

Analisis Kepatuhan Belanja Dana BOS Melalui Klasterisasi K-Means++: Studi Kasus Pada Satuan Pendidikan Tingkat Menengah Pertama

Jayyid Jiddan, Ilka Zufria*

Fakultas Sains dan Teknologi, Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: 'jayyidjiddan87@gmail.com, *ilkazufria@uinsu.ac.id

Email Penulis Korespondensi: ilkazufria@uinsu.ac.id

Submitted 18-12-2025; Accepted 09-02-2026; Published 24-02-2026

Abstrak

Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 18 Tahun 2022 telah menetapkan kewajiban bagi seluruh satuan pendidikan untuk melaksanakan pengadaan barang dan jasa menggunakan Dana Bantuan Operasional Sekolah (BOS) melalui platform SIPLah. Kebijakan ini merupakan langkah strategis pemerintah untuk meningkatkan transparansi, akuntabilitas, dan efisiensi dalam pengelolaan dana pendidikan nasional. Namun, implementasi kebijakan ini di lapangan masih menghadapi tantangan besar yang ditandai dengan tingkat adopsi platform yang masih rendah, di mana mayoritas sekolah cenderung memprioritaskan belanja manual konvensional yang rentan terhadap risiko penyimpangan administratif. Penelitian ini bertujuan untuk menganalisis tingkat kepatuhan belanja Dana BOS melalui klasterisasi pola belanja guna mengidentifikasi karakteristik adopsi SIPLah dan merumuskan rekomendasi kebijakan yang tepat sasaran. Metode yang digunakan adalah algoritma K-Means++ yang dipilih karena keunggulan dalam inisialisasi centroid yang lebih optimal dibandingkan K-Means standar, sehingga menghasilkan klaster yang lebih akurat dengan konvergensi yang cepat. Data penelitian bersumber dari Rencana Kegiatan dan Anggaran Sekolah (RKAS) Tahun Anggaran 2024 pada 40 sampel Sekolah Menengah Pertama Negeri (SMPN) yang dipilih menggunakan rumus Slovin dengan tingkat kesalahan 10%. Proses klasterisasi menghasilkan tiga kelompok karakteristik kepatuhan yang distingsif, yaitu: Klaster 1 (Kepatuhan Rendah) yang mendominasi dengan 30 sekolah (75%) pada rentang penggunaan SIPLah 0-5%, Klaster 2 (Kepatuhan Sedang) sebanyak 8 sekolah (20%) pada rentang >5-15%, dan Klaster 3 (Kepatuhan Tinggi) hanya sebanyak 2 sekolah (5%) pada rentang >15%. Validasi hasil menggunakan Silhouette Coefficient menghasilkan nilai 0,687 yang mengindikasikan struktur klaster terbentuk dengan baik (good structure). Temuan ini memberikan kontribusi penting bagi dinas pendidikan sebagai dasar evaluasi kebijakan, di mana sekolah pada Klaster 1 memerlukan audit infrastruktur dan pendampingan intensif, sedangkan Klaster 3 dapat dijadikan model percontohan (best practice) dalam tata kelola keuangan sekolah.

Kata Kunci: K-Means++; Kepatuhan Belanja; Dana Bos; SIPLah; Klasterisasi; Transparansi

Abstract

The Ministry of Education, Culture, Research, and Technology Regulation Number 18 of 2022 mandates all educational units to procure goods and services using School Operational Assistance (BOS) funds through the SIPLah platform. This policy is a strategic government step to enhance transparency, accountability, and efficiency in the management of national education funds. However, field implementation still faces significant challenges characterized by low platform adoption rates, where the majority of schools tend to prioritize conventional manual procurement prone to administrative irregularity risks. This study aims to analyze the compliance level of BOS fund spending through spending pattern clustering to identify SIPLah adoption characteristics and formulate targeted policy recommendations. The method used is the K-Means++ algorithm, selected for its superiority in optimal centroid initialization compared to standard K-Means, resulting in more accurate clusters with fast convergence. Research data sourced from the 2024 School Activity Plan and Budget (RKAS) of 40 State Junior High School (SMPN) samples selected using Slovin's formula with a 10% error rate. The clustering process produced three distinct compliance characteristic groups: Cluster 1 (Low Compliance) dominating with 30 schools (75%) in the SIPLah usage range of 0-5%, Cluster 2 (Medium Compliance) comprising 8 schools (20%) in the range of >5-15%, and Cluster 3 (High Compliance) comprising only 2 schools (5%) in the range of >15%. Validation using the Silhouette Coefficient yielded a score of 0.687, indicating a well-formed cluster structure. These findings provide a significant contribution to the education office as a basis for policy evaluation, where schools in Cluster 1 require infrastructure audits and intensive assistance, while Cluster 3 can serve as a best practice model in school financial governance.

Keywords: K-Means++; Spending Compliance; BOS Funds; SIPLah; Clustering; Transparency

1. PENDAHULUAN

Dana Bantuan Operasional Sekolah (BOS) merupakan program pendanaan pendidikan yang bersumber dari Anggaran Pendapatan dan Belanja Negara (APBN) untuk mendukung biaya operasional non-personalia di sekolah dasar dan menengah. Program ini bertujuan meringankan beban masyarakat dalam rangka wajib belajar yang berkualitas, mencakup pembiayaan operasional seperti bahan habis pakai, pemeliharaan sarana prasarana, dan honorarium tenaga pendidik [1], [2]. Seiring perkembangan teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi menerbitkan Permendikbudristek Nomor 18 Tahun 2022 yang mewajibkan penggunaan Sistem Informasi Pengadaan Barang dan Jasa di Sekolah (SIPLah) [3], [4]. Kebijakan ini dimaksudkan untuk meningkatkan transparansi, akuntabilitas, dan efisiensi belanja sekolah, serta meminimalisir potensi penyimpangan anggaran yang kerap terjadi pada metode pengadaan konvensional [5].

Namun, fakta di lapangan menunjukkan adanya kesenjangan yang signifikan antara regulasi dan implementasi. Observasi awal pada data realisasi anggaran menunjukkan bahwa tingkat adopsi platform digital ini masih sangat rendah. Sebagian besar satuan pendidikan cenderung masih memprioritaskan belanja manual di luar sistem resmi. Kondisi ini menciptakan risiko ketidaktransparanan dan menyulitkan proses audit. Permasalahan mendasar yang dihadapi saat ini

adalah ketiadaan pemetaan objektif mengenai profil kepatuhan belanja sekolah, sehingga pemerintah daerah kesulitan menentukan strategi intervensi yang tepat sasaran.

