

**ANALISIS PERBANDINGAN PERFORMA ALGORITMA K-MEANS &  
PENINGKATAN DBSCAN PADA CLUSTERING DATA  
PENGELOMPOKAN SEKOLAH DI SULAWESI SELATAN UNTUK  
PENYALURAN MERATA BOP (BANTUAN OPERASIONAL  
PENDIDIKAN) : STUDI KASUS DINAS PENDIDIKAN SULAWESI  
SELATAN**



**MUHAMMAD ZULFIKAR ARIS  
D121 19 1064**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2025**

**ANALISIS PERBANDINGAN PERFORMA ALGORITMA K-MEANS &  
PENINGKATAN DBSCAN PADA CLUSTERING DATA  
PENGELOMPOKAN SEKOLAH DI SULAWESI SELATAN UNTUK  
PENYALURAN MERATA BOP (BANTUAN OPERASIONAL  
PENDIDIKAN) : STUDI KASUS DINAS PENDIDIKAN SULAWESI  
SELATAN**

**MUHAMMAD ZULFIKAR ARIS  
D121 19 1064**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2025**

**ANALISIS PERBANDINGAN PERFORMA ALGORITMA K-MEANS &  
PENINGKATAN DBSCAN PADA CLUSTERING DATA  
PENGELOMPOKAN SEKOLAH DI SULAWESI SELATAN UNTUK  
PENYALURAN MERATA BOP (BANTUAN OPERASIONAL  
PENDIDIKAN) : STUDI KASUS DINAS PENDIDIKAN SULAWESI  
SELATAN**

MUHAMMAD ZULFIKAR ARIS  
D121 19 1064

Skripsi

Sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Sarjana Teknik Informatika

pada

**DEPARTEMEN TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2025**

## SKRIPSI

# ANALISIS PERBANDINGAN PERFORMA ALGORITMA K-MEANS & PENINGKATAN DBSCAN PADA CLUSTERING DATA PENGELOMPOKAN SEKOLAH DI SULAWESI SELATAN UNTUK PENYALURAN MERATA BOP (BANTUAN OPERASIONAL PENDIDIKAN) : STUDI KASUS DINAS PENDIDIKAN SULAWESI SELATAN

MUHAMMAD ZULFIKAR ARIS  
D121 19 1064

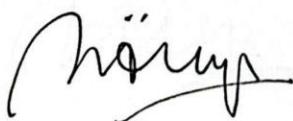
Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Teknik Informatika pada  
28/05/2025 dan dinyatakan telah memenuhi syarat kelulusan

Program Studi Sarjana Teknik Informatika  
Departemen Teknik Informatika  
Fakultas Teknik  
Universitas Hasanuddin  
Makassar

Mengesahkan,

Pembimbing Utama



Novy N.R.A. Mokobombang, S.T.,  
Ms.T.M., Ph.D.  
NIP 19721114 200501 2 001

Ketua Program Studi,



Prof. Dr.-Ir. Indrabayu, ST., MT.,  
M.BUS.SYS, MPM, ASEAN, Eng.  
NIP 19750716 200212 1 004

## PERNYATAAN KEASLIAN

Dengan ini saya menyatakan bahwa, skripsi berjudul "Analisis Perbandingan Performa Algoritma K-Means dan DBSCAN pada Clustering Data Pengelompokan Sekolah di Sulawesi Selatan Untuk Penyaluran Merata BOP (Bantuan Operasional Pendidikan) : Studi Kasus Dinas Pendidikan Sulawesi Selatan" adalah benar karya saya dengan arahan dari pembimbing Novy N.R.A. Mokobombang, S.T., Ms.T.M., Ph.D. sebagai Pembimbing Utama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Gowa, 03 Juni 2025

Yang Menyatakan



Muhammad Zulfikar Aris  
D121191064

## UCAPAN TERIMA KASIH

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga saya dapat menyelesaikan skripsi yang berjudul "Analisis Perbandingan Performa Algoritma K-Means & Peningkatan DBSCAN pada Clustering Data Pengelompokan Sekolah di Sulawesi Selatan untuk Penyaluran Merata BOP (Bantuan Operasional Pendidikan) : Studi Kasus Dinas Pendidikan Sulawesi Selatan" dengan lancar. Penelitian ini disusun sebagai salah satu syarat untuk meraih gelar sarjana pada Departement Teknik Informatika, Universitas Hasanuddin dan Fakultas Anda.

Penelitian yang saya lakukan dapat terlaksana dengan baik atas bimbingan dan arahan dari ibu Novy N.R.A. Mokobombang, S.T., Ms.T.M., Ph.D. sebagai pembimbing utama saya. Saya mengucapkan banyak terima kasih kepada beliau. Ucapan terima kasih juga saya sampaikan kepada segenap dosen dan Staff Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin yang telah membantu kelancaran selama proses perkuliahan.

Terima kasih dan Penghargaan yang tinggi juga saya sampaikan kepada seluruh staff Dinas Pendidikan Sulawesi Selatan 2023/2024 khususnya bidang PTK dan Fasilitas yang telah membantu penulis selama proses pengumpulan data.

Akhirnya, kepada ketua orang tua tercinta, ibu Dra. Nurmini dan Alm. Bapak Drs. Muh Aris M.Si. saya mengucapkan banyak terima kasih atas dukungan dan do'a restu nya yang mengalir tiada henti kepada penulis. Tidak lupa juga, terima kasih penulis ucapkan kepada kakak penulis, Anita Lestari Aris S.E., Dr. Valentino Aris S.Kom., MM., Arini Lestari Aris S.E., M.M., dan Ardianti Lestari Aris S.E. Terima kasih yang besar juga saya sampaikan ke teman-teman Teknik Informatika Angkatan 2019 Universitas Hasanuddin selaku rekan yang telah memberi bantuan, dukungan dan semangat selama masa perkuliahan dan penyusunan tugas akhir ini. Penghargaan yang besar juga untuk teman-teman Founders dan CO-Founders CV. Digisaurus Juara Digital dan IndigoHub Makassar, khususnya Khairul Umam, ST, MT. selaku Mentor diluar kampus,Akmal Hidayat,Chaerunnisa,Sunarti,Arsal, dan Akmal Jaya yang senantiasa memberi dukungan mental dalam pembuatan tugas akhir ini. Terima kasih selanjutnya penulis ucapkan untuk klub bola kesayangan penulis yaitu FC Barcelona yang senantiasa memberi hiburan dan semangat kepada penulis. Akhir kata, Penulis berharap semoga Tuhan Yang Maha Esa berkenan untuk membalas segala kebaikan dari semua pihak yang terlibat. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, penulis berharap segala bentuk kritik, saran, dan masukan yang membangun dari berbagai pihak, dan juga semoga tugas akhir ini dapat bermanfaat bagi para pembaca.

Gowa, 20 Januari 2025  
Penulis,

Muhammad Zulfikar Aris

## ABSTRAK

MUHAMMAD ZULFIKAR ARIS. **Analisis Perbandingan Performa Algoritma K-Means & Peningkatan DBSCAN pada Clustering Data Pengelompokan Sekolah di Sulawesi Selatan untuk Penyaluran Merata BOP (Bantuan Operasional Pendidikan) : Studi Kasus Dinas Pendidikan Sulawesi Selatan** (dibimbing oleh Novy N.R.A. Mokobombang, S.T., Ms.T.M., Ph.D.).

**Latar Belakang.** Pemerataan BOP Bantuan Operasional Pendidikan merupakan tugas berat bagi Dinas Pendidikan Sulawesi Selatan mengingat setiap sekolah di provinsi ini memiliki perbedaan dalam berbagai aspek sosial, ekonomi, dan geografis. Penelitian ini akan membahas dan membandingkan kinerja peningkatan DBSCAN dan algoritma K-Means Only dalam proses pengelompokan data sekolah sebagai upaya untuk mengoptimalkan penyaluran BOP agar mencapai paritas dan ketepatan sasaran yang maksimal. Pencatatan meliputi 9 variabel utama yang menjadi kriteria bantuan BOP yaitu luas wilayah, daya listrik, jumlah ruang, jumlah pegawai, jumlah guru, jumlah siswa, dan koordinat 9201 sekolah di Sulawesi Selatan beserta garis lintang dan garis bujurnya. **Metode** ini diawali dengan melakukan proses data praprocessing berupa penggabungan dataset, pembersihan data, penskalaan menggunakan StandardScaler, normalisasi dan pengurangan dimensionalitas menggunakan teknik Principal Component Analysis untuk menyederhanakan variabel dan mempertahankan data penting. Penyempurnaan lebih lanjut dari proses ini dipastikan melalui pengelompokan yang dilakukan menggunakan algoritma K-Means dan peningkatan DBSCAN. Pengujian dilakukan menggunakan uji *Silhouette Score*, sedangkan parameter epsilon dan MinPts untuk DBSCAN dihitung menggunakan K-Nearest Neighbors (KNN). Hasil penelitian membuktikan bahwa algoritma peningkatan DBSCAN-lah yang dapat mengelompokkan sekolah dengan kepadatan data yang bervariasi secara efektif, dan pada saat yang sama dapat mengidentifikasi outlier/noise secara akurat. Sedangkan K-Means terbukti efisien dalam membentuk klaster pada data yang terstruktur, namun kesulitan menangkap pola non-linear. Validasi menggunakan *Silhouette Score* mengonfirmasi bahwa peningkatan DBSCAN memperoleh skor lebih tinggi dibandingkan K-Means, menandakan performa clustering yang lebih baik. Dengan demikian, Kajian ini diharapkan dapat memberikan dampak positif bagi Dinas Pendidikan Sulawesi Selatan dalam memperkuat kebijakan pengurangan kesenjangan pendidikan dan meningkatkan efisiensi penyaluran dana BOP atau bentuk bantuan lainnya. Kajian ini juga akan memperkaya khasanah penelitian ilmiah melalui penerapan teknik klasterisasi dalam sektor pendidikan.

Kata Kunci: *K-Means*, Peningkatan DBSCAN, *Clustering*, BOP, Sulawesi Selatan, *Silhouette Score*

## ABSTRACT

MUHAMMAD ZULFIKAR ARIS. **Performance Comparison Analysis of K-Means Algorithm & Enhanced DBSCAN in Clustering School Data in South Sulawesi for Equitable Distribution of BOP (Educational Operational Assistance): A Case Study of the South Sulawesi Education Office** (supervised by Novy N.R.A. Mokobombang, S.T., Ms.T.M., Ph.D.).

**Background.** Ensuring the equitable distribution of BOP (Educational Operational Assistance) is a significant challenge for the South Sulawesi Education Office, given the diverse social, economic, and geographical aspects of schools in the province. This study aims to analyze and compare the performance of the enhanced DBSCAN and the standard K-Means algorithm in clustering school data to optimize the allocation of BOP, ensuring maximum parity and precision in distribution. The dataset includes nine key variables that determine BOP eligibility: area size, electricity capacity, number of rooms, number of staff, number of teachers, number of students, and geographic coordinates (latitude and longitude) of 9,202 schools across South Sulawesi. **The methodology** begins with data preprocessing, which involves dataset merging, data cleaning, scaling using StandardScaler, normalization, and dimensionality reduction via Principal Component Analysis (PCA) to simplify variables while retaining essential data. The refined dataset is then clustered using the K-Means algorithm and enhanced DBSCAN. The clustering performance is evaluated using the *Silhouette Score*, while the epsilon and MinPts parameters for DBSCAN are determined using the K-Nearest Neighbors (KNN) method. The results indicate that the enhanced DBSCAN algorithm effectively clusters schools with varying data densities while accurately identifying outliers/noise. In contrast, K-Means efficiently forms clusters in structured data but struggles to capture non-linear patterns. The validation using the *Silhouette Score* confirms that enhanced DBSCAN achieves a higher score than K-Means, signifying superior clustering performance. This study is expected to provide valuable insights for the South Sulawesi Education Office in refining policies to reduce educational disparities and enhance the efficiency of BOP or other financial aid distribution. Additionally, it contributes to the scientific literature by applying clustering techniques in the education sector.

