhitungan SSE sebesar 61.2027. Hasil perbandingan nilai *error* pada ketiga metode perhitungan jarak memperlihatkan nilai *error* terkecil dimiliki oleh *Euclidean Distance* dan *Minkowski Distance*. Hasil penelitian menunjukkan bahwa K-Means *clustering* dengan tiga *cluster* merupakan jumlah ideal dibandingkan hasil clustering berjumlah empat dan lima *cluster*. Komparasi dan optimalisasi algoritma K-Means untuk menghasilkan akurasi pengelompokkan yang lebih akurat menjadi penelitian selanjutnya.

Ucapan Terima Kasih

Terima kasih kepada pihak Lembaga Penjamin Mutu Pendidikan Kalimantan Timur, khususnya kepada Bapak Drs. Ali Sadikin, M.AP selaku kepala Seksi Pemetaan Mutu dan Supervisi yang memberikan ijin pengambilan dan pengolahan data mutu sekolah sebagai data penelitian. Bapak Syahrul selaku pembimbing

lapangan yang memberikan masukan dan bantuan selama proses penelitian. Kedua pembimbing, Bapak Haviluddin, S.Kom., M.Kom., Ph.D dan Ibu Masna Wati, S.Si., M.T yang telah membimbing penelitian ini sehingga dapat terlaksana dengan baik serta kedua penguji Ibu Ir. Novianti Puspitasari, S.Kom., M.Eng dan Ibu Ummul Hairah, S.Pd., M.T. yang memberikan masukan yang membangun selama penelitian.

Daftar Pustaka

- Ghosh, S., & Dubey, S. K. (2012). A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms. *International Journal of Computational Engineering Research*, 2(3), 737–739.
- James, W. (2019). *How to Calculate the Sum of Squares for Error (SSE)*. PROBABILITY AND STATISTICS. [https://www.wikihow.com/Calculate-the-Sum-of-Squares-for-Error-\(SSE\)](https://www.wikihow.com/Calculate-the-Sum-of-Squares-for-Error-(SSE))
- Mohamad. (2016). *SSE - Sum Square Error*. NUMXL Support. <https://support.numxl.com/hc/en-us/articles/215653803-SSE-Sum-of-Squared-Errors>
- Murtioso, O. A. (2019). *Clustering Pemakaian Obat Menggunakan Algoritma K-Means Pada Puskesmas Lempake*.
- MY, N. (2019). Meraih Mutu Sekolah Melalui Data Mutu: Suatu Tinjauan. *Jurnal Ilmiah Universitas Batanghari Jambi*, 19(2), 347. <https://doi.org/10.33087/jiubj.v19i2.677>
- Nasari, F., & Darma, S. (2015). Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama). *Seminar Nasional Teknologi Informasi Dan Multimedia 2015*, 6–8.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>
- Nur, F., Zarlis, M., & Nasution, B. B. (2017). Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 1(2), 100–105. <https://doi.org/10.30743/infotekjar.v1i2.70>
- Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno.Com*, 16(1), 17–24. <https://doi.org/10.33633/tc.v16i1.1284>
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1), 62. <https://doi.org/10.26418/jp.v5i1.29611>
- Puspitasari, N., & Haviluddin. (2016). Penerapan metode k-means dalam pengelompokkan curah hujan di kalimantan timur. *Proceedings Snrik 2016, Volume 1*(December 2016), 2–7.
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteks*, 4(1), 85–92. <https://doi.org/10.33330/jurteks.v4i1.28>
- SK Mendikbud Nomor 0240a/O/1991 tanggal 2 Mei 1991*. (n.d.).
- Talakua, M. W., Leleury, Z. A., & Talluta, A. W. (2017). *Analisis Cluster Dengan Menggunakan Metode Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014 Cluster Analysis By Using K-Means Method for Grouping of District / City in Maluku Province Industrial Based on Indicators of Maluku Dev. 11*, 119–128.
- Taslim, T., & Fajrizal, F. (2016). Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 7(2), 108–114. <https://doi.org/10.31849/digitalzone.v7i2.602>