Beberapa penelitian terdahulu telah menerapkan metode klasterisasi untuk pengelompokan data di berbagai sektor. Nugroho & Adhinata (2022) menerapkan algoritma K-Means++ untuk klasterisasi penyebaran COVID-19 dan membuktikan bahwa metode ini memiliki akurasi inisialisasi centroid yang lebih baik dibandingkan K-Means standar [6]. Penelitian lain oleh Syahkur et al. (2024) menggunakan metode serupa untuk evaluasi gizi balita [7]. Sementara itu, studi terkait SIPLah oleh Herinanto (2024) lebih berfokus pada aspek pemasaran produk UMKM [11], namun belum menyentuh aspek evaluasi kepatuhan belanja.

Berdasarkan tinjauan literatur tersebut, secara eksplisit teridentifikasi celah penelitian (Research Gap) yang belum tergal. Mayoritas penelitian clustering di bidang pendidikan cenderung berfokus pada performa akademik siswa atau data kesehatan, dan belum ada yang secara spesifik menganalisis pola belanja Dana BOS berbasis kepatuhan digital (e-procurement) menggunakan algoritma optimasi. Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan mengimplementasikan algoritma K-Means++ guna mengelompokkan sekolah berdasarkan tingkat kepatuhan penggunaan SIPLah. Algoritma K-Means++ dipilih secara spesifik karena keunggulannya dalam strategi inisialisasi probabilistik yang mampu mengatasi kelemahan local optima pada K-Means biasa [8], [9].

Kontribusi utama dari penelitian ini adalah memberikan bukti empiris berupa pemetaan kluster sekolah berdasarkan tingkat kepatuhan belanjanya yang selama ini belum terpetakan. Hasil ini diharapkan dapat menjadi landasan bagi Dinas Pendidikan dalam merumuskan rekomendasi kebijakan yang berbasis data (evidence-based policy), seperti penentuan prioritas audit bagi sekolah di kluster rendah dan pemberian penghargaan bagi sekolah di kluster tinggi, demi terwujudnya tata kelola keuangan sekolah yang transparan dan akuntabel.

2. METODOLOGI PENELITIAN

2.1 Kerangka Penelitian

Gambar 1 menunjukkan kerangka penelitian secara keseluruhan yang menggambarkan alur sistematis dari setiap tahapan yang dilakukan, perhatikan gambar 1 dibawah ini :



Gambar 1. Kerangka Penelitian

Penelitian ini menggunakan pendekatan sistematis yang terdiri dari lima tahap utama: perencanaan, analisis data, desain sistem, penerapan, dan evaluasi. Pada tahap perencanaan, dilakukan identifikasi masalah terkait rendahnya kepatuhan sekolah dalam penggunaan SIPLah untuk pembelanjaan dana BOS yang berpotensi menimbulkan ketidaktransparanan dan risiko penyimpangan dana. Pengumpulan data dilakukan menggunakan data sekunder berupa RKAS (Rencana Kegiatan dan Anggaran Sekolah) beserta realisasinya dari Dinas Pendidikan Kabupaten Deli Serdang tahun 2024. Data ini mencakup informasi detail tentang alokasi dan realisasi dana BOS untuk tiga kategori belanja: Barang dan Jasa, Peralatan dan Mesin, serta Aset Tetap. Selain data RKAS, penelitian juga mengumpulkan data persentase penggunaan SIPLah dari Buku Kas Umum (BKU) setiap sekolah untuk mengukur tingkat adopsi platform digital dalam pengadaan barang dan jasa. Pada tahap analisis data, populasi penelitian adalah 64 SMPN di Kabupaten Deli Serdang yang memiliki data RKAS lengkap tahun 2024.

2.2 Algoritma K-Means++

Algoritma K-Means++ adalah penyempurnaan dari algoritma K-Means standar dengan fokus pada inisialisasi centroid yang lebih cerdas. Perbedaan mendasar terletak pada tahap pemilihan centroid awal: K-Means memilih k centroid secara acak sepenuhnya, sedangkan K-Means++ menggunakan metode probabilistik sampling berdasarkan jarak kuadrat (D^2 -

sampling) untuk memastikan centroid tersebar secara optimal di ruang data. Langkah-langkah algoritma K-Means++ dalam penelitian ini adalah sebagai berikut:

Tahap 1: Inisialisasi Centroid K-Means++. Pilih centroid pertama (c_1) secara acak dari dataset. Untuk setiap centroid berikutnya ($c_2, c_3, \dots, c_k$), hitung jarak kuadrat $D(x_i)^2$ dari setiap data point x_i ke centroid terdekat yang sudah terpilih. Kemudian hitung probabilitas pemilihan setiap data point sebagai centroid berikutnya menggunakan formula:

$$P(x_i) = D(x_i)^2 / \sum D(x_i)^2 \quad (1)$$

Di mana $\sum D(x_i)^2$ adalah jumlah total jarak kuadrat semua data point. Data point dengan jarak lebih jauh dari centroid yang sudah ada memiliki probabilitas lebih tinggi untuk terpilih sebagai centroid berikutnya. Proses ini diulang hingga k centroid awal terpilih. Dalam penelitian ini, $k=3$ sesuai dengan tiga kategori yang diinginkan: rendah, sedang, dan tinggi.

Tahap 2: Assignment Step. Setelah k centroid awal terpilih, lakukan iterasi standar K-Means. Pada setiap iterasi, assign setiap data point ke centroid terdekat berdasarkan jarak Euclidean. Untuk data satu dimensi (persentase SIPLah), jarak Euclidean disederhanakan menjadi nilai absolut selisih:

$$d(x_i, c_j) = |x_i - c_j| \quad (2)$$

Setiap sekolah ditugaskan ke kluster j jika $d(x_i, c_j)$ adalah jarak minimum di antara semua centroid.

Tahap 3: Update Step. Setelah semua data point ter-assign ke kluster masing-masing, hitung centroid baru sebagai rata-rata nilai semua anggota kluster:

$$c_j(\text{baru}) = (1/n_j) \sum x_i \text{ untuk semua } x_i \in \text{kluster } j \quad (3)$$

Di mana n_j adalah jumlah anggota kluster j .

Tahap 4: Konvergensi. Ulangi Assignment Step dan Update Step hingga centroid tidak berubah posisi (konvergensi tercapai) atau hingga mencapai jumlah iterasi maksimum. Kriteria konvergensi ditetapkan jika perubahan posisi centroid kurang dari threshold $\epsilon = 0,001$.

2.3 Evaluasi Kualitas Kluster dengan Silhouette Coefficient

Silhouette Coefficient adalah metrik evaluasi klusterisasi yang mengukur seberapa baik setiap objek ditempatkan dalam klusternya dan seberapa terpisah kluster-kluster tersebut satu sama lain. Nilai Silhouette berkisar dari -1 hingga +1, di mana nilai mendekati +1 menunjukkan objek sangat cocok dengan klusternya dan jauh dari kluster tetangga, nilai mendekati 0 menunjukkan objek berada di perbatasan antara dua kluster, dan nilai negatif menunjukkan objek mungkin ditempatkan di kluster yang salah.