Keywords: *K-Means*, Enhanced DBSCAN, *Clustering*, BOP, South Sulawesi, *Silhouette Score*

**DAFTAR ISI**

	Halaman
HALAMAN JUDUL .....	i
PERNYATAAN PENGAJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN .....	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK .....	vi
ABSTRACT .....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR .....	x
DAFTAR TABEL .....	xv
DAFTAR SINGKATAN DAN ARTI SIMBOL.....	xvi
DAFTAR LAMPIRAN .....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Teori.....	3
1.3 Rumusan Masalah .....	15
1.4 Tujuan Penelitian.....	16
1.5 Manfaat Penelitian.....	16
1.6 Ruang Lingkup .....	17
BAB II METODE PENELITIAN.....	18
2.1 Lokasi dan Waktu Penelitian .....	18
2.2 Instrumen Penelitian.....	19
2.3 Objek Penulisan .....	19
2.4 Diagram Alur Kerja .....	21
2.5 Gambaran Umum Sistem.....	25
BAB III HASIL DAN PEMBAHASAN .....	28
3.1 Pengumpulan Data.....	28
3.2 Data Preprocessing.....	28
3.3 Normalisasi Data.....	32
3.4 Principal Component Analysis (PCA) .....	34

3.5 DBSCAN Clustering .....	36
3.6 Peningkatan DBSCAN Dengan K-Means Clustering.....	43
3.7 Metode K-Means Only.....	49
3.8 Perbandingan Hasil Grafik Algoritma Peningkatan DBSCAN dan K-Means Only55	
3.9 Perbandingan Hasil Uji Validasi DBSCAN Only, Peningkatan DBSCAN, dan K-Means Only .....	56
4.10 Uji Validasi Pembobotan Menggunakan Klaster Optimal Peningkatan DBSCAN (3 klaster) .....	57
4.11 Uji Validasi Pembobotan Menggunakan Klaster Optimal K-Means Only (2 klaster).....	62
4.12 Perbandingan Uji Validasi Pembobotan Klaster Optimal Peningkatan DBSCAN (3 klaster) dan K-Means Only (2 klaster).....	66
4.13 Pembuatan Peta Interaktif Menggunakan Pustaka Folium .....	67
BAB 4 KESIMPULAN DAN SARAN.....	71
4.1 Kesimpulan.....	71
4.2 Saran.....	72
DAFTAR PUSTAKA.....	73
LAMPIRAN .....	76

## DAFTAR GAMBAR

Nomor Urut	Halaman
1. Contoh perbandingan algoritma DBSCAN & K-Means .....	5
2. Flowchart Alur Kerja K-Means Clustering .....	8
3. Flowchart Alur Kerja DBSCAN Clustering.....	9
4. Contoh teknik PCA (Wei. L, Tang. W, Huang. H, dan Chen. T, 2021) .....	12
5. Indeks silhouette penentuan klaster optimal .....	13
6. Lokasi penelitian (Source : Google Maps) .....	18
7. Lokasi pengambilan data pada (Source : Google Maps) .....	18
8. Tampak depan lokasi pengambilan data (Source : Google Maps).....	19
9. Diagram alur kerja.....	21
10. Gambaran umum sistem penelitian .....	25
11. Mentahan dataset sebelum di gabung. ....	29
12. Gambaran salah satu dataset sekolah sebelum digabung dari kota Makassar .	30
13. data setelah digabung menjadi 1 dan diberi nama data_gabungan_sekolah_sulsel.csv .....	31
14. Dataset sebelum di bersihkan.....	32
15. Dataset setelah dilakukan pembersihan .....	32
16. Pengimportan library yang di butuhkan dan menampilkan dataset Yang telah di bersihkan sebelumnya .....	33
17. Normaliasi data dengan StandardScaler .....	33
18. Hasil dari normalisasi data.....	33
19. Mengonversi data array ke DataFrame.....	34

20. Download data array menjadi data_normalisasi_sekolah_sulsel.csv.....	34
21. Import library dan pemanggilan data.....	35
22. Pengimplementasian teknik PCA.....	35
23. Hasil analisis menggunakan teknik PCA.....	36
24. Import library dan pemanggilan data PCA .....	36
25. Mendeklarasikan features dan nilai x1 .....	37
26. Penggunaan algoritma Nearest Neighbors .....	37
27. Grafik visualisasi <i>Silhouette Score</i> .....	39
28. DBSCAN Clustering .....	39
29. Klaster total .....	39
30. Membuat DataFrame dbscan_kluster .....	40
31. Total data yang dianggap sebagai noise/outlier.....	40
32. Penyederhanaan data noise/outlier .....	41
33. Hasil uji validasi <i>Silhouette Score</i> DBSCAN .....	41
34. Penggabungan data hasil DBSCAN dengan PCA dan data awal .....	42
35. Grafik jumlah klaster optimal.....	43
36. Import library dan KMeans dan pemanggilan dataset data_dari_dbSCAN.csv ....	44
37. Penentuan kolom untuk peningkatan DBSCAN .....	44
38. Pengimplementasian algoritma K-Means pada peningkatan DBSCAN .....	45
39. DataFrame dari hasil K-Means Clustering pada peningkatan DBSCAN .....	45
40. Grafik hasil K-Means Clustering pada peningkatan DBSCAN .....	46
41. Penggabungan dataset data_dari_dbSCAN.csv dengan DataFrame klaster serta uji validasi n_cluster = 3.....	46
42. Pengimplementasian algoritma K-Means Clustering dalam peningkatan DBSCAN dan hasil DataFrame pada n_cluster = 4 .....	47

43. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n_cluster = 4.....	47
44. Pengimplementasian algoritma K-Means Clustering dalam peningkatan DBSCAN dan hasil DataFrame pada n_cluster = 2 .....	47
45. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n_cluster = 2.....	48
46. Hasil uji validasi <i>Silhouette Score</i> n_cluster = 2.....	48
47. Hasil uji validasi <i>Silhouette Score</i> n_cluster = 4.....	48
48. Import library,pemanggilan data, dan penentuan variabel K-Means Only.....	50
49. Implementasi K-Means Clustering pada K-Means Only.....	50
50. Pembuatan DataFrame hasil K-Means Only.....	51
51. Grafik hasil K-Means only dengan n_cluster = 3.....	51
52. Penggabungan dataset data_pca_sekolah_sulsel dan DataFrame klaster_kmeans_only .....	52
53. Pengimplementasian algoritma K-Means Clustering pada K-Means Only dan hasil DataFrame pada n_cluster = 2 .....	52
54. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n_cluster = 2.....	53
55. Pengimplementasian algoritma K-Means Clustering pada K-Means Only dan hasil DataFrame pada n_cluster = 4 .....	53
56. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n_cluster = 4.....	53
57. Hasil uji validasi <i>Silhouette Score</i> n_cluster = 2.....	54
58. Hasil uji validasi <i>Silhouette Score</i> n_cluster = 4.....	54
59. Grafik perbandingan hasil terbaik K-Means Only dan peningkatan DBSCAN ...	55
60. Perbandingan tabel hasil uji validasi <i>Silhouette Score</i> DBSCAN Only,peningkatan DBSCAN, dan K-Means Only.....	56
61. Import library dan pemanggilan dataset data_hasil_peningkatan_dbSCAN_3cluster .....	57

62. Bobot perangkingan .....	58
63. Normalisasi data menggunakan Min-Max Scaling .....	58
64. Menjumlahkan hasil perkalian total performa sekolah dan normalisasi ulang agar rentangnya menjadi 0 hingga 100.....	59
65. Pengurutan dan penentuan peringkat dengan skor tertinggi .....	59
66. Penggabungan data skor akhir dan kolom asli serta menampilkan 10 data teratas hasil perangkingan dari model peningkatan DBSCAN 3 klaster .....	60
67. Menyimpan dataset dengan nama  Sekolah_Rekomendasi_Bantuan_BOP_PeningkatanDSBCAN3Cluster.csv.....	61
68. Tampilan dataset hasil rekomendasi sekolah untuk bantuan BOP dan Grafik (Peningkatan DBSCAN 3 Cluster) .....	61
69. Import library dan pemanggilan dataset kmeans_only_2cluster.csv .....	62
70. Bobot Perangkingan .....	63
71. Normalisasi Data dan Pengurutan .....	63
72. Menampilkan 10 Data Teratas Hasil Perangkingan .....	64
73. Menyimpan dataset dengan nama  Sekolah_Rekomendasi_Bantuan_BOP_KMeansOnly2Cluster.csv .....	64
74. Tampilan dataset hasil rekomendasi sekolah untuk bantuan BOP dan Grafik (K-Means Only 2 klaster) .....	65
75. Perbandingan hasil pembobotan K-Means Only (2 klaster) & Peningkatan DBSCAN (3 klaster) .....	67
76. Pembuatan peta interaktif menggunakan pustaka Folium .....	67
77. Memasukkan marker sekolah ke peta .....	68
78. Fitur pencarian sekolah dan pencarian peringkat sekolah .....	68
79. Menampilkan fitur filter berbentuk layer dengan 4 layer yaitu SD,SMP,SMA,	

dan SMK .....	69
80. Tampilan hasil peta interaktif menggunakan pustaka Folium .....	69
81. Heatmap korelasi antara indikator dan skor rangking rekomendasi.....	70

**DAFTAR TABEL**

Nomor Urut	Halaman
1. Pembobotan kriteria BOP berdasarkan kriteria bantuan BOP dari Dinas Pendidikan Sulawesi Selatan.....	15
2. Tabel total jumlah sekolah sesuai jenjang .....	22
3. Parameter pengujian.....	26
4. Tabel nilai epsilon, min_samples, dan <i>Silhouette Score</i> .....	38
5. Hasil uji validasi peningkatan DBSCAN .....	48
6. Hasil uji validasi K-Means Only .....	54

**DAFTAR SINGKATAN DAN ARTI SIMBOL**

---

Lambang/Singkatan	Arti dan Keterangan
PCA	Principal Component Analysis
S(i)	<i>Silhouette Index</i>
a(i)	Rata-rata jarak antara i dan semua objek lain dalam klaster sendiri
b(i)	Jarak minimum rata-rata antara i dan semua klaster lain
KNN	K-Nearest Neighbor
HTML	Hypertext Markup Language
csv	Comma Separated Value
k	<i>Klaster</i>
DBSCAN	Density-Based Spatial Clustering of Applications with Noise

---

**DAFTAR LAMPIRAN**

Nomor Urut	Halaman
1. Surat permohonan pengambilan data 5 sekolah yang belum masuk di dataset ..	76
2. <i>Source code</i> program .....	77
3. Source Code keseluruhan .....	94
4. Gambaran dataset sari Dinas Pendidikan Sulawesi Selatan .....	94
5. Riwayat Hidup.....	95

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Pendidikan merupakan salah satu dari empat pilar dasar pembangunan suatu negara. Hingga saat ini, kualitas pendidikan yang merata di seluruh daerah masih menjadi jalur utama penyediaan kesempatan belajar bagi seluruh warga negara. Indonesia, seperti sejumlah negara lain, masih menghadapi tugas berat untuk mencapai pemerataan pendidikan yang masih menjadi fokus utama upaya pembangunan. Sulawesi Selatan merupakan salah satu provinsi di Indonesia yang Kompleksitas di antara provinsi-provinsi ini disebabkan oleh berbagai masalah termasuk pendidikan karena Sulawesi Selatan termasuk dalam lima provinsi teratas dengan jumlah sekolah terbanyak. Menurut Harian Fajar sekaligus wawancara langsung ke Dinas Pendidikan SULSEL, praktik di lapangan tidak banyak berubah. Dana BOP masih didistribusikan secara bergilir ke setiap sekolah tiap Kota meskipun seharusnya ada peraturan yang mengatur bahwa BOP didistribusikan secara berbeda untuk memenuhi kebutuhan setiap sekolah berdasarkan beberapa kriteria. Untuk menghasilkan pola, wawasan, dan segmentasi peserta didik atau lingkungan akademik, seseorang dapat menggunakan teknik pengelompokan dengan data pendidikan (Adhikari, A. dan Sen, S. 2023). Provinsi Sulawesi Selatan memiliki banyak daerah dengan karakteristik sosial, ekonomi, dan geografis yang berbeda-beda. Meskipun telah dilakukan berbagai upaya untuk meningkatkan akses dan kualitas pendidikan di daerah ini, kesenjangan antara wilayah perkotaan dan pedesaan serta antarwilayah masih menjadi kendala utama.

Dengan demikian, BOP merupakan bantuan keuangan yang diberikan pemerintah kepada sekolah untuk digunakan dalam pengelolaan operasional peningkatan mutu pendidikan di sekolah. Faktor-faktor tersebut dapat memengaruhi efektivitas dan efisiensi penyaluran BOP yang didasarkan pada pemahaman yang lebih mendalam tentang karakteristik sekolah di wilayah tersebut. Jadi, yang menjadi perhatian bukan hanya alokasinya, tetapi seberapa efisien dan efektif alokasi tersebut.

Dalam konteks Sulawesi Selatan yang merupakan wilayah Dinas Pendidikan Daerah, daerah ini memiliki tanggung jawab yang sangat besar untuk memastikan penyaluran BOP merata dan tepat sasaran. Agar semua itu dapat terlaksana, maka Dinas Pendidikan Daerah harus terlebih dahulu melakukan pemetaan yang lebih mendalam terhadap sebaran sekolah di daerahnya. Dengan demikian, ke depannya petugas akan lebih mudah menentukan titik-titik dukung mana saja yang harus disalurkan. Oleh karena itu, untuk memberikan kontribusi yang maksimal dalam meningkatkan efektivitas penyaluran BOP, penelitian ini menggunakan teknik clustering atau pengelompokan data untuk mengungkap pola yang terdapat pada data sekolah di Sulawesi Selatan. Dua algoritma clustering yang diperbandingkan dalam penelitian ini adalah K-Means dan DBSCAN. Dengan

membandingkan kedua algoritma tersebut, maka Dinas Pendidikan akan lebih memahami bagaimana mengoptimalkan penyaluran BOP dengan lebih baik.

Proses penambangan data dilakukan secara manual dalam mengekstraksi nilai tambah berupa informasi yang sebelumnya tidak diketahui dari suatu basis data. Penambangan data telah berkembang pesat sejak tahun 1990-an sebagai cara yang sangat akurat dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antar data sehingga dapat dikelompokkan menjadi salah satu dari beberapa kluster dimana objek dalam satu kluster akan memiliki tingkat kesamaan yang tinggi satu sama lain. Dengan demikian penambangan data dianggap sebagai proses penemuan pengetahuan dari basis data Knowledge Discovery in Databases

Clustering adalah pengelompokan data yang termasuk dalam beberapa klaster atau kelompok, sehingga data dalam klaster memiliki tingkat homogenitas tertinggi, dan data antar klaster memiliki tingkat kesamaan terendah. Homogeneous clustering adalah menemukan beberapa klaster objek yang homogen dari sekumpulan data. Sasaran utama metode ini adalah mengelompokkan banyak kumpulan data/objek dalam klaster (kelompok) yang berusaha membuat setiap klaster memiliki data yang serupa; proses ini mengatakan bahwa algoritma clustering memeriksa seluruh kumpulan data dan menjadikannya menjadi subkelompok, di mana homogenitas data dalam klaster dimaksimalkan dan heterogenitas yang terletak di luar klaster diminimalkan. Proses pengelompokan data sering menggunakan dua jenis utama clustering data: clustering data hierarkis dan non-hierarkis.