Perhitungan Silhouette Coefficient untuk setiap data point i terdiri dari tiga langkah. Pertama, hitung $a(i)$ sebagai rata-rata jarak dari data point i ke semua data point lain dalam kluster yang sama:

$$a(i) = (1/|A|-1) \sum d(i,j) \text{ untuk semua } j \in A, j \neq i \quad (4)$$

Di mana A adalah kluster yang berisi data point i , dan $d(i,j)$ adalah jarak antara data point i dan j .

Kedua, hitung $b(i)$ sebagai minimum dari rata-rata jarak data point i ke semua data point di setiap kluster lain:

$$d(i,C) = (1/|C|) \sum d(i,j) \text{ untuk semua } j \in C \quad (5)$$

$$b(i) = \min\{d(i,C)\} \text{ untuk semua kluster } C \neq A \quad (6)$$

Ketiga, hitung Silhouette Coefficient $s(i)$ untuk data point i :

$$s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\} \quad (7)$$

Silhouette Coefficient keseluruhan adalah rata-rata $s(i)$ untuk semua data point. Interpretasi nilai: 0,71-1,00 menunjukkan struktur kuat, 0,51-0,70 struktur baik, 0,26-0,50 struktur lemah, dan $\leq 0,25$ tidak terstruktur.

2.4 Penentuan Parameter Klusterisasi

Salah satu tantangan terbesar dalam clustering adalah menentukan jumlah kluster (k) yang paling optimal. Dalam penelitian ini, nilai $k=3$ ditetapkan berdasarkan pendekatan domain knowledge dan kebutuhan praktis kebijakan. Tujuannya adalah memetakan sekolah ke dalam tiga strata intervensi manajerial yang jelas: (1) Kepatuhan Rendah (Low Compliance), (2) Kepatuhan Sedang (Medium Compliance), dan (3) Kepatuhan Tinggi (High Compliance). Pembagian ini memudahkan Dinas Pendidikan dalam merumuskan strategi berjenjang, mulai dari audit investigatif hingga pemberian penghargaan [7]. Parameter eksekusi algoritma diatur dengan batas maksimum iterasi sebanyak 100 kali untuk memastikan algoritma memiliki waktu yang cukup mencapai titik konvergensi.

2.5 Penentuan Sampel

Populasi penelitian mencakup 64 Sekolah Menengah Pertama Negeri (SMPN). Mengingat keterbatasan sumber daya dan efisiensi analisis, ukuran sampel (n) ditentukan menggunakan formula Slovin dengan batas toleransi kesalahan (e) sebesar 10% (0,10):

$$n = N / (1 + N \times e^2) = 64 / (1 + 64 \times 0,10^2) = 64 / 1,64 \approx 39,02 \approx 40 \text{ sekolah}$$

Teknik penarikan sampel dilakukan secara purposive sampling. Prioritas sampel diberikan kepada sekolah yang memiliki variansi data transaksi (nilai persentase >0) agar model klasterisasi dapat mempelajari pola belanja yang nyata dan representatif, bukan sekadar mengelompokkan data nol.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data dan Statistik Deskriptif

Dataset penelitian bersumber dari data Rencana Kegiatan dan Anggaran Sekolah (RKAS) beserta realisasinya dari 40 sampel Sekolah Menengah Pertama Negeri di Kabupaten Deli Serdang tahun 2024. Total realisasi dana BOS mencapai Rp43.488.968.032 yang terdiri dari belanja Barang dan Jasa sebesar Rp36.327.197.053, Peralatan dan Mesin Rp2.159.704.068, dan Aset Tetap Rp5.002.066.911. Dari total belanja tersebut, hanya Rp949.769.024 atau 2,18% yang dilakukan melalui platform SIPLah, menunjukkan tingkat adopsi yang sangat rendah meskipun penggunaan platform telah diwajibkan melalui Permendikbudristek Nomor 18 Tahun 2022.

Analisis statistik deskriptif terhadap persentase penggunaan SIPLah menunjukkan nilai mean sebesar 5,89%, dengan modus 0,00% yang mengindikasikan bahwa nilai yang paling sering muncul adalah sekolah yang sama sekali tidak menggunakan SIPLah. Nilai minimum berada di 0,00% dan maksimum mencapai 89,30%, menunjukkan kesenjangan yang sangat lebar dalam tingkat adopsi platform digital antar sekolah. Distribusi frekuensi data menunjukkan bahwa 15 sekolah atau 37,50% dari total sampel sama sekali tidak menggunakan SIPLah, 9 sekolah atau 22,50% berada di rentang penggunaan 0,01% hingga 2,50%, dan hanya 3 sekolah atau 7,50% yang berada di rentang 2,51% hingga 5,00%. Sekolah dengan penggunaan SIPLah di atas 10% hanya berjumlah 8 sekolah atau 20% dari total sampel, menunjukkan bahwa mayoritas besar sekolah berada pada tahap implementasi yang sangat rendah atau bahkan tidak mengimplementasikan sama sekali.

Berdasarkan konsultasi dengan Kasubbag Keuangan Dinas Pendidikan Kabupaten Deli Serdang dan pertimbangan distribusi data aktual, threshold klaster ditetapkan untuk memetakan sekolah ke dalam tiga kategori intervensi yang berbeda sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Threshold dan Interpretasi Klaster

Klaster	Threshold Persentase SIPLah	Interpretasi	Intervensi yang Direkomendasikan
Klaster 1 (Rendah)	0% - 5%	Implementasi Minimal/Tidak Ada	Diperlukan audit khusus dan pendampingan intensif
Klaster 2 (Sedang)	>5% - 15%	Implementasi Moderat	Diberikan pelatihan intensif untuk peningkatan
Klaster 3 (Tinggi)	>15%	Implementasi Baik	Dipertahankan dan dijadikan best practice

Penetapan threshold ini mempertimbangkan bahwa 60% sekolah memiliki persentase SIPLah kurang dari 2,50%, sehingga batas 5% untuk klaster rendah mencakup mayoritas sekolah yang memerlukan intervensi mendesak. Batas 15% untuk klaster sedang dipilih untuk mengidentifikasi sekolah yang sudah melewati fase adopsi awal namun masih memerlukan pembinaan untuk mencapai tingkat optimal, sementara threshold di atas 15% menandai sekolah dengan kinerja baik yang dapat dijadikan model *best practice*.

3.2 Penerapan K-Means++ untuk Menyelesaikan Permasalahan Segmentasi

Permasalahan utama yang dihadapi Dinas Pendidikan Kabupaten Deli Serdang adalah ketiadaan pemetaan objektif tentang karakteristik adopsi SIPLah di sekolah-sekolah, sehingga kebijakan pembinaan bersifat seragam tanpa mempertimbangkan kebutuhan spesifik setiap kelompok sekolah. Untuk menyelesaikan permasalahan ini, diperlukan metode yang mampu mengelompokkan sekolah secara otomatis berdasarkan pola belanja Dana BOS dengan kriteria yang terukur dan dapat dipertanggungjawabkan. Algoritma K-Means++ dipilih sebagai solusi karena tiga keunggulan utamanya dalam konteks penelitian ini. Pertama, algoritma ini mampu menangani data dengan distribusi tidak merata melalui strategi inisialisasi centroid yang cerdas menggunakan probabilistic sampling berbasis jarak kuadrat, sehingga cocok untuk data persentase SIPLah yang sangat timpang dimana mayoritas sekolah terkonsentrasi di rentang rendah. Kedua, K-Means++ menghasilkan klaster yang stabil dan konsisten karena menghindari local optima yang sering terjadi pada K-Means standar, memastikan hasil segmentasi dapat diandalkan untuk basis pengambilan keputusan kebijakan. Ketiga, algoritma ini efisien secara komputasional sehingga dapat diintegrasikan ke dalam sistem monitoring berkelanjutan yang memerlukan pembaruan hasil klasterisasi secara periodik tanpa beban komputasi yang berlebihan.

Penerapan algoritma dimulai dengan tahap preprocessing data yang mencakup validasi kelengkapan data, pengecekan missing values, dan normalisasi format persentase penggunaan SIPLah. Setelah memastikan kualitas data, tahap berikutnya adalah inisialisasi centroid menggunakan metode probabilistik D²-sampling yang menjadi keunggulan utama K-Means++ dibandingkan K-Means standar. Proses inisialisasi dimulai dengan pemilihan centroid pertama secara acak dari dataset, dalam penelitian ini sistem memilih SMP N 2 Hamparan Perak dengan persentase 1,83% sebagai

centroid awal karena posisinya yang representatif di zona rendah. Untuk memilih centroid kedua, algoritma menghitung jarak kuadrat D^2 dari setiap sekolah terhadap centroid pertama menggunakan formula $D^2 = (x_i - 1,83)^2$. Total jarak kuadrat keseluruhan adalah 8.243,17, dan setiap sekolah memiliki probabilitas terpilih sebanding dengan nilai D^2 -nya. Meskipun SMP N 5 Lubuk Pakam memiliki probabilitas tertinggi sebesar 92,82%, melalui proses sampling probabilistik, SMP N 3 Tanjung Morawa dengan persentase 10,96% terpilih sebagai centroid kedua dengan probabilitas 1,01%. Pemilihan ini menguntungkan karena posisinya yang representatif di zona sedang dengan rentang 5% hingga 15%. Untuk centroid ketiga, algoritma menghitung jarak setiap sekolah ke centroid terdekat di antara dua centroid yang sudah terpilih, menghasilkan total jarak minimum 6.438,22. SMP N 5 Lubuk Pakam akhirnya terpilih sebagai centroid ketiga dengan persentase 89,30% dan probabilitas dominan 95,33%, mewakili zona tinggi dalam distribusi data.

Setelah tiga centroid awal terbentuk, algoritma melanjutkan ke tahap iterasi standar K-Means yang terdiri dari assignment step dan update step. Pada iterasi pertama, setiap sekolah ditugaskan ke centroid terdekat berdasarkan perhitungan jarak Euclidean. Untuk data satu dimensi seperti persentase SIPLah, jarak Euclidean disederhanakan menjadi nilai absolut selisih antara nilai persentase sekolah dengan nilai centroid. Hasil assignment pertama menghasilkan Klaster 1 dengan 28 anggota yang mencakup sekolah dengan persentase 0% hingga 5,14%, Klaster 2 dengan 10 anggota di rentang 7,16% hingga 23,62%, dan Klaster 3 dengan 2 anggota di 89,30%. Tahap update step kemudian menghitung posisi centroid baru sebagai rata-rata dari semua anggota yang tergabung di masing-masing klaster. Centroid pertama bergeser dari 1,83% menjadi 0,69%, centroid kedua dari 10,96% menjadi 11,21%, sedangkan centroid ketiga tetap di 89,30% karena hanya memiliki satu anggota. Perubahan posisi centroid sebesar 1,14% pada klaster pertama dan 0,25% pada klaster kedua menunjukkan bahwa centroid awal hasil K-Means++ sudah cukup optimal, sehingga pergeseran yang terjadi relatif kecil. Pada iterasi kedua, dilakukan re-assignment dengan centroid baru, namun tidak ada satu pun sekolah yang berpindah klaster, mengindikasikan bahwa konvergensi telah tercapai. Kriteria konvergensi ditetapkan ketika tidak ada lagi perpindahan anggota klaster dan perubahan posisi centroid lebih kecil dari threshold yang ditetapkan.

Kontribusi metode K-Means++ terhadap penyelesaian permasalahan penelitian dapat dilihat dari tiga aspek utama yang saling mendukung. Aspek pertama adalah efisiensi komputasional dimana kecepatan konvergensi yang tercapai hanya dalam dua iterasi jauh lebih baik dibandingkan K-Means standar yang biasanya memerlukan lima hingga sepuluh iterasi untuk mencapai hasil yang stabil. Efisiensi ini sangat penting untuk implementasi sistem monitoring real-time yang memerlukan pembaruan hasil klasterisasi secara berkala tanpa membebankan sumber daya komputasi yang berlebihan. Aspek kedua adalah kualitas segmentasi dimana strategi inisialisasi centroid yang cerdas melalui probabilistic sampling menghasilkan klaster dengan karakteristik yang jelas dan terpisah dengan baik, terbukti dari nilai Silhouette Coefficient keseluruhan sebesar 0,687 yang berada dalam kategori struktur baik. Kualitas ini memastikan bahwa setiap kelompok sekolah memiliki karakteristik yang homogen internal namun berbeda signifikan antar kelompok, sehingga rekomendasi kebijakan dapat disesuaikan dengan tepat untuk setiap segmen. Aspek ketiga adalah stabilitas hasil dimana penghindaran terhadap local optima melalui inisialisasi cerdas memastikan bahwa hasil klasterisasi konsisten dan dapat diandalkan sebagai basis pengambilan keputusan strategis jangka panjang, bukan sekadar snapshot sesaat yang berpotensi berubah drastis pada iterasi berikutnya.

3.3 Hasil Klasterisasi dan Evaluasi

Tabel 2 menunjukkan distribusi akhir sekolah ke dalam tiga kelompok hasil klasterisasi. Yang paling mencolok dari distribusi ini adalah konsentrasi masif di Klaster 1 yang mencakup tiga perempat total sampel, mengindikasikan bahwa mayoritas sekolah berada dalam kondisi kritis dengan tingkat adopsi SIPLah yang sangat minimal atau bahkan nihil. Ketimpangan distribusi ini bukan anomali metodologis melainkan cerminan akurat dari realitas di lapangan dimana implementasi kebijakan digital menghadapi hambatan struktural yang serius. Centroid akhir Klaster 1 yang berada di 0,69% bahkan lebih rendah dari threshold minimum yang ditetapkan di 5%, menunjukkan bahwa sekolah-sekolah di kelompok ini tidak hanya gagal mencapai target minimal namun berada jauh di bawahnya dengan gap yang sangat signifikan.