Algoritma K-Means sebenarnya diposisikan sebagai salah satu algoritma terbesar yang akan diaplikasikan di bidang data mining. K-Means membagi data, mengelompokkannya ke dalam beberapa cluster yang memiliki kesamaan, dan memisahkan setiap cluster berdasarkan perbedaan antar cluster. Algoritma ini telah diusulkan oleh lebih kurang sekitar peneliti di berbagai bidang. Ini adalah metode yang termasuk dalam algoritma clustering berbasis jarak yang membagi data ke dalam sejumlah cluster tertentu dan hanya bekerja dengan atribut numerik (Tandy, J., & Assegaff, S., 2019). Keuntungan utama pengelompokan adalah pola dan struktur tersembunyi dapat dideteksi dengan menganalisis kumpulan data besar dengan sedikit atau tanpa pengetahuan latar belakang. K-Means adalah teknik yang banyak digunakan untuk memprediksi dan pengelompokan di masa depan (Alawi, S. J. S., Shaharanee, I. N. M., & Jamil, J. M., 2023).

DBSCAN merupakan salah satu algoritma pengelompokan berbasis kepadatan paling awal yang muncul selama tahun 1990-an. Ada tiga karakteristik utama DBSCAN. Pertama, ia menemukan klaster yang memiliki bentuk dan ukuran yang berbeda. Kedua, ia mengidentifikasi outlier dengan benar. Ketiga, ia tidak memerlukan jumlah klaster yang ditentukan sebelumnya. Jumlah klaster ditentukan oleh algoritma sesuai dengan karakteristik data input (Sutramiani et al., 2024). Prinsip utama DBSCAN didasarkan pada konsep kepadatan data, dan ini menjadikannya bentuk pengembangan paling awal dari teknik pengelompokan

berbasis kepadatan. Tidak seperti pengelompokan k-means, DBSCAN tidak memerlukan penentuan jumlah klaster sebelumnya. Hanya dua hal yang perlu ditentukan, yaitu jumlah tetangga minimum untuk membentuk grup dan jarak epsilon hingga titik mana yang harus dianggap sebagai grup yang sama. Hal ini meningkatkan keuntungan karena biasanya bentuk dan jumlah kelompok yang harus diberikan pada data berdimensi tinggi tidak dapat diketahui melalui analisis visual data. DBSCAN juga dapat mengenali noise dengan baik tanpa perlu mendapatkan input awal berupa jumlah cluster ( $k$ ) seperti k-Means atau k-Medoids, dan dapat mengenali bentuk cluster yang tidak teratur.

Dengan uraian permasalahan di atas, maka dapat dikatakan bahwa pada penelitian ini akan menggunakan metode K-means Clustering dan peningkatan algoritma DBSCAN Clustering dalam membentuk daerah-daerah di Provinsi Sulawesi Selatan, yang dalam hal ini sekolah-sekolah digabung. Kemudian, akan memasukkan sekolah-sekolah tersebut ke dalam beberapa klaster kualitas dan kuantitas secara langsung untuk selanjutnya memberikan alternatif solusi dalam mendukung pemberian Bantuan Operasional Pendidikan (BOP) yang lebih merata dan bertarget di Provinsi Sulawesi Selatan. Dengan demikian, hasil analisis tersebut dapat bermanfaat untuk lebih mendukung kebijakan yang bertujuan untuk pemerataan pendidikan di wilayah Sulawesi Selatan dan sekaligus memperkuat basis data Dinas Pendidikan Sulawesi Selatan.

## 1.2 Teori

### 1.2.1 Clustering

Dalam analisis data, pengelompokan sebagai pengelompokan sekumpulan objek atau data berdasarkan kesamaan atau kedekatannya satu sama lain cukup populer. Data yang termasuk dalam kelompok atau klaster yang sama memiliki atribut atau properti yang sama, sedangkan data dalam klaster yang berbeda memiliki atribut yang bervariasi. Pelatihan ini ditujukan untuk mengekstraksi pola atau struktur yang melekat dalam data yang dapat dieksplorasi lebih lanjut, baik secara deskriptif maupun preskriptif (Jain, 2020). Karena tidak memerlukan pelabelan atau kelas respons apa pun dalam kumpulan data selama analisis, clustering ini termasuk dalam kategori pembelajaran tanpa pengawasan (Unsupervised Learning). Ada banyak algoritma yang digunakan untuk pengelompokan, dan tidak semua algoritma harus bekerja dengan cara yang sama. Berikut ini adalah beberapa algoritma yang paling populer:

1. **Pengelompokan K-Means:** K-Means merupakan salah satu algoritma pengelompokan yang paling populer dan mudah. Metode ini membagi data ke dalam sejumlah kluster tertentu, atau kluster " $k$ ". Setiap objek ditetapkan ke dalam kluster tergantung pada jarak rata-ratanya dari semua titik kluster, atau centroid. Proses ini berlanjut hingga titik pembaruan atau pergerakan centroid. Salah satu keuntungan menjalankan algoritma K-Means adalah kemampuannya untuk menangani big data dengan mudah dan cepat;

meskipun demikian, algoritma ini memiliki sejumlah kluster yang perlu ditentukan sebelumnya oleh seseorang, dan pilihan titik centroid awal dapat memengaruhi hasil (Lloyd, 2021).

2. **Pengelompokan Hirarkis:** Algoritma pengelompokan hierarkis menghasilkan struktur pohon yang dikenal sebagai dendrogram, yang secara fundamental berbeda dari K-Means karena tidak memerlukan pengaturan jumlah klaster sebelumnya. Algoritme ini memiliki dua versi populer yang disebut aglomeratif, yang bekerja secara top-down, dan divisif, yang juga bekerja secara top-down. Aglomeratif dimulai dengan satu klaster besar dan memecahnya lebih lanjut menjadi klaster yang lebih kecil. Meskipun cukup fleksibel, pendekatan ini bisa jadi sangat tidak efektif jika diterapkan pada kumpulan data yang lebih besar karena intensitas komputasinya seperti yang diamati oleh Jain et al. (2019).
3. **DBSCAN** : Density-Based Spatial Clustering of Applications with Noise, sebuah algoritma pengelompokan yang mencoba mengelompokkan objek berdasarkan seberapa rapat mereka. Tidak seperti K-means, metode ini tidak memerlukan jumlah kluster yang harus ditentukan sebelumnya. Metode ini mendefinisikan kluster berdasarkan emanasi titik ke dalam ruang. Hal ini bahkan memungkinkannya untuk menemukan kluster dengan bentuk yang berubah-ubah. Salah satu masalahnya adalah parameter ini sulit dipilih. Sebagian besar parameter tersebut terkait dengan epsilon dan jumlah titik minimum yang diperlukan untuk membentuk kluster (Ester et al., 2020).
4. Dikenal sebagai **Fuzzy C-Means (FCM)**, yang merupakan pendekatan modifikasi dari K-Means yang mempertimbangkan bahwa setiap data memiliki tingkat keanggotaan di semua klaster. Dalam hal ini, tidak seperti K-Means, suatu objek diperbolehkan untuk menjadi bagian dari lebih dari satu klaster dengan tingkat keanggotaan tertentu. Namun, FCM sangat bagus untuk data tidak murni yang tidak sepenuhnya menjadi bagian dari satu klaster tetapi membutuhkan waktu.

Clustering juga sangat bermanfaat dalam beberapa bidang. Clustering memiliki banyak aplikasi, aplikasi utamanya meliputi:

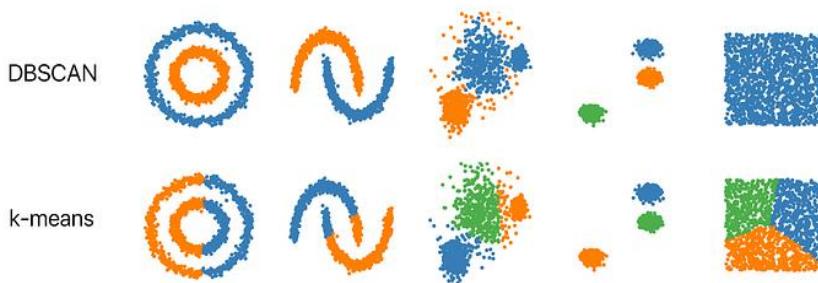
1. **Dalam bidang Pendidikan** : pengelompokan juga dapat digunakan untuk meningkatkan efisiensi penyaluran dana BOP di bidang pendidikan. Sekolah dapat dikelompokkan secara luas ke dalam dua kategori: satu kelompok memiliki buku teks yang diadopsi kurikulum, sementara yang lain tidak. "Kelompok mana yang perlu ditingkatkan infrastrukturnya?", "dengan prestasi akademik yang rendah?", atau "jumlah siswa yang banyak?", adalah pertanyaan yang harus diajukan. Ini akan membantu menentukan prioritas pendanaan BOP secara lebih tepat.
2. **Biologi dan Genetika**: Dalam biologi dan genetika, clustering mencakup pengelompokan spesies atau gen berdasarkan fitur genetik atau fenotipik

tertentu. Misalnya, pengelompokan mikroorganisme atau organisme laut berdasarkan urutan DNA-nya. Pengelompokan juga diterapkan untuk mengelompokkan penyakit berdasarkan pola ekspresi gen (Zhu et al., 2022).

3. Clustering juga digunakan dalam **analisis sosial** sebagai metode untuk memahami pola interaksi dalam jaringan sosial dengan mengelompokkan individu atau kelompok berdasarkan interaksi sosial mereka. Pengelompokan memberi kita wawasan tentang dinamika antar individu serta untuk menemukan komunitas atau kelompok yang memiliki dampak besar pada suatu jaringan. Teknik ini terutama digunakan dalam analisis komunikasi dan media sosial (Hassan et al., 2022).
4. **Pengenalan Pola Clustering:** teknik ini akan sangat membantu sistem Anda menghasilkan pola tulisan tangan dan gambar wajah karena akan membantu sistem mengelompokkan fitur visual yang serupa di antara objek untuk meningkatkan kemanjuran sistem yang terlibat dalam pengenalan tulisan tangan atau wajah. Sistem biometrik dan keamanan juga menggunakan (Wang & Xie, 2023).

Adapun beberapa metrik yang umum digunakan untuk menilai kualitas hasil pengelompokan adalah sebagai berikut:

1. **Skor Silhouette:** Nilai seberapa baik objek telah ditempatkan di klaster yang benar dibandingkan dengan klaster lain, dengan kemungkinan nilai bervariasi dari -1 hingga maksimum +1. Nilai yang mendekati +1 menunjukkan bahwa objek tersebut terklaster dengan baik.
2. **Indeks Davies-Bouldin (DBI):** Kesamaan dalam klaster dibandingkan dengan klaster terdekatnya. Nilai yang lebih rendah dari kuantitas ini lebih baik dan menunjukkan pengelompokan yang lebih baik.
3. **Indeks Dunn:** Jika Anda menetapkan jarak minimum antar klaster ke jarak terjauh, rasionalnya akan menjadi indeks ini. Semakin tinggi nilainya, semakin baik pengelompokannya.



Gambar 1. Contoh perbandingan algoritma DBSCAN & K-Means

Saat ini, clustering terus berkembang seiring kemajuan teknologi yang disertai dengan meningkatnya kebutuhan akan analisis data yang lebih rumit. Kemajuan terkini dalam clustering meliputi hal-hal berikut:

1. **Penggunaan dalam Big Data dan Cloud Computing:** Pengelompokan telah banyak digunakan dalam pemrosesan big data melalui cloud computing karena volume data terus meningkat. Algoritme pengelompokan skala besar, seperti Mini-Batch K-Means dan DBSCAN yang diparalelkan, diimplementasikan dalam Apache Spark yang lebih efisien dalam memproses big data. Kerangka kerja lain yang tersedia akan membutuhkan lebih banyak waktu.
2. **Pengelompokan Berbasis Pembelajaran Mendalam:** Pendekatan baru yang menggabungkan pengelompokan dan pembelajaran mendalam, hanya saja lebih cocok untuk data yang kompleks dan besar; pendekatan ini menggunakan Pengelompokan Tertanam Mendalam (DEC) dan Variational Autoencoders (VAE) agar bekerja lebih baik dengan data gambar dan teks. Teknik ini memberikan hasil yang lebih baik dalam pengelompokan data gambar dan teks.
3. **Pendekatan Baru untuk Pengelompokan Data Non-Euclidean:** Data teks dan grafik memerlukan pendekatan baru karena keduanya memiliki hubungan non-Euclidean yang, memang, memerlukan pendekatan baru untuk pengelompokan. Dua teknik yang semakin populer untuk jenis data ini adalah pengelompokan berbasis grafik dan pendekatan berdasarkan Alokasi Dirichlet Laten (LDA).
4. **Tantangan Pengelompokan Data Berdimensi Tinggi:** Masalah utama dan lama yang sangat memengaruhi pengelompokan data berdimensi tinggi adalah "kutukan dimensionalitas." Untuk membuat proses pengelompokan lebih mudah dikelola, pengurangan dimensionalitas dapat diterapkan dengan menggunakan algoritme seperti t-SNE dan PCA. Algoritme yang dimodifikasi juga kini menjadi pusat perhatian, seperti DBSCAN yang dimodifikasi.
5. **Penanganan Noise dan Outlier:** Untuk mengurangi dampak outlier pada hasil pengelompokan, algoritma yang dapat digunakan adalah Robust K-Means dan Adaptive DBSCAN, yang akan membuat pengelompokan lebih mampu menangani data dengan noise dan banyak outlier.
6. **Aplikasi Industri dan Bisnis:** Analisis bisnis, khususnya analisis perilaku konsumen dan segmentasi pasar area di mana pengelompokan semakin populer. Selain itu, pengelompokan cukup berguna dalam perawatan kesehatan untuk analisis data genomik serta pengobatan berbasis pola.
7. **Interpretabilitas dan Visualisasi:** Hasil pengelompokan lunak dan keras atau penugasan sembarang pada data yang sangat kompleks dapat memperoleh representasi yang sangat mudah dipahami yang tersedia bagi

mata manusia melalui t-SNE, UMAP, dan sejenisnya. Ini juga menemukan intuisi yang lebih baik untuk memahami hasil pengelompokan.