Hasil detail klasterisasi yang mencakup persentase penggunaan SIPLah, jarak ke setiap centroid, serta penentuan label klaster final disajikan pada Tabel 2 dibawah ini :

Tabel 2. Distribusi Sekolah Per Klaster

Klaster	Jumlah Sekolah	Persentase dari Total	Range Persentase SIPLah	Centroid Akhir
Klaster 1 (Rendah)	30	75,00%	0,00% - 4,47%	0,69%
Klaster 2 (Sedang)	8	20,00%	5,14% - 14,65%	11,21%
Klaster 3 (Tinggi)	2	5,00%	23,62% - 89,30%	89,30%

Jika dilihat pada Tabel 2, Klaster 2 dengan centroid di 11,21% merepresentasikan sekolah yang telah melewati fase adopsi awal dan menunjukkan komitmen terhadap implementasi platform digital, meskipun masih jauh dari optimal. Range persentase yang relatif sempit antara 5,14% hingga 14,65% menunjukkan homogenitas karakteristik dalam kelompok ini, dimana sekolah-sekolah menghadapi tantangan serupa dalam meningkatkan adopsi dari level moderat ke level tinggi. Posisi centroid yang berada di tengah rentang threshold klaster sedang mengindikasikan bahwa kelompok ini memerlukan strategi akselerasi yang tepat untuk mencapai lompatan ke level berikutnya. Klaster 3 menampilkan fenomena yang menarik dimana hanya dua sekolah berhasil melampaui threshold 15% namun dengan gap yang sangat ekstrem antara 23,62% dan 89,30%. Heterogenitas internal yang tinggi ini tercermin dari range yang sangat lebar,

menunjukkan bahwa meskipun kedua sekolah sama-sama berkinerja tinggi, faktor-faktor pendorong keberhasilan dan tingkat pencapaian mereka sangat berbeda secara kuantitatif.

Evaluasi kualitas klaster menggunakan metrik Silhouette Coefficient menghasilkan temuan yang ditampilkan pada Tabel 3. Nilai rata-rata score keseluruhan sebesar 0,687 yang berada dalam kategori struktur baik memvalidasi bahwa proses klasterisasi berhasil memisahkan sekolah ke dalam kelompok yang bermakna dan dapat dibedakan dengan jelas. Namun analisis mendalam terhadap score per klaster mengungkap dinamika yang lebih kompleks. Klaster 2 menunjukkan performa terbaik dengan rata-rata score 0,782 dan konsistensi tinggi yang terlihat dari rentang sempit antara 0,692 hingga 0,821, mengindikasikan bahwa sekolah-sekolah di kelompok ini memiliki karakteristik yang sangat homogen dan terpisah jelas dari klaster lain. Homogenitas ini memudahkan perancangan intervensi karena kebutuhan dan tantangan yang dihadapi relatif seragam. Berikut Tabel 3 ditampilkan dibawah ini :

Tabel 3. Silhouette Coefficient Per Klaster

Klaster	Jumlah Sekolah	Rata-rata Silhouette Score	Min Score	Max Score	Interpretasi
Klaster 1	30	0,724	0,556	0,880	Struktur Baik
Klaster 2	8	0,782	0,692	0,821	Struktur Baik
Klaster 3	2	0,151	0,151	0,151	Struktur Lemah
Overall	40	0,687	0,151	0,880	Struktur Baik

Pada Tabel 3 Terlihat Klaster 1 meskipun memiliki rata-rata score yang baik di 0,724, menampilkan variabilitas yang lebih tinggi dengan rentang 0,556 hingga 0,880. Variabilitas ini wajar mengingat kelompok ini mencakup spektrum yang lebar dari sekolah dengan 0% adopsi hingga mendekati 5%, dimana sub-kelompok dengan 0% adopsi memiliki karakteristik yang berbeda dari sub-kelompok di 4-5% yang hampir mencapai threshold klaster sedang. Klaster 3 menunjukkan score terendah di 0,151 yang mengindikasikan struktur lemah, namun ini bukan kelemahan metodologis melainkan konsekuensi logis dari jumlah anggota yang sangat sedikit dan jarak yang sangat ekstrem antar anggota. Dengan hanya dua anggota yang memiliki gap 65,68 poin persentase, kohesi internal secara matematis memang rendah, namun pengelompokan keduanya tetap valid karena sama-sama merepresentasikan outlier positif yang jauh dari mayoritas data.

Perbandingan dengan penelitian terdahulu sebagaimana ditampilkan pada Tabel 4 memberikan konteks penting untuk memahami kualitas dan keterbatasan hasil yang dicapai. Nilai Silhouette Coefficient penelitian ini sebesar 0,687 memang lebih rendah dibandingkan penelitian Nugroho dan Adhinata yang mencapai 0,882, namun perbedaan ini dapat dijelaskan secara sistematis melalui analisis karakteristik data yang fundamental berbeda. Penelitian Nugroho dan Adhinata menggunakan tiga variabel yang saling berkorelasi dengan distribusi yang relatif merata di seluruh kabupaten dan kota Pulau Jawa, menciptakan struktur data yang naturally conducive untuk klasterisasi berkualitas tinggi. Sebaliknya, penelitian ini berhadapan dengan distribusi yang sangat skewed dimana tiga perempat data terkonsentrasi di satu klaster pada rentang sempit, ditambah keberadaan outlier ekstrem yang menciptakan tantangan struktural dalam pembentukan klaster kohesif.

Berikut kami sajikan perbandingan hasil dengan penelitian terdahulu yang bisa dilihat pada Tabel 4 dibawah ini :

Tabel 4. Perbandingan Hasil dengan Penelitian Terdahulu

Aspek	Penelitian ini	Nugroho & Adhinata (2022)	Syahkur et al. (2024)
Domain Aplikasi	Pola Belanja Dana BOS	Clustering COVID-19	Pengelompokan Gizi Balita
Jumlah Data	40 sekolah	119 kabupaten/kota	Data gizi balita
Jumlah Fitur	1 (Persentase SIPLah)	3 (kasus positif, vaksin dosis 1 & 2)	Multiple fitur gizi
Jumlah Klaster Optimal	3	5	2
Silhouette Coefficient	0,687	0,882	Tidak dilaporkan (DBI: 0,682)
Jumlah Iterasi	2	Tidak dilaporkan	Tidak dilaporkan
Distribusi Data	Sangat tidak merata (75% di satu klaster)	Relatif merata	Relatif merata
Karakteristik Khusus	Outlier ekstrem (89,30%)	Distribusi normal	Distribusi normal

Tabel 4 menunjukkan meskipun nilai Silhouette lebih rendah, penelitian ini menunjukkan keunggulan signifikan dalam efisiensi konvergensi dengan hanya dua iterasi dibandingkan lima hingga sepuluh iterasi yang umumnya dibutuhkan K-Means standar. Keunggulan ini membuktikan bahwa strategi inisialisasi cerdas K-Means++ tetap efektif bahkan pada data dengan karakteristik challenging seperti distribusi sangat tidak merata dan keberadaan outlier ekstrem. Perbandingan dengan penelitian Syahkur dkk yang menggunakan Davies-Bouldin Index sebesar 0,682 menunjukkan konsistensi bahwa K-Means++ mampu menghasilkan klaster berkualitas baik pada berbagai domain aplikasi dengan karakteristik data yang berbeda. Nilai Silhouette 0,687 yang sedikit lebih tinggi dari DBI Syahkur mengindikasikan bahwa meskipun menghadapi tantangan struktural data, hasil klasterisasi penelitian ini tetap robust dan dapat diandalkan untuk tujuan segmentasi sekolah berdasarkan pola belanja Dana BOS.

3.4 Interpretasi dan Implikasi Hasil

Temuan paling kritis yang terungkap dari hasil klasterisasi adalah bahwa setengah dari sekolah di kelompok adopsi rendah sama sekali tidak pernah melakukan satu pun transaksi melalui SIPLah meskipun platform telah diwajibkan sejak tahun 2022 melalui Permendikbudristek Nomor 18 Tahun 2022. Fenomena ini bukan sekadar angka statistik melainkan indikasi kegagalan sistemik dalam implementasi mandat kebijakan digital di tingkat lokal yang memerlukan perhatian serius dari pengambil kebijakan. Rata-rata penggunaan SIPLah keseluruhan yang hanya 2,18% dari total belanja Rp43,4 miliar mengungkap fakta bahwa hampir 98% transaksi Dana BOS masih dilakukan secara manual di luar platform resmi, menciptakan risiko serius terhadap transparansi dan akuntabilitas pengelolaan dana pendidikan karena transaksi manual tidak terdokumentasi secara digital dan sulit dilacak oleh sistem pengawasan.

Perbandingan kontras antara SMP N 5 Lubuk Pakam dan SMP N 1 Batang Kuis mengungkap insight fundamental tentang faktor penentu keberhasilan implementasi SIPLah yang bukan bersumber dari ketersediaan sumber daya finansial. Sekolah pertama dengan realisasi dana yang sangat modest hanya Rp74,37 juta berhasil mencapai penetrasi platform 89,30%, sementara sekolah kedua dengan realisasi hampir lima belas kali lipat yaitu Rp1,08 miliar justru tidak menggunakan platform sama sekali. Paradoks ini mengkonfirmasi bahwa variabel kunci keberhasilan terletak pada dimensi non-finansial yaitu komitmen kepemimpinan kepala sekolah dan bendahara dalam mengadopsi transformasi digital, serta kapasitas sumber daya manusia dalam literasi digital dan pemahaman sistem e-procurement. Temuan ini memiliki implikasi strategis bahwa intervensi pemerintah daerah seharusnya mengalokasikan sumber daya lebih besar untuk program capacity building dan change management daripada sekadar penyediaan infrastruktur atau anggaran tambahan yang terbukti bukan faktor determinan utama.

Implikasi kebijakan untuk kelompok adopsi rendah yang mencakup tiga perempat total sekolah memerlukan pendekatan intervensi yang komprehensif dan multi-level. Audit mendalam harus dilakukan untuk mengidentifikasi akar penyebab non-adopsi yang mencakup dimensi infrastruktur seperti verifikasi kualitas konektivitas internet, dimensi kapasitas seperti penilaian literasi digital kepala sekolah dan bendahara, serta dimensi persepsi seperti investigasi terhadap hambatan psikologis dan resistensi terhadap perubahan. Khusus untuk sub-kelompok 15 sekolah dengan 0% adopsi yang merepresentasikan kasus paling kritis, diperlukan program pendampingan khusus dimana fasilitator dari Dinas Pendidikan melakukan hand-holding langsung dari transaksi pertama hingga sekolah mampu menyelesaikan minimal lima transaksi penuh secara mandiri. Program pendampingan harus dirancang dengan timeline jelas dan mekanisme evaluasi berkala untuk memastikan transfer knowledge yang efektif dan keberlanjutan adopsi setelah pendampingan berakhir, menghindari ketergantungan jangka panjang terhadap fasilitator eksternal.

Kelompok adopsi sedang yang telah melewati fase kritis awal memerlukan strategi akselerasi yang berbeda berupa pelatihan tingkat lanjut yang fokus pada optimalisasi penggunaan platform. Konten pelatihan harus melampaui tutorial dasar dan mencakup modul-modul strategis seperti teknik negosiasi harga dengan vendor untuk mendapatkan penawaran kompetitif, prosedur penyelesaian masalah teknis umum seperti transaksi menggantung yang sering menjadi sumber frustrasi, pemahaman mendalam tentang regulasi perpajakan PMK 58/2022 yang mengatur aspek legal transaksi digital, serta best practice manajemen dokumen digital untuk efisiensi administratif. Penetapan target terukur sangat krusial untuk memberikan arah jelas bagi sekolah, misalnya kenaikan penggunaan SIPLah minimal menjadi 20% pada semester pertama dan 30% pada semester kedua, dengan mekanisme monitoring bulanan untuk tracking progress dan identifikasi dini hambatan yang muncul.

Strategi untuk kelompok adopsi tinggi harus fokus pada leveraging keberhasilan mereka sebagai model pembelajaran bagi sekolah lain melalui mekanisme peer learning yang terbukti lebih efektif dibanding pelatihan formal top-down. Pemberian penghargaan publik dari Bupati Deli Serdang sebagai Satuan Pendidikan Transparan dan Akuntabel akan memberikan insentif non-finansial yang bermakna sekaligus menciptakan social pressure positif bagi sekolah lain untuk meningkatkan kinerja. Sekolah-sekolah ini perlu difasilitasi untuk mendokumentasikan praktik terbaik mereka dalam bentuk panduan operasional yang konkret dan kontekstual, bukan sekadar narasi sukses yang inspiratif namun tidak actionable. Keterlibatan mereka sebagai narasumber dalam forum-forum pelatihan kabupaten atau provinsi harus difasilitasi dengan proper incentive structure untuk memastikan partisipasi berkelanjutan, dan dibentuk mekanisme mentoring terstruktur dimana setiap sekolah berkinerja tinggi mendampingi dua hingga tiga sekolah berkinerja rendah dalam program sister school yang dimonitor ketat oleh Dinas Pendidikan.