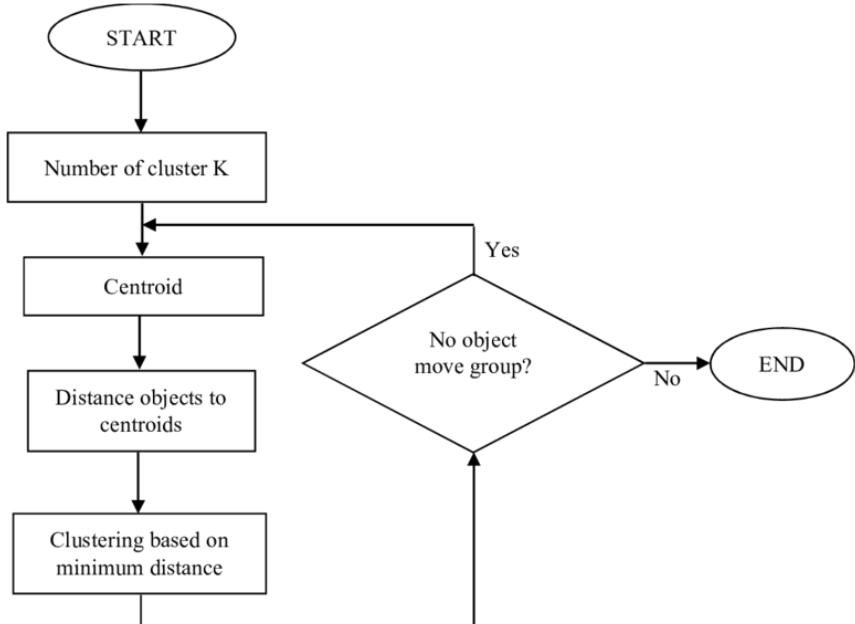
### 1.2.2 K-MEANS Clustering

K-Means adalah algoritma pembelajaran tanpa pengawasan yang digunakan untuk mengelompokkan data dalam beberapa kelompok berdasarkan kesamaan yang dimiliki oleh data. Sasaran utama yang dikejar oleh algoritma ini adalah mencoba untuk menjaga semua objek dalam klaster tertentu sedekat mungkin satu sama lain dan membuatnya sebisa mungkin berbeda dari objek dari klaster lain. Biasanya, algoritma ini bekerja dengan data yang ekstensif atau besar, dengan asumsi bahwa setiap klaster memiliki dispersi yang sama dan melingkar. (Kumar & Sharma, 2021).

Algoritma K-Means pertama-tama dimulai dengan nilai yang telah ditentukan sebelumnya dari jumlah klaster  $K$ . Kemudian,  $K$  titik acak dalam ruang dipilih sebagai pusat (disebut centroid dalam konteks K-Means) pengelompokan. Kemudian, setiap titik dalam kumpulan data ditetapkan ke klaster yang centroid-nya paling dekat dengan klaster tersebut melalui jarak atau ukuran kesamaan tertentu. Setelah itu, centroid ditentukan ulang sebagai posisi rata-rata data dalam klaster dan proses ini berlanjut hingga posisi centroid tidak banyak berubah (Lee & Park, 2022). K-means yang relatif sederhana akan menghadapi masalah. Masalah utamanya adalah menemukan jumlah pengguna  $K$  yang sering mencoba berbagai nilai  $K$  sendiri untuk menemukan angka optimal. Masalah lain adalah titik pusat awal memengaruhi hasil akhir; variasi K-means yang lebih efektif telah dikembangkan untuk menangani data yang tidak terstruktur dan kompleks, namun, untuk mengatasi masalah ini (Gupta & Desai, 2023).

K-Means memiliki kecepatan komputasi yang sangat menguntungkan yang harus diterapkan dengan tepat untuk menghindari terjebak dalam solusi lokal yang tidak ideal. Untuk beberapa pilihan centroid awal yang ceroboh, K-Means tidak akan dapat digunakan (Wang & Yang, 2024).

Beberapa aplikasi K-Means adalah pengelompokan gambar, segmentasi pasar, dan analisis pola perilaku konsumen. Dalam bioinformatika, K-Means digunakan untuk mengelompokkan data genetik atau organisme berdasarkan kesamaan karakteristik genetik. Pengelompokan pengguna berdasarkan preferensi mereka dalam sistem rekomendasi di mana bisnis kemudian dapat menawarkan saran yang lebih bermanfaat juga merupakan sebuah aplikasi (Kumar & Sharma, 2021).



Gambar 2. Flowchart Alur Kerja K-Means Clustering

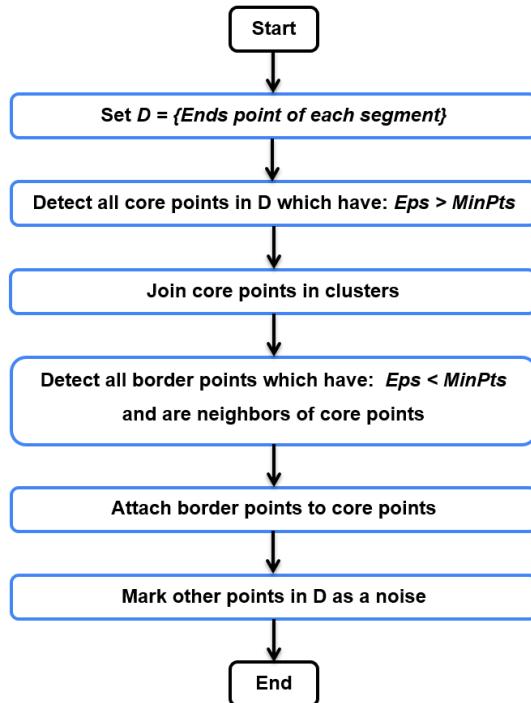
### 1.2.3 DBSCAN dan Prinsip Kerjanya

DBSCAN merupakan algoritma pengelompokan berbasis kepadatan yang mengelompokkan data ke dalam kelompok bagian ruang data yang memiliki titik yang relatif banyak ada keuntungan berbasis data dibandingkan algoritma seperti K-Means karena tidak memerlukan K yang ditentukan pengguna dan lebih tangguh terhadap distribusi data yang tidak terstruktur, ditambah lagi akan mencari data yang bukan bagian dari suatu kelompok sebagai gangguan/outlier (Sharma & Raj, 2023).

eps dan MinPts adalah dua parameter dasar yang digunakan oleh DBSCAN. MinPts berarti jumlah titik minimum yang harus ada dalam jarak eps untuk mempertimbangkan pembentukan klaster. Sementara itu, eps adalah jarak maksimum hingga titik data dapat dianggap sebagai tetangga. Titik tersebut akan berada di bawah set tetangga, berdasarkan jumlah titik yang tersedia dalam radius yang dipertimbangkan. Jika tidak, titik yang tidak termasuk dalam kriteria akan dianggap sebagai noise atau outlier. Untuk menjalankan algoritma pengelompokan, mulailah mengambil titik secara acak dan putuskan titik ini apakah itu titik inti atau bukan atas nama kedekatan titik itu dengan titik-titik lain yang menonjol. Pada dasarnya, titik inti dalam klaster memiliki setidaknya tetangga eps dalam jarak MinPts. Titik yang terhubung langsung ke titik inti dan berada dalam rentang eps juga termasuk dalam klaster yang sama. Proses berlanjut hingga semua titik telah diproses. Titik yang tidak memenuhi syarat sebagai titik inti atau dekat dengan titik inti adalah outlier atau noise. Tidak seperti K-Means, yang terbatas pada bentuk klaster bulat, DBSCAN dapat mendeteksi klaster dengan bentuk yang berubah-

ubah. Oleh karena itu, DBSCAN merupakan pilihan yang baik untuk kumpulan data yang tidak memiliki distribusi seragam atau bentuk yang mendekati bulat. Selain itu, DBSCAN dapat mendeteksi gangguan. Ini merupakan keuntungan penting, terutama saat digunakan untuk menganalisis kumpulan data yang memiliki banyak outlier.

Keunggulan utama DBSCAN mencakup kapasitasnya untuk pengelompokan bentuk heterogen dan agak tidak teratur serta kapasitas untuk menangani gangguan dalam data. DBSCAN tidak memerlukan jumlah input klaster dan dengan demikian mengurangi kemungkinan menggunakan nilai K yang tidak tepat atau salah. Dari keunggulan tersebut DBSCAN diterapkan dalam berbagai aplikasi, mulai dari pengelompokan spasial dalam studi geografis hingga analisis data gambar dalam visi komputer dan juga analisis jaringan sosial. Penggunaan bioinformatika melibatkan pengelompokan data genetik berdasarkan kesamaan yang berkaitan dengan pola ekspresi gen dan penemuan outlier dalam kumpulan data yang sangat besar.



Gambar 3. Flowchart Alur Kerja DBSCAN Clustering

#### 1.2.4 Data Mining

Penambangan data dikenal sebagai proses mengekstraksi informasi yang sebelumnya tidak diketahui dari sejumlah besar data yang beragam dan kompleks. Informasi ini kemudian dapat digunakan untuk pengambilan keputusan. Penambangan data menstandardisasi apa yang sering kali merupakan proses berlapis-lapis yang mencakup pembersihan, pemilihan, transformasi, pemodelan,

dan akhirnya menemukan data. Tujuan utamanya adalah menemukan informasi yang dapat diterjemahkan ke dalam pengetahuan yang berguna untuk mendukung pengambilan keputusan atau prediksi masa depan. Penambangan data sebagai proses yang digerakkan oleh pencarian tertentu termasuk pengelompokan, klasifikasi, regresi, atau asosiasi untuk mengekstraksi informasi implisit yang sebelumnya tidak diketahui dari sejumlah besar data (Han et al., 2021).

Agar menghasilkan hasil yang akurat dan bermanfaat, proses penambangan data harus dipecah menjadi tugas-tugas yang terdefinisi dengan baik untuk dijalankan dalam urutan tertentu. Tugas-tugas ini terdiri dari hal-hal berikut:

1. Pemilihan Data: Pemilihan data dipandang sebagai tugas pertama dalam pemrosesan data. Data yang kuat untuk tujuan atau masalah yang sedang dipertimbangkan akan lebih bijaksana untuk ditafsirkan sebagai relevan (Indahsari et al., 2023).
2. Pembersihan Data: Hapus data duplikat, tangani data yang hilang, dan perbaiki setiap ketidakkonsistenan yang terdeteksi. Karena data biasanya kotor. Ini adalah rutinitas awal yang penting karena tanpanya analisis tersebut dapat menyebabkan hasil yang tidak akurat dan tidak informatif.
3. Transformasi Data: Setelah pembersihan data, ada tahap berikutnya, yang biasanya disebut transformasi data. Di sini, data akan diformat sedemikian rupa sehingga algoritme analisis dapat menggunakannya. Ini dapat melibatkan atau tidak melibatkan normalisasi data, pengkodean data kategoris, atau penggabungan beberapa bentuk data yang terurai menjadi bentuk yang lebih terstruktur.
4. Proses Penambangan Data: Pada tahap ini, berbagai metode akan dipertimbangkan yang dapat menemukan pola dan menarik informasi berharga dari data. Mungkin ada berbagai algoritma dengan tipe yang berbeda; misalnya, algoritma klasifikasi untuk membuat prediksi atau algoritma pengelompokan untuk mengelompokkan data berdasarkan seberapa miripnya. DBSCAN, Apriori, dan K-Means adalah beberapa algoritma yang sering diterapkan (Tarigan et al., 2022).

**Interpretasi dan Evaluasi:** Setelah pola ditemukan, langkah selanjutnya dan terakhir adalah menginterpretasikan hasil dan mengevaluasi nilainya. Kemudian, temuan ini dimasukkan ke dalam masukan keputusan dalam penelitian dan aplikasi bisnis. Pada titik ini, sangat penting bahwa hasilnya benar-benar relevan dan dapat diterapkan. Selain itu ada beberapa teknik utama yang biasanya digunakan, berikut ini juga beberapa teknik utama yang digunakan dalam penambangan data:

1. Klasifikasi : Metode ini digunakan untuk mengklasifikasikan data ke dalam kategori yang telah ditentukan sebelumnya, misalnya, klasifikasi akan digunakan untuk mengklasifikasikan transaksi di sektor keuangan sebagai penipuan atau sah.
2. Pengelompokan : Pengelompokan diterapkan untuk mengelompokkan data ke dalam kelompok homogen berdasarkan karakteristik. Pengelompokan

dapat diterapkan dalam pemasaran untuk mengidentifikasi berbagai segmen pasar.

3. Regresi : akan menjadi pilihan yang lebih baik untuk memprediksi nilai kontinu berdasarkan variabel input. Misalnya, harga properti dapat diprediksi berdasarkan lokasi dan ukuran properti.
4. Asosiasi: Prosedur pembuatan aturan ini dimasukkan untuk menyoroti pola atau hubungan antara item dalam data. Misalnya, alasan pengecer menggunakan asosiasi adalah mencari tahu daftar produk yang paling sering dibeli sebagai satu item.

Peningkatan terkini dalam penambangan data difokuskan pada penskalaan ke big data dan penggunaan algoritma pembelajaran mendalam untuk analisis data yang lebih canggih. Teknologi komputasi awan dan algoritma pembelajaran mesin yang lebih baik kini memungkinkan penambangan data dilakukan dengan mempertimbangkan volume data yang lebih baik dan konteks yang lebih kompleks

### **1.2.5 Unsupervised Learning**

Unsupervised learning adalah metode pembelajaran mesin untuk membantu model menemukan pola dan struktur dalam data tanpa label apa pun yang telah ditetapkan sebelumnya. Itulah perbedaannya dengan Supervised Learning; Unsupervised Learning bekerja secara independen dan tidak untuk memaksimalkan kriteria informasi yang jelas mengenai kategori data. Dengan kata lain, pembelajaran tanpa pengawasan dapat diterapkan dalam praktik karena data yang diberi label tidak mudah ditemukan.

Clustering dan Association merupakan dua metodologi utama Unsupervised Learning. K-Means sejauh ini merupakan salah satu algoritma paling populer untuk mengelompokkan data berdasarkan kesamaan (Rias et al., 2023). Dalam konteks tersebut, teknik Association digunakan untuk mengetahui bagaimana analisis keranjang belanja dapat diterapkan untuk menentukan hubungan yang ada antara variabel dalam suatu kumpulan data (Roihan et al., 2020).

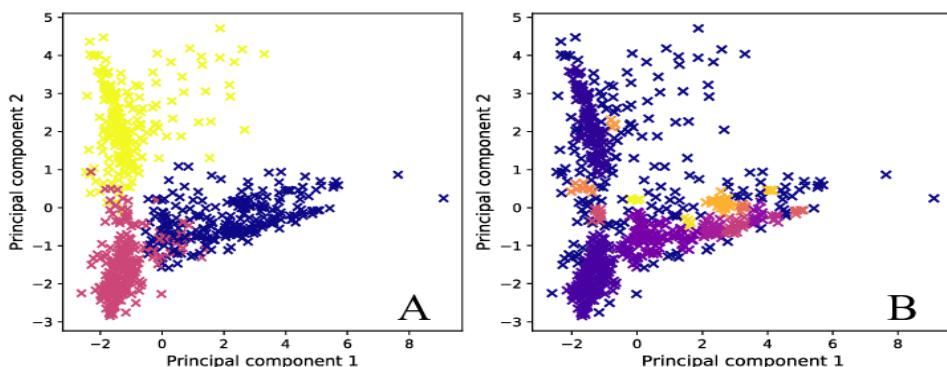
Salah satu keuntungan utama pembelajaran tanpa pengawasan adalah kemampuannya untuk bekerja dengan data yang tidak berlabel. Namun, metode ini memiliki beberapa masalah, termasuk interpretasi hasil yang tepat yang sering kali subjektif dan bergantung pada orang yang melakukan penelitian di domain tersebut. Lebih jauh, implementasi dalam praktik sering kali menghadapi kompleksitas algoritma dan kemungkinan munculnya pola yang tidak relevan, yang dapat mengurangi efektivitas model dalam aplikasi dunia nyata (Fahmi, 2023).

### **1.2.6 Principal Component Analysis (PCA)**

Salah satu metode statistik, Principal Component Analysis (PCA), dapat digunakan untuk mempertahankan sebanyak mungkin informasi yang terkandung dalam data yang memiliki banyak variabel. Mengubah sejumlah variabel yang saling berkorelasi menjadi sejumlah variabel baru yang tidak berkorelasi adalah tujuan

utama PCA. Dengan mengurangi jumlah variabel, proses ini memungkinkan penyederhanaan kompleksitas data. Tujuan PCA adalah "menemukan arah (komponen utama) yang memaksimalkan varians data dan mengurangi dimensi tanpa kehilangan informasi yang terlalu banyak" (Jolliffe & Cadima, 2019). PCA digunakan dalam banyak bidang, seperti pengajaran mesin, pengolahan gambar, dan analisis data besar.

PCA pada dasarnya dimulai dengan menghitung matriks korelasi, atau kovarians, antar variabel dalam kumpulan. Kemudian, eigenvektor dan eigenvalue dihitung dari matriks tersebut. Eigenvektor menunjukkan arah dimensi baru dalam data, sementara eigenvalue menunjukkan kontribusi setiap komponen terhadap variasi data. Komponen utama yang dihasilkan dari PCA adalah kombinasi linier dari variabel asli dengan variansi terbesar.



Gambar 4. Contoh teknik PCA dari perbandingan algoritma K-Means & DBSCAN  
(Wei. L, Tang. W, Huang. H, dan Chen. T, 2021)

Standardisasi data biasanya dilakukan saat proses PCA dimulai, terutama ketika variabel yang digunakan memiliki satuan yang berbeda. Ini penting untuk memastikan bahwa setiap variabel memiliki berat yang sama dalam analisis. Selanjutnya, matriks kovarians dihitung untuk mengetahui hubungan antar variabel. Selanjutnya, eigenvektor dan eigenvalue dihitung untuk mendapatkan bagian penting yang akan digunakan untuk transformasi data. Menurut literatur, PCA meningkatkan efisiensi analisis data dan mengurangi dimensi data. Ini terutama berlaku untuk kasus data besar (Zhang & Zhang , 2022). Hal ini sangat bermanfaat untuk mempermudah visualisasi data yang kompleks dan menghindari masalah overfitting dalam model statistik dan machine learning.

### 1.2.7 Indeks Silhouette

Dalam analisis data, *indeks Silhouette* adalah alat yang digunakan untuk menilai kualitas hasil klasterisasi. Indeks Silhouette, yang diperkenalkan oleh Rousseeuw pada tahun 1987, berfungsi untuk mengukur seberapa mirip suatu objek dengan klasternya sendiri dibandingkan dengan klaster lain. Nilai yang lebih besar dari -1 menunjukkan bahwa objek tersebut terkelompok dengan baik, sedangkan nilai

yang lebih rendah menunjukkan bahwa objek mungkin lebih cocok berada di klaster lain. Secara matematis, untuk setiap objek  $i$ , jarak rata-rata antara  $i$  dan semua objek lain dalam klaster yang sama dihitung; jarak rata-rata ini juga dihitung untuk setiap klaster lain, dan nilai terkecil dari rata-rata jarak ini disebut  $b(i)$ . Selanjutnya, rumus berikut digunakan untuk menghitung indeks *Silhouette* objek  $i$ :

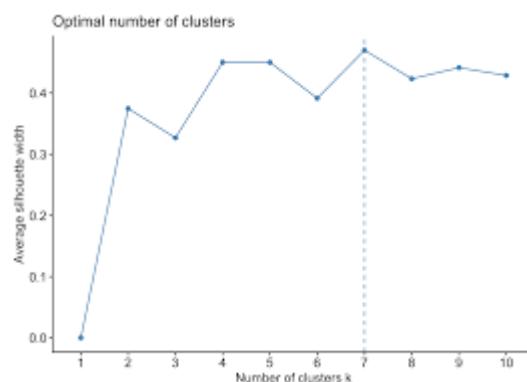
$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Di mana:

- $a(i)$  adalah jarak rata-rata antara  $i$  dan semua objek dalam klasternya sendiri.
- $b(i)$  adalah jarak minimum rata-rata antara  $i$  dan semua objek dalam klaster lain.

Nilai  $S(i)$  yang lebih besar dari 1 menunjukkan bahwa objek berada dalam klaster yang tepat, sementara nilai yang lebih rendah dari -1 menunjukkan bahwa objek mungkin lebih cocok di klaster lain. Nilai yang lebih rendah dari 0 menunjukkan bahwa objek berada di antara dua klaster. Dengan menggunakan indeks *Silhouette* untuk menentukan jumlah klaster yang ideal, peneliti dapat mengevaluasi kualitas keseluruhan dari skema klasterisasi yang digunakan dengan menghitung nilai rata-rata untuk setiap objek dalam dataset. Nilai rata-rata yang lebih tinggi menunjukkan klasterisasi yang lebih baik.

Dalam beberapa tahun terakhir, Indeks *Silhouette* telah digunakan secara luas dalam berbagai penelitian untuk mengevaluasi hasil klasterisasi. Hasil klasterisasi yang dihasilkan oleh algoritma K-Means Enhanced dan K-Means Maximum Minimum Criterion dievaluasi dengan Indeks *Silhouette*; penelitian ini menekankan pentingnya penggunaan *Indeks Silhouette* untuk menentukan jumlah klaster yang optimal dan menilai kualitas klasterisasi yang dihasilkan (Khairati et al., 2019).



Gambar 5. Indeks silhouette penentuan klaster optimal

### 1.2.8 Python

Python adalah bahasa pemrograman tingkat tinggi yang berfokus pada sintaks yang bersih dan keterbacaan kode. Python pyton di buat pertama kali pada akhir tahun 1980-an oleh Guido van Rossum, tetapi sekarang telah menjadi salah satu bahasa pemrograman paling populer di dunia karena sintaksisnya yang cukup sederhana dan mudah dipahami yang dapat digunakan untuk menulis kode yang jelas dan sangat mudah dibaca oleh para pengembang. Selain itu, Python mendukung berbagai paradigma pemrograman berorientasi objek, fungsi, dan input imperatif, sehingga pengembang dapat membangun berbagai jenis aplikasi dengan konstruksi yang lebih adaptif. Selama beberapa tahun terakhir, bahasa ini telah mendapatkan pengakuan sebagai salah satu bahasa utama untuk aplikasi mulai dari pengembangan web hingga analisis data, kecerdasan buatan, dan pembelajaran mesin karena mendukung setiap pengembang dengan pustaka standar yang besar dan koleksi pustaka pihak ketiga yang lebih besar yang digerakkan oleh komunitas. Contoh pustaka ini termasuk NumPy dan Pandas yang sangat populer untuk manipulasi data, serta TensorFlow dan PyTorch untuk model pembelajaran mesin, dan masih banyak lagi. Python mendapatkan popularitas terutama karena sangat ramah bagi pendatang baru di dunia pemrograman. Tanpa kesederhanaan dasar, dukungan komunitas yang luas, sintaksis yang sederhana, dan fleksibilitas dalam berbagai domain, Python tidak akan pernah masuk dalam daftar bahasa yang ideal untuk pemula (Cutting dan Stephen, 2021). Fitur-fitur ini menjadikan Python sebagai bahasa pilihan baik bagi pemula maupun profesional berpengalaman yang ingin mengembangkan solusi perangkat lunak dengan cepat.

### 1.2.9 BOP Provinsi (Bantuan Operasional Pendidikan)

Untuk mendukung proses pembelajaran yang efektif dan efisien, Kementerian Pendidikan melalui Bantuan Operasional Pendidikan (BOP) memberikan dukungan dana operasional nonpersonalia sekolah. Bantuan Operasional Sekolah ini disalurkan melalui 2 tahap yaitu 31 Januari dan 31 juli setiap tahunnya dan tidak semua sekolah yang ada di SULSEL mendapatkan bantuan ini, berdasarkan data tahun lalu yang dijelaskan langsung oleh beberapa Staff bidang PTK dan Fasilitas DISDIK SULSEL bahwa total ada 627 sekolah yang menerima bantuan di Sulawesi Selatan. Namun, 627 sekolah yang menerima bantuan tahun 2024 dan Dinas Pendidikan SULSEL rekomendasikan ke pusat berasal dari kota Makassar karena metode yang mereka gunakan saat ini yaitu dengan penyaluran bergilir tiap kota padahal di peraturan BOP meyarankan untuk memberikan bantuan ke sekolah yang benar-benar membutuhkan berdasarkan kriteria yang telah ditentukan sehingga metode yang digunakan kurang efektif jika menyalurkannya bergilir ke setiap kota yang ada di SULSEL tiap tahunnya. Dengan adanya penelitian ini sangat bermanfaat untuk memberikan rekomendasi sekolah yang memang sangat membutuhkan bantuan berdasarkan kriteria akreditasi, total siswa, jumlah ruangan, total wilayah, daya listrik, jumlah guru, dan jumlah staf. Kriteria BOP dari pusat meliputi skor akreditasi, jumlah siswa, jumlah ruangan, dan total wilayah.

Berdasarkan permintaan data penilaian tambahan dari Dinas Pendidikan Sulawesi Selatan untuk kriteria BOP meliputi daya listrik (watt), jumlah guru, dan jumlah staff dengan bobot perangkingan pada tabel 1.

Tabel 1. Pembobotan kriteria BOP berdasarkan kriteria bantuan BOP dari Dinas Pendidikan Sulawesi Selatan

Urutan <b>Prioritas</b>	Variabel Kriteria BOP	Jenis Kriteria	Bobot
1	Skor Akreditasi	Kemendikbudristek	29
2	Total Siswa	Kemendikbudristek	23
3	Jumlah Ruangan	Kemendikbudristek	19
4	Total Wilayah (m <sup>2</sup> )	Kemendikbudristek	14
5	Daya Listrik (Watt)	DISDIK SULSEL	5
6	Jumlah Guru	DISDIK SULSEL	5
7	Jumlah Staff	DISDIK SULSEL	5
<b>Total</b>			100

Program ini bertujuan untuk meningkatkan kualitas pendidikan secara merata di berbagai tingkat pemerintahan, termasuk tingkat provinsi. Barlian et al. (2023) meneliti dan mengkaji kebijakan Bantuan Operasional Pendidikan Daerah (BOPD) yang digunakan Pemerintah Provinsi Jawa Barat, dalam upayanya meningkatkan mutu pendidikan di SMA Negeri 1 Purwakarta. Kajian mereka menitikberatkan pada implementasi daerah di provinsi tersebut. Hasil kajian menunjukkan bahwa implementasi BOPD secara signifikan meningkatkan mutu pendidikan, khususnya pada aspek sarana dan prasarana yang memadai serta guru dan murid yang semakin meningkat performanya. Oleh karena itu, disarankan agar pemangku kepentingan lebih berkolaborasi. Singkatnya, implementasi BOP di tingkat provinsi sangat mendukung fungsi pendidikan dan peningkatan mutu. Apa yang seharusnya dilakukan secara terus-menerus adalah evaluasi dan perbaikan proses implementasi untuk menghilangkan berbagai hambatan.

### 1.3 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah yang timbul adalah berikut :

1. Bagaimana penerapan metode K-means Clustering dengan algoritma peningkatan DBSCAN Clustering dalam pengelompokan sekolah di Sulawesi Selatan?
2. Apakah penerapan metode K-means Clustering dengan algoritma peningkatan DBSCAN Clustering dapat bekerja dengan baik dan benar pada penelitian ini?

3. Apakah terdapat perbedaan signifikan antara hasil pengelompokan sekolah menggunakan algoritma K-Means dan peningkatan DBSCAN dalam konteks penyaluran BOP di Sulawesi Selatan?
4. Bagaimana performa algoritma K-Means dan peningkatan DBSCAN dalam mengelompokkan data sekolah di Sulawesi Selatan untuk tujuan penyaluran Bantuan Operasional Pendidikan (BOP) yang merata?

#### **1.4 Tujuan Penelitian**

Penelitian ini bertujuan untuk:

1. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik kepada Dinas Pendidikan Sulawesi Selatan tentang pola-pola yang ada dalam data sekolah di wilayah tersebut.
2. Dengan pemahaman yang lebih baik tentang pola distribusi sekolah, Dinas Pendidikan dapat mengoptimalkan penyaluran BOP mereka agar lebih merata dan tepat sasaran.
3. Penerapan metode K-means Clustering dan Peningkatan DBSCAN Clustering berjalan dengan baik dan benar.
4. Membandingkan hasil metode K-means Clustering dan pengembangan metode K-means Clustering dengan Peningkatan DBSCAN Clustering dalam pengelompokan sekolah di Sulawesi Selatan.

#### **1.5 Manfaat Penelitian**

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Bagi Penulis, Penulis dapat lebih mendalami dan memahami ilmu tentang data mining dan data science terkhusus pada metode clustering.
2. Bagi Universitas, Mengetahui kemampuan mahasiswa dalam menerapkan akademis maupun non-akademis di lingkungan masyarakat dan ilmu Lembaga, serta sebagai arsip literasi kampus.
3. Bagi Dinas Pendidikan Sulawesi Selatan, Penelitian ini dapat membantu Dinas Pendidikan Sulawesi Selatan dalam pengelompokan sekolah yang ada di Sulawesi Selatan yang jumlahnya sangat banyak dan belum pernah di lakukan pengelompokan sekolah daerah Sulawesi Selatan sebelumnya, khususnya untuk penyaluran BOP. Hasil Penelitian ini dapat memberikan alternatif solusi untuk mendukung penyaluran Bantuan Operasional Pendidikan (BOP) yang lebih merata dan tepat sasaran di provinsi Sulawesi Selatan. Selain itu hasil dari analisis ini dapat memberikan kontribusi positif dalam mendukung kebijakan pemerataan pendidikan di Sulawesi Selatan serta memperkuat basis data Dinas Pendidikan Sulawesi Selatan.

## 1.6 Ruang Lingkup

Penelitian ini memiliki beberapa batasan, yaitu:

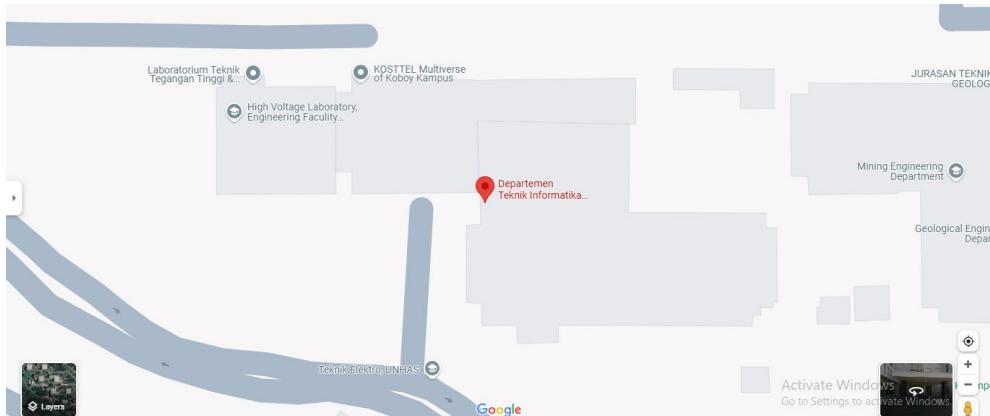
1. Menggunakan Dataset Seluruh Sekolah SD,SMP,SMA,dan SMK swasta dan negri
2. Menggunakan algoritma DBSCAN untuk penentuan jumlah klaster dan mendeteksi data outlier/noise.
3. Menggunakan Teknik Analisis Komponen Utama (PCA).
4. Menggunakan algoritma K-means Only dan Peningkatan DBSCAN untuk pengelompokan data dan membandingkan hasilnya untuk mencari pemodelan yang baik dan akurat.
5. Menggunakan algoritma Nearest Neighbors untuk menentukan parameter epsilon metode peningkatan DBSCAN.
6. Menggunakan Tools Jupyter Notebook pada Anaconda untuk seleksi fitur dan implementasi metode dengan bahasa python.
7. Melakukan pembobotan sesuai kriteria BOP dari Dinas Pendidikan Sulawesi Selatan dari hasil pemodelan algoritma yang terbaik untuk penentuan sekolah yang paling membutuhkan rekomendasi.
8. Validasi dengan menggunakan *Silhouette Score*.

## BAB II

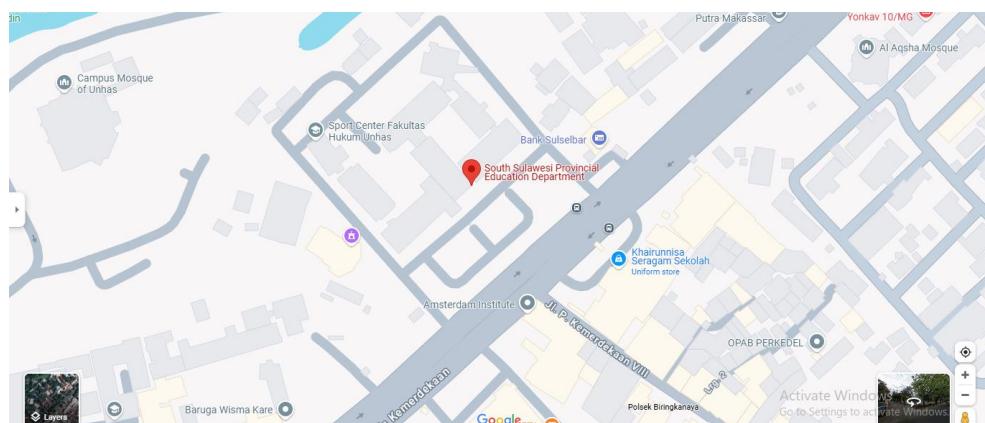
### METODE PENELITIAN

#### 2.1 Lokasi dan Waktu Penelitian

Lokasi pelaksanaan tugas akhir di lakukan di Laboratorium Ubquitos Computing & Networking Departement Teknik Informatika Fakultas Teknik, Universitas Hasanuddin serta penelitian ini dilakukan pada juli 2024 hingga 02 Maret 2025. Sedangkan lokasi pengambilan data di lakukan di Dinas Pendidikan Sulawesi Selatan.



Gambar 6. Lokasi penelitian pada Laboratorium Ubquitos Computing & Networking Departement Teknik Informatika, Universitas Hasanuddin (Source : Google Maps)



Gambar 7. Lokasi pengambilan data pada Dinas Pendidikan Sulawesi Selatan yang berlokasi di KM 10, Jl. Perintis Kemerdekaan, Tamalanrea Indah, Kec. Tamalanrea, Kota Makassar, Sulawesi Selatan (Source : Google Maps)



Gambar 8. Tampak depan lokasi pengambilan data Dinas Pendidikan Sulawesi Selatan yang berlokasi di KM 10, Jl. Perintis Kemerdekaan, Tamalanrea Indah, Kec. Tamalanrea, Kota Makassar, Sulawesi Selatan (Source : Google Maps)

## 2.2 Instrumen Penelitian

Pada penelitian ini digunakan dua jenis instrumen yaitu instrumen perangkat lunak dan instrumen perangkat keras. Berikut rincian mengenai setiap instrumen penelitian yang digunakan.

### 2.3.1 Instrumen perangkat lunak

Perangkat lunak yang digunakan pada penelitian ini adalah sebagai berikut :

1. Operation System windows 10, os yang digunakan untuk menjalankan python.
2. Jupyter Notebook, code editor/tools yang digunakan untuk melakukan analisis data.
3. Bahasa Python, Sebagai bahasa yang di gunakan dalam analisis data.
4. MS. Excel untuk mengubah data mentah menjadi suatu dataset
5. SKlearn,pandas,matplotlib,Folium,dan NumPy yaitu library yang digunakan untuk mendukung proses analisis data.

### 2.3.2 Instrumen perangkat keras

Perangkat keras yang digunakan selama penelitian adalah sebagai berikut :

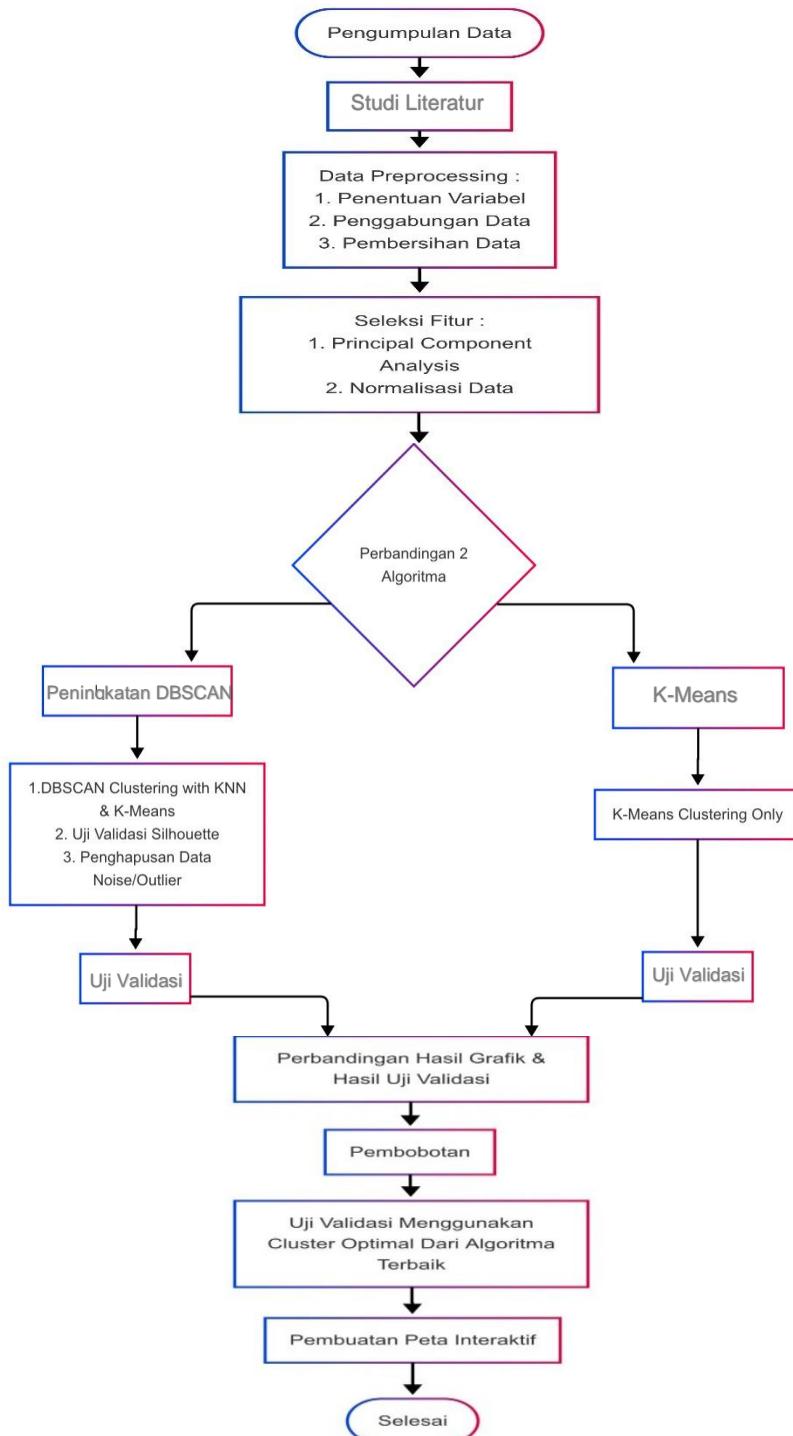
1. Hardware berupa laptop DELL VOSTRO 14" 5468, Intel Core i7-7500U CPU @ 2.70GHz 2.90 GHz, 8 GB DDR4 RAM, 64-bit operating system, Windows 10.

## 2.3 Objek Penulisan

Objek dalam penulisan ini adalah data Sekolah SD,SMP,dan SMA/SMK Negeri ataupun swasta periode 2023/2024 di Sulawesi Selatan yang terdata di Dinas Pendidikan Sulawesi Selatan dan Kementerian Pendidikan & Kebudayaan dengan

total 9201 sekolah. Selain itu data yang diperoleh dalam penelitian ini adalah semua sekolah SD,SMP,SMA,dan SMK baik swasta maupun negeri berdasarkan 21 Kabupaten, 2 Kota, dan 1 Kepulauan yang ada di Sulawesi Selatan Parameter data dari sekolah yang digunakan meliputi, luas wilayah (wilayah\_total), akreditasi (akreditasi\_total), daya listrik (daya\_listrik), jumlah ruangan (ruangan\_total), jumlah staf (staf\_total), jumlah guru (teacher\_total), dan jumlah siswa (siswa\_total), Lintang (lintang), dan Bujur (bujur). Data lintang & bujur berdasarkan hasil scraping dari google maps untuk pembentukan peta interaktif

## 2.4 Diagram Alur Kerja



Gambar 9. Diagram alur kerja

#### **2.4.1 Karakteristik data**

Mengumpulkan data yang digunakan dalam penulisan ini, yaitu data statistik dan asset sekolah periode 2023/2024 di Sulawesi Selatan yang bersumber dari Dinas Pendidikan Sulawesi Selatan. Data yang berhasil di kumpulkan dari Dinas Pendidikan Sulawesi Selatan yaitu data dari seluruh sekolah SD,SMP,SMA,dan SMK yang ada di Sulawesi Selatan baik itu Swasta maupun Negeri. Dengan total 24 dataset berdasarkan 21 Kabupaten, 2 Kota, dan 1 Kepulauan yang ada di Sulawesi Selatan. Total keseluruhan data tersebut memiliki 17 kolom, 9197 baris, dan total isi data sebanyak 119.560. Untuk Kec. Wajo, Makassar dibutuhkan kunjungan langsung untuk mengambil data dari 5 sekolah (4 SMA & 1 SMK) karena data untuk Kec.Wajo, Kota Makassar tidak ada didalam dataset yang diberikan oleh dinas Pendidikan. Selain itu penambahan 2 kolom kedalam dataset berdasarkan hasil scraping dari google maps untuk pembentukan peta interaktif. Jika digabung datanya memiliki 19 kolom, 9202 baris, dan total isi data sebanyak 119.605.

Tabel 2. Tabel total jumlah sekolah sesuai jenjang

No	Jenjang	Jumlah
1	SD (Sekolah Dasar)	6470 (6076 Negeri & 394 Swasta)
2	SMP (Sekolah Menengah Pertama)	1716 (1256 Negeri & 460 Swasta)
3	SMA (Sekolah Menengah Atas)	601 (329 Negeri & 272 Swasta)
4	SMK (Sekolah Menengah Kejuruan)	414 (163 Negeri & 251 Swasta)
	<b>TOTAL</b>	<b>9201 (7824 Negeri &amp; 1377 Swasta)</b>

#### **2.4.2 Studi Literatur**

Studi literatur dilakukan dengan mengkaji jurnal, e-book, artikel ilmiah, serta sumber dari situs internet yang berkaitan dengan algoritma DBSCAN, K-Means, KNN, Partitioning Method, Metode Berbasis Hirarki (*Hierarchical Method*), Metode Berbasis Kepadatan (Density-Based Method), Metode Berbasis Kisi (Grid-Based Method), dan Menggunakan Teknik Analisis Komponen Utama (PCA).

#### **2.4.3 Data Preprocessing**

Melakukan data preprocessing untuk menyiapkan dataset dengan mengubah data mentah menjadi data yang layak untuk diproses. Berikut adalah tahapan yang dilakukan dalam data preprocessing :

##### **1. Penentuan Variabel**

Tahap memilih dan menentukan variabel yang digunakan sebagai parameter dalam pembentukan klaster. Penentuan variabel ini berdasarkan kelengkapan data dan kebutuhan pengelompokan data berdasarkan kuantitas dan kualitas yang masuk kedalam kriteria peraturan pembagian BOP.

##### **2. Penggabungan Data**

Menggabungkan data yang berasal dari dataset yang berbeda.

### 3. Pembersihan Data

Melakukan pembersihan data dengan menseleksi yang diterima. Hal ini penting dilakukan agar tidak ada *missing values*.

#### **2.4.4 Seleksi Fitur**

Seleksi fitur berfungsi untuk mengurangi dimensi data dan fitur-fitur yang sudah tidak relevan. Tools yang akan digunakan adalah Jupyter Notebook. Berikut adalah tahapan yang dilakukan dalam seleksi fitur :

1. Normalisasi Data (Feature Scaling)

Mengubah data yang dimiliki agar memiliki rentang nilai yang sama. Menyamakan rentang nilai ini penting dilakukan agar tidak ada nilai yang terlalu besar atau terlalu kecil agar membuat proses analisis lebih mudah.

2. Teknik Analisis Komponen Utama (PCA)

Mereduksi variabel yang ada didalam dataset yang dimiliki. Dengan mengubah dimensi data yang terlalu besar menjadi dua dimensi dengan menggunakan metode Teknik Analisis Komponen Utama (PCA). Hal ini dilakukan untuk mempermudah proses clustering dan memaksimalkan metode algoritma yang akan digunakan.

#### **2.4.5 Seleksi Clustering**

Metode DBSCAN disini untuk menghilangkan data noise/outlier dan menentukan jumlah klaster yang paling optimal. Metode DBSCAN yang digunakan akan dibantu oleh Algoritma Nearest Neighbor (KNN) untuk menentukan nilai epsilon atau nilai jarak maksimum yang optimal. Ada 4 tahapan yaitu :

1. Melakukan clustering menggunakan metode DBSCAN dengan K-Nearest Neighbors (KNN).
2. Melakukan uji validasi menggunakan *Silhouette Score*.
3. Menghapus data noise atau outlier yang terdeteksi.
4. Melakukan uji validasi kembali setelah pembersihan data.

#### **2.4.6 K-Means Only**

Klasterisasi dilakukan dengan menggunakan metode K- means Clustering tanpa menggunakan metode DBSCAN Clustering. Klasterisasi dilakukan dengan menggunakan data hasil pre processing.

#### **2.4.7 Uji Validasi K-Means Only**

Uji validasi dilakukan pada klasterisasi dengan menggunakan metode K-means Clustering. Uji validasi dilakukan dengan menggunakan *Silhouette Score* pada hasil klasterisasi K-means Only.

#### **2.4.8 Perbandingan Hasil Uji Validasi**

Membandingkan hasil uji dari setiap algoritma sehingga akan menghasilkan algoritma terbaik dan efisien pada penelitian ini. Serta Menyimpulkan metode clustering yang memberikan hasil terbaik berdasarkan validasi dan visualisasi.

#### **2.4.9 Pembobotan**

Melakukan pembobotan setelah mengetahui hasil algoritma terbaik dan cluster terbaik untuk menentukan urutan sekolah yang akan mendapatkan bantuan BOP.

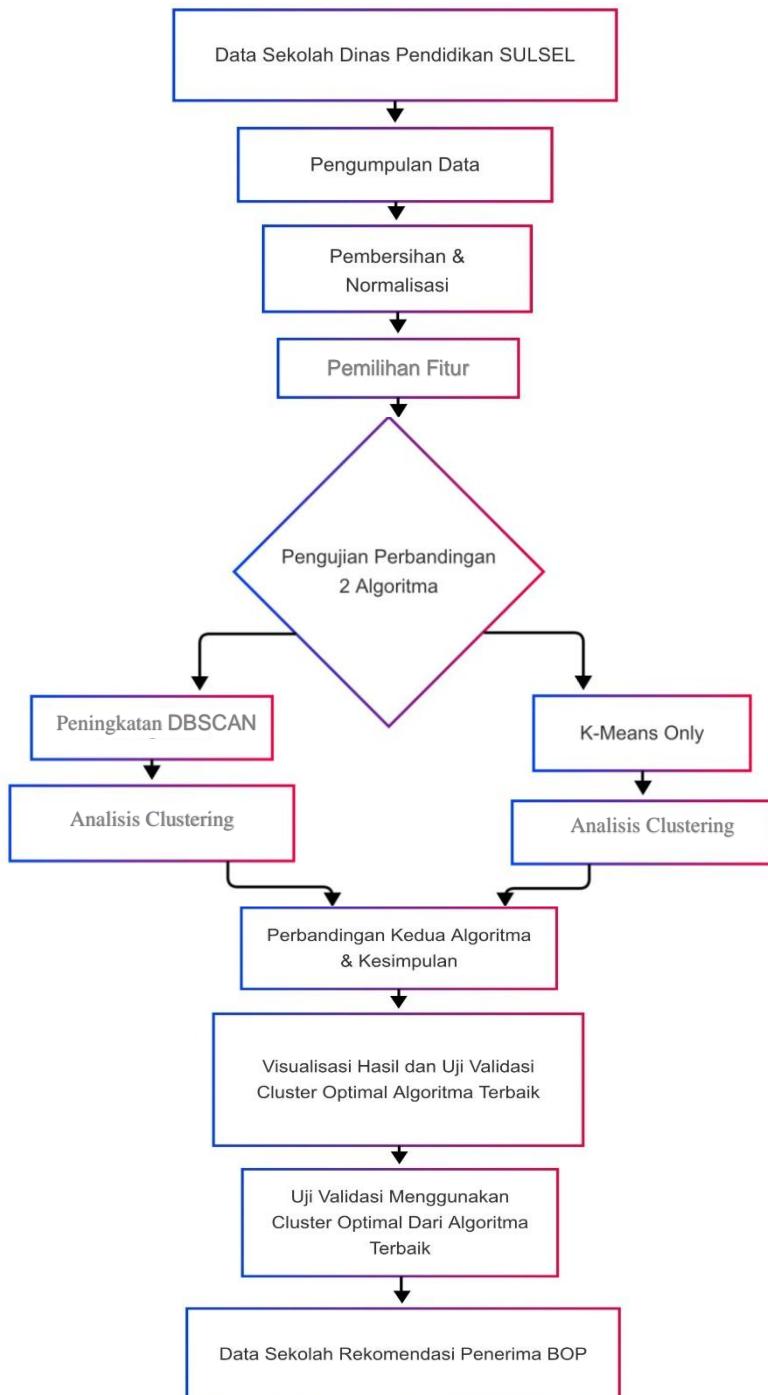
#### **2.4.10 Uji Validasi Menggunakan Cluster Optimal Dari Algoritma Terbaik**

Melakukan uji validasi dari cluster optimal algoritma terbaik untuk mengetahui sekolah yang masuk di cluster 1, cluster 2, atau cluster 3, dan seterusnya. Cluster tersebut nantinya menunjukkan penyebaran data sekolah/pengelompokan data sekolah yang sangat membutuhkan BOP, tidak terlalu membutuhkan, dan tidak membutuhkan BOP.

#### **2.4.11 Pembuatan Peta Interaktif**

Membuat peta interaktif berdasarkan hasil uji validasi menggunakan cluster optimal dari algoritma terbaik dengan warna sebagai penunjuk sekolah tersebut berada di kategori sangat membutuhkan BOP, tidak terlalu membutuhkan, dan tidak membutuhkan BOP.

## 2.5 Gambaran Umum Sistem



Gambar 10. Gambaran umum sistem penelitian

Sistem yang dikembangkan dalam penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma K-Means dan peningkatan DBSCAN dalam melakukan clustering data sekolah di Sulawesi Selatan. Clustering ini bertujuan untuk mengelompokkan sekolah berdasarkan berbagai faktor seperti luas wilayah (wilayah\_total), akreditasi (akreditasi\_total), daya listrik (daya\_listrik), jumlah ruangan (ruangan\_total), jumlah staf (staf\_total), jumlah guru (teacher\_total), dan jumlah siswa (siswa\_total). Dengan demikian, hasil clustering dapat digunakan sebagai dasar dalam merancang strategi penyaluran BOP yang lebih merata dan tepat sasaran oleh Dinas Pendidikan Sulawesi Selatan.

Dengan mengumpulkan data sekolah dari sumber resmi seperti Dinas Pendidikan Sulawesi Selatan, data yang diperoleh akan melalui tahap preprocessing, termasuk pembersihan data, normalisasi, dan pemilihan fitur yang relevan. Setelah itu, dua algoritma clustering, yaitu K-Means dan peningkatan DBSCAN, akan diterapkan untuk mengelompokkan sekolah berdasarkan pola yang ditemukan dalam dataset.

Untuk mengukur performa kedua algoritma, metrik evaluasi yang akan digunakan adalah *Silhouette Score*. Hasil dari analisis ini akan membantu menentukan algoritma mana yang lebih efektif dalam mengelompokkan sekolah secara optimal, sehingga dapat mendukung pengambilan keputusan dalam pendistribusian BOP secara lebih adil dan efisien. Selain itu dari penelitian ini akan divisualisasikan dalam bentuk peta clustering sekolah yang menunjukkan pola distribusi berdasarkan kategori yang telah ditentukan. Dengan adanya sistem ini, diharapkan Dinas Pendidikan Sulawesi Selatan dapat memiliki dasar ilmiah dalam menentukan kebijakan yang lebih baik terkait pendanaan pendidikan di wilayah tersebut.

### 2.5.1 Perbandingan Hasil Uji Validasi

Proses analisis dimulai dengan pengumpulan data sekolah yang berasal dari Dinas Pendidikan Sulawesi Selatan. Data yang digunakan mencakup luas wilayah (wilayah\_total), akreditasi (skor\_akreditasi), daya listrik (daya\_listrik), jumlah ruangan (ruangan\_total), jumlah staf (staf\_total), jumlah guru (jumlah\_guru), jumlah siswa (siswa\_total), Lintang (lintang), dan Bujur (bujur). Setelah data terkumpul, dilakukan preprocessing untuk memastikan kelayakan analisis, termasuk proses pembersihan data dari duplikasi atau nilai yang hilang, normalisasi agar setiap variabel berada dalam skala yang sama, serta transformasi variabel kategori menjadi bentuk numerik jika diperlukan.

Tabel 3. Parameter pengujian

No	Algoritma	Parameter	Pengukuran Evaluasi
1	K-Means	K (Jumlah klaster)	<i>Silhouette Score</i>
2	DBSCAN	Epsilon, MinPts	<i>Silhouette Score</i>
3	Perbandingan	Waktu eksekusi	Analisis runtime dan <i>Silhouette Score</i>

Penentuan variabel dipenelitian ini berdasarkan kriteria untuk penyaluran BOP yang dijelaskan oleh Dinas Pendidikan Sulawesi Selatan yaitu kriteria paling utama dengan pembobotan akreditasi mempunyai nilai penilaian yang tinggi, selanjutnya di ikuti dengan total siswa, jumlah ruangan, total wilayah ( $m^2$ ), lalu dengan nilai penilaian yang sama besar yaitu jumlah guru,daya listrik, dan jumlah staf. Akreditasi, total siswa, jumlah ruangan, dan total wilayah berdasarkan kriteria dari pusat sebagai kriteria utama sedangkan jumlah guru, daya listrik, dan jumlah staf sebagai kriteria tambahan untuk mendukung kriteria utama untuk diajukan ke Kemendikbud. Variabel lainnya seperti lintang dan bujur digunakan untuk pembuatan peta interaktif.

Proses pengujian dilakukan melalui tiga eksperimen utama. Pertama, eksperimen variasi jumlah klaster pada K-Means untuk menentukan nilai K yang paling optimal berdasarkan evaluasi *Silhouette Score*. Kedua, eksperimen variasi parameter epsilon dan MinPts pada DBSCAN guna mengidentifikasi konfigurasi terbaik dalam membentuk klaster yang berkualitas. Terakhir, eksperimen perbandingan antara kedua algoritma berdasarkan hasil clustering yang dihasilkan, baik dari segi kualitas maupun efisiensi waktu eksekusi.

Dari hasil analisis, penelitian ini akan menyajikan perbandingan yang jelas mengenai kelebihan dan kekurangan masing-masing algoritma dalam konteks distribusi rekomendasi sekolah di Sulawesi Selatan. Jika K-Means menunjukkan performa lebih baik dalam pembentukan klaster yang terstruktur dengan distribusi yang relatif seimbang, maka algoritma ini dapat menjadi pilihan utama untuk kebijakan penyaluran BOP. Namun, jika peningkatan DBSCAN lebih efektif dalam menangani variasi kepadatan data dan mengelompokkan sekolah dengan distribusi yang lebih heterogen, maka metode ini lebih direkomendasikan.

## **BAB III**

### **HASIL DAN PEMBAHASAN**

Hasil dari penelitian ini akan mencari performa dan akurasi terbaik dari 2 algoritma K-Means dan peningkatan DBSCAN untuk pengelompokan sekolah yang membutuhkan rekomendasi bantuan BOP. Adapun dataset yang digunakan berasal dari Dinas Pendidikan Sulawesi Selatan serta hasil scraping dari google maps.

#### **3.1 Pengumpulan Data**

Pengumpulan data dilakukan dengan mengajukan surat penelitian dari Universitas Hasanuddin, Fakultas Teknik, Departement Teknik Informatika kepada pihak Dinas Pendidikan Sulawesi Selatan. Data yang dibutuhkan adalah data seluruh sekolah SD,SMP,SMA,dan SMK yang ada di Sulawesi Selatan baik itu Swasta maupun Negeri. Pada tahap ini ada total 24 dataset berdasarkan 21 Kabupaten, 2 Kota, dan 1 Kepulauan yang ada di Sulawesi Selatan.

Setelah melakukan wawancara dengan pihak Dinas Pendidikan Sulawesi Selatan terkait pembentukan kelompok rekomendasi sekolah untuk bantuan BOP di provinsi Sulawesi Selatan. Dalam wawancara ini menghasilkan tujuan, yaitu membentuk kelompok pada seluruh sekolah SD,SMP,SMA,dan SMK di Sulawesi Selatan dengan hanya membentuk kelompok yang membutuhkan rekomendasi bantuan karena dijelaskan bahwa proses penyalurannya saat ini bergiliran tiap Kota dengan cara memilih kota tersebut lalu membagikan ke sekolah yang ada di Kota tersebut padahal setiap tahun anggaran harus mempertimbangkan seluruh sekolah yang ada di Sulawesi Selatan untuk penyalurannya. Pengelompokan sekolah-sekolah ini dikelompokan berdasarkan aset yang dimiliki oleh sekolah tersebut. Aset sekolah yang ditentukan adalah luas wilayah (wilayah\_total), akreditas (skor\_akreditasi), daya listrik (daya\_listrik), jumlah ruangan (ruangan\_total), jumlah staf (staf\_total), jumlah guru (jumlah\_guru), jumlah siswa (siswa\_total), serta Lintang (lintang) dan Bujur (bujur) digunakan untuk pembuatan peta interaktif.

#### **3.2 Data Preprocessing**

Data yang digunakan didalam penelitian ini dalam bentuk \*.csv (Comma Separated Value). Dinas Pendidikan memberikan 24 dataset, yaitu Data data 21 Kabupaten dan 3 Kota periode 2023/2024 yang ada di Sulawesi Selatan. Total keseluruhan data tersebut memiliki 19 kolom, 9202 baris, dan total isi data sebanyak 119.605.

##### **3.2.1 Penentuan Variabel**

Dari banyaknya variabel yang ada pada dataset, berdasarkan kriteria BOP dan yang di inginkan Dinas Pendidikan Sulawesi Selatan adalah variabel luas wilayah (wilayah\_total), akreditas (skor\_akreditasi), daya listrik (daya\_listrik), jumlah ruangan (ruangan\_total), jumlah staf (staf\_total), jumlah guru (jumlah\_guru), jumlah siswa (siswa\_total), serta Lintang (lintang) dan Bujur (bujur) hanya digunakan pada

pembuatan peta interaktif. Itu adalah variabel yang akan digunakan dalam penelitian ini. Variabel yang dipilih ini berdasarkan kriteria untuk bantuan BOP, kelengkapan data yang dimiliki oleh setiap sekolah, dan kebutuhan untuk mengelompokan data.

### 3.2.2 Penggabungan Dataset

Pada tahap ini akan menggabungkan total 24 dataset yang ada, dataset sekolah SD,SMP,SMA,dan SMK dari 21 Kabupaten dan 3 Kota yang ada di Sulawesi Selatan.

Name	Date Modified	Type	Size
Kab. Bantaeng.xlsx	2/21/2025 3:48 PM	Microsoft Excel W...	37 KB
Kab. Barru.xlsx	2/21/2025 4:02 PM	Microsoft Excel W...	41 KB
Kab. Bone.xlsx	2/21/2025 4:44 PM	Microsoft Excel W...	120 KB
Kab. Bulukumba.xlsx	2/21/2025 8:29 PM	Microsoft Excel W...	68 KB
Kab. ENrekang.xlsx	2/21/2025 8:42 PM	Microsoft Excel W...	47 KB
Kab. Gowa.xlsx	2/22/2025 9:06 AM	Microsoft Excel W...	90 KB
Kab. Jeneponto.xlsx	2/22/2025 9:49 AM	Microsoft Excel W...	61 KB
Kab. Luwu Timur.xlsx	2/22/2025 10:13 AM	Microsoft Excel W...	42 KB
Kab. Luwu Utara.xlsx	2/22/2025 10:29 AM	Microsoft Excel W...	55 KB
KAb. Luwu.xlsx	2/22/2025 9:57 AM	Microsoft Excel W...	68 KB
Kab. Pangkajene Kepulauan.xlsx	2/22/2025 11:24 AM	Microsoft Excel W...	68 KB
Kab. Pinrang.xlsx	2/22/2025 12:43 PM	Microsoft Excel W...	61 KB
Kab. Sinjai.xlsx	2/23/2025 7:16 AM	Microsoft Excel W...	50 KB
Kab. Soppeng.xlsx	2/23/2025 7:36 AM	Microsoft Excel W...	49 KB
Kab. Takalar.xlsx	2/23/2025 7:46 AM	Microsoft Excel W...	51 KB
Kab. Tana Toraja.xlsx	2/23/2025 7:57 AM	Microsoft Excel W...	53 KB
Kab. Toraja Utara.xlsx	2/23/2025 8:10 AM	Microsoft Excel W...	47 KB
Kab. Wajo.xlsx	2/23/2025 8:24 AM	Microsoft Excel W...	73 KB
Kab.Sidrap.xlsx	2/22/2025 1:17 PM	Microsoft Excel W...	50 KB
Kep. Selayar.xlsx	2/23/2025 9:43 AM	Microsoft Excel W...	38 KB
Kota Makassar.xlsx	2/12/2025 4:55 AM	Microsoft Excel W...	15 KB
Kota Palopo.xlsx	2/23/2025 9:52 AM	Microsoft Excel W...	28 KB
Maros.xlsx	2/22/2025 10:56 AM	Microsoft Excel W...	60 KB
Pare-pare.xlsx	2/23/2025 12:10 PM	Microsoft Excel W...	27 KB

Gambar 11. Mentahan dataset sebelum di gabung.

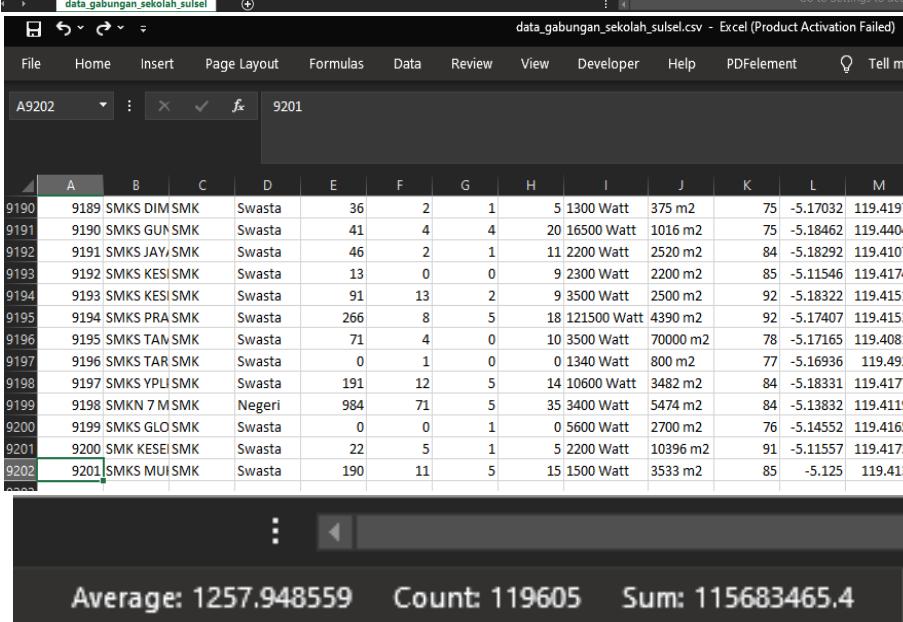
24 dataset ini memiliki kolom yang sama dengan total 19 kolom, yaitu no, nama sekolah, NPSN,BP, (Kategori), Status, jumlah sync,PD (Jumlah Siswa),Rombel (rombongan belajar), Pegawai,ruangan kelas,ruangan lab, ruangan perpus,jumlah ruangan,daya listrik, luas wilayah,skor akreditasi, koordinat lintang, dan koordinat bujur.

No	Nama Sekolah	NPSN	BP	Data Sekolah Kota Makassar							
				Status	Last Sync	Jml Sync	PD	Rombel	Guru	Pegawai	I
1	UPT SPF SD INPRES BARRANG CADDI 2	40312637	SD	Negeri	15 Feb 2024 11:34:16	8	80	6	6	3	
2	UPT SPF SD INPRES BARRANG LOMPO	40312029	SD	Negeri	18 Apr 2024 18:46:23	6	263	12	15	2	
3	UPT SPF SD INPRES BERTINGKAT TABARINGAN I	40313725	SD	Negeri	23 Apr 2024 08:12:49	11	250	12	16	3	
4	UPT SPF SD INPRES CAMBAYA 1	40312024	SD	Negeri	27 Apr 2024 09:02:01	19	328	12	13	3	
5	UPT SPF SD INPRES CAMBAYA 2	40312038	SD	Negeri	27 Apr 2024 16:25:59	21	265	11	15	6	
6	UPT SPF SD INPRES TABARINGAN	40313727	SD	Negeri	18 Apr 2024 15:16:15	10	289	11	15	4	
7	UPT SPF SD INPRES TABARINGAN I	40313726	SD	Negeri	26 Apr 2024 16:14:04	22	248	10	14	4	
8	UPT SPF SD NEGERI BARRANG CADDI	40313875	SD	Negeri	01 May 2024 17:32:18	15	175	6	9	4	
9	UPT SPF SD NEGERI BARRANG LOMPO	40312198	SD	Negeri	23 Jan 2024 09:29:19	2	343	12	18	4	
10	UPT SPF SD NEGERI CAMBAYA	40312083	SD	Negeri	22 Apr 2024 22:38:08	23	286	12	16	5	
11	UPT SPF SD NEGERI KODINGARENG	40313833	SD	Negeri	30 Apr 2024 15:21:02	17	502	20	27	6	
12	UPT SPF SD NEGERI LANGKAI	40313755	SD	Negeri	19 Apr 2024 17:12:05	15	131	6	8	2	
13	UPT SPF SD NEGERI LUMU-LUMU	40312090	SD	Negeri	16 Feb 2024 21:23:53	2	131	6	9	1	
14	UPT SPF SD NEGERI PATTINGALOONG 1	40312091	SD	Negeri	21 Apr 2024 09:54:44	22	320	12	16	3	
15	UPT SPF SD NEGERI TABARINGAN 5	40313724	SD	Negeri	17 Mar 2024 20:53:39	12	178	8	9	4	
16	UPT SPF SD NEGERI UIUNG TANAH 2	40312094	SD	Negeri	30 Apr 2024 12:49:11	19	378	12	16	6	
17	UPT SPF SD NEGERI UIUNG TANAH I	40312093	SD	Negeri	06 May 2024 08:51:27	27	315	12	15	3	
18	SD HANG TUAH MAKASSAR	40313109	SD	Swasta	04 Apr 2024 11:36:32	15	279	12	16	5	
19	SD MUHAMMADIYAH 7	40314042	SD	Swasta	04 May 2024 14:20:45	13	192	6	5	1	
R. Kelas	R. Lab	R. Perpus	Jumlah Ruang	Data Listrik