3.5 Analisis Kendala Implementasi SIPLah

Berdasarkan hasil klasterisasi dan analisis laporan Dinas Pendidikan, teridentifikasi empat kategori kendala utama yang bersifat saling terkait dan memperkuat satu sama lain:

Kendala Infrastruktural. Konektivitas internet yang buruk atau tidak ada merupakan hambatan paling fundamental, terutama untuk sekolah di zona risiko tinggi seperti Gunung Meriah, Sibolangit, dan STM Hulu. Dari 30 sekolah di Klaster 1, estimasi 40-50% berada di kecamatan dengan profil risiko tinggi yang menghadapi kesenjangan digital signifikan. Program pemerintah kabupaten seperti "Kampung Internet" mengonfirmasi bahwa konektivitas belum merata dan masih menjadi prioritas kebijakan. Tanpa infrastruktur internet yang memadai, sekolah tidak dapat mengakses platform SIPLah meskipun memiliki komitmen dan kapasitas SDM yang baik.

Kendala Ekonomi. Harga barang di SIPLah yang sering kali lebih mahal dibanding toko lokal menciptakan konflik langsung antara regulasi dan prinsip manajemen keuangan. Dari data distribusi UMKM di 22 kecamatan, terlihat kesenjangan tajam: Tanjung Morawa memiliki 23.196 UMKM, sementara Gunung Meriah hanya 473 UMKM. Minimnya

vendor lokal di SIPLah memaksa sekolah di wilayah itu untuk membeli dari vendor luar daerah dengan ongkir tinggi, menjadikan total harga tidak kompetitif. Ini menciptakan siklus negatif: sekolah tidak menggunakan SIPLah karena mahal, volume transaksi rendah, UMKM lokal tidak tertarik join platform, vendor lokal makin sedikit, harga makin tidak kompetitif, dan sekolah makin tidak tertarik.

Kendala Kapasitas SDM. Penggunaan SIPLah bukan sekadar keterampilan komputer dasar, melainkan literasi digital administratif yang mencakup pemahaman tentang e-procurement, regulasi perpajakan (PMK 58/2022), dan manajemen dokumen digital. Tingkat keahlian ini tidak dimiliki oleh banyak bendahara atau kepala sekolah. Fenomena "transaksi menggantung" (transaksi yang tidak selesai) adalah gejala paling jelas dari kesenjangan kapasitas. Pengguna berhasil memulai transaksi namun gagal menyelesaikan langkah krusial seperti validasi BAST atau finalisasi pembayaran. Ini menciptakan rasa frustrasi yang membuat pengguna enggan mencoba lagi.

Kendala Persepsi. Resistensi terhadap perubahan dan kenyamanan dengan pola belanja manual yang sudah familiar menjadi hambatan psikologis. Banyak kepala sekolah dan bendahara merasa bahwa sistem lama (belanja manual ke toko langganan) lebih mudah, fleksibel, dan tidak ribet. Ada ketakutan terhadap transparansi penuh yang dituntut SIPLah, karena setiap transaksi tercatat digital dan dapat diaudit. Bagi sebagian pihak, transparansi ini dianggap mengurangi fleksibilitas dalam pengelolaan dana.

3.6 Rekomendasi Kebijakan Berdasarkan Hasil Klusterisasi

Rekomendasi untuk Klaster 1 (Rendah: 0-5%). Target: 30 sekolah (75% sampel). Intervensi Prioritas: Audit Khusus dan Pendampingan Intensif. Dinas Pendidikan harus melakukan audit mendalam untuk mengidentifikasi akar penyebab non-adopsi di setiap sekolah, mencakup: (1) Audit Infrastruktur untuk verifikasi ketersediaan dan kualitas konektivitas internet; (2) Audit Kapasitas SDM untuk mengukur literasi digital kepala sekolah dan bendahara. Untuk 15 sekolah dengan persentase 0% (tidak ada transaksi sama sekali), diperlukan pendampingan khusus: fasilitator mendampingi langsung proses transaksi pertama di SIPLah dan dampingi hingga sekolah berhasil menyelesaikan minimal 5 transaksi penuh (dari pemesanan hingga pembayaran).

Rekomendasi untuk Klaster 2 (Sedang: >5-15%). Target: 8 sekolah (20% sampel). Intervensi Prioritas: Pelatihan Intensif untuk Tingkat Lanjut. Sekolah di klaster ini sudah melewati fase adopsi awal. Fokus pelatihan adalah tingkat lanjut: (1) Modul 1: Strategi negosiasi harga dengan vendor untuk mendapat deal terbaik; (2) Modul 2: Penyelesaian masalah umum (transaksi menggantung dan perpajakan). Set target realistis untuk setiap sekolah: Semester 1 minimal naik ke 20% penggunaan SIPLah, Semester 2 target 30%.

Rekomendasi untuk Klaster 3 (Tinggi: >15%). Target: 2 sekolah (5% sampel). Intervensi Prioritas: Replikasi Best Practice dan Apresiasi. Pemberian Penghargaan Satuan Pendidikan Transparan dan Akuntabel dari Bupati Deli Serdang, undangan sebagai narasumber SIPLah di forum-forum kabupaten/provinsi, dan plakat penghargaan yang dipasang di sekolah. Jadikan sekolah ini sebagai mentor bagi sekolah di Klaster 1 dan 2, dokumentasikan best practice mereka untuk disebarluaskan.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means++ untuk mengklusterisasi pola belanja Dana BOS SMP Negeri di Kabupaten Deli Serdang berdasarkan persentase penggunaan platform SIPLah. Hasil klusterisasi dari 40 sampel sekolah menghasilkan tiga kelompok yang memiliki karakteristik berbeda secara signifikan. Klaster 1 (Rendah) beranggotakan 30 sekolah (75%) dengan rata-rata penggunaan SIPLah hanya 0,69% dan rentang persentase 0,00% hingga 4,47%, dimana 15 sekolah di dalamnya sama sekali tidak pernah bertransaksi melalui SIPLah. Klaster 2 (Sedang) terdiri dari 8 sekolah (20%) dengan centroid 11,21% dan rentang 5,14% hingga 14,65%, menunjukkan adopsi yang moderat. Sementara itu, Klaster 3 (Tinggi) hanya beranggotakan 2 sekolah (5%) dengan centroid 89,30%, menjadikannya contoh penerapan yang baik. Kualitas pengelompokan ini divalidasi dengan Silhouette Coefficient sebesar 0,687 yang mengindikasikan struktur klaster yang baik, didukung oleh konvergensi algoritma yang cepat hanya dalam 2 iterasi berkat inisialisasi centroid yang optimal metode K-Means++. Temuan inti penelitian mengungkap fakta kritis bahwa rata-rata penggunaan SIPLah secara keseluruhan hanya mencapai 2,18% dari total belanja Dana BOS sebesar Rp43,4 miliar, yang berarti 97,82% transaksi masih dilakukan secara manual di luar platform. Hasil ini memberikan kontribusi berupa pemetaan objektif dan berbasis data pertama kali yang mengidentifikasi dengan jelas kesenjangan implementasi kebijakan digital di tingkat sekolah, sekaligus menyediakan dasar empiris bagi Dinas Pendidikan untuk menyusun intervensi yang tepat sasaran, terukur, dan berbeda untuk setiap kelompok sekolah guna mendorong transparansi dan akuntabilitas pengelolaan keuangan pendidikan.

REFERENCES

- [1] M. Umdatul and Z. A. Siti Apipah, "Attractive : Innovative Education Journal," *Students' Difficulties Elem. Sch. Increasing Lit. Abil.*, vol. 4, no. 1, pp. 1–12, 2024.
- [2] G. Sania, I. Zufria, and M. Fakhriza, "Penerapan Metode AHP dan SAW Pada Sistem Pendukung Keputusan Untuk Menentukan Pengalokasian Dana BOS," *J. Inf. Technol.*, vol. 4, no. 1, pp. 117–125, 2024, doi: 10.46229/jifotech.v4i1.869.
- [3] K. M. Tahir, "KEBIJAKAN ALOKASI PEMBAGIAN DANA BOS SESUAI JUMLAH STANDAR MINIMAL BOS FUNDING ALLOCATION POLICY ACCORDING TO MINIMUM pendidikan , seperti gaji guru dan gaji tenaga Pemberian

- bantuan BOS mengacu pada yang siswanya sedikit mendapatkan alokasi Anggaran Pend,” pp. 79–98, 2023.
- [4] D. S. Diantari, S. O. M. Bani, A. S. Fatiqh, and S. Trihantoyo, “Pengoperasian Aplikasi RKAS Dalam Pengelolaan Dana Bantuan Operasional Sekolah (BOS),” *J. Ilmu Pendidik.*, vol. 1, no. 2, pp. 280–299, 2024, [Online]. Available: <https://doi.org/10.62383/hardik.v1i2.370>
 - [5] D. Herinanto, M. Gumanti, and B. H. S. Utami, “Pemanfaatan SIPLah Dalam Memasarkan Produk Ekonomi Kreatif Bagi Para Pengusaha Mikro, Kecil, dan Menengah di Desa Podosari,” *SWARNA J. Pengabd. Kpd. Masy.*, vol. 3, no. 3, pp. 280–289, 2024, doi: 10.55681/swarna.v3i3.1247.
 - [6] Nugroho and F. D. Adhinata, “Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa,” *Teknika*, vol. 11, no. 3, pp. 170–179, 2022, doi: 10.34148/teknika.v11i3.502.
 - [7] Y. Li, “User Role Discovery and Optimization Method Based on K-means++ and Reinforcement Learning in Mobile Applications,” *C. - Comput. Model. Eng. Sci.*, vol. 131, no. 3, pp. 1365–1386, 2022, doi: 10.32604/cmescs.2022.019656.
 - [8] C. V. Addad, S. Lattanzi, N.-A. Fard, C. Sohler, and O. Svensson, “Fast and accurate k-means++ via rejection sampling,” *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.
 - [9] B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, “Scalable κ -means++,” *Proc. VLDB Endow.*, vol. 5, no. 7, pp. 622–633, 2024, doi: 10.14778/2180912.2180915.
 - [10] A. Deshpande, P. Kacham, and R. Pratap, “Robust k-means++,” *Proc. 36th Conf. Uncertain. Artif. Intell. UAI 2020*, vol. 124, pp. 819–828, 2020.
 - [11] J. Hämäläinen, T. Kärkkäinen, and T. Rossi, “Improving scalable k-means++,” *Algorithms*, vol. 14, no. 1, pp. 1–20, 2021, doi: 10.3390/a14010006.
 - [12] M. R. Syahkur, D. Hartama, and Solikhun, “Evaluasi Jumlah Cluster pada Algoritma K-Means ++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita,” vol. 13, no. 3, pp. 487–496, 2024.
 - [13] F. Nuraeni, H. Susilawati, and Y. H. Agustin, “Perbandingan Implementasi Algoritma K-Means++ Dan Fuzzy C-Means Pada Segmentasi Citra Wajah,” *JuTI “Jurnal Teknol. Informasi,”* vol. 1, no. 2, p. 47, 2023, doi: 10.26798/juti.v1i2.722.
 - [14] F. A. Zurfani, Sawaluddin, Mardiningsih, and M. R. Syahputra, “Analisis Metode Clustering K-Means pada Zonasi Daerah Terdampak Banjir di Kota Medan dengan Evaluasi Silhouette Coefficient bahkan menjadi tradisi tahunan . Telah banyak upaya dilakukan untuk menanggulangnya,” no. 6, 2024.
 - [15] Mohammad Ferdiansyah and Umi Chotijah, “Implementasi Algoritme K-Means++ Untuk Clustering Penjualan Bahan Bangunan,” *J. Ilm. Tek. Inform. dan Komun.*, vol. 4, no. 1, pp. 181–193, 2024, doi: 10.55606/juitik.v4i1.767.
 - [16] I. S. Tinendung and I. Zufria, “Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering,” *J. Media Inform. Budidarma*, vol. 7, no. 4, p. 2014, 2023, doi: 10.30865/mib.v7i4.6908.
 - [17] F. D. Pratama, I. Zufria, and Triase, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 77–84, 2022, doi: 10.36341/rabit.v7i1.2217.
 - [18] A. A. Tunru, R. Ilahi, and N. Hikmah, “Analisis Minat Belajar Siswa Pada Mata Pelajaran Pendidikan Jasmani Di SDN 027 Samarinda Ulu Kota Samarinda,” *J. Pendidik. IPS*, vol. 4, no. 2, pp. 53–60, 2023, [Online]. Available: [http://jurnal.ut.ac.id/index.php/jp/search/authors/view?givenName=Mery Noviyanti &familyName=&affiliation=Universitas Terbuka&country=ID&authorName=Mery Noviyanti](http://jurnal.ut.ac.id/index.php/jp/search/authors/view?givenName=Mery%20Noviyanti%20&familyName=&affiliation=Universitas%20Terbuka&country=ID&authorName=Mery%20Noviyanti)
 - [19] E. Kavlakoglu and V. Winland, “Apa itu k-means clustering? | IBM,” IBM INDONESIA. Accessed: Apr. 18, 2025. [Online]. Available: <https://www.ibm.com/id-id/topics/k-means-clustering#f5>
 - [20] N. Jain, “Menguasai pengelompokan data: Panduan Anda untuk K-means & K-means++.” Accessed: Apr. 18, 2025. [Online]. Available: <https://www.aiacceleratorinstitute.com/mastering-data-clustering-your-comprehensive-guide-to-k-means-and-k-means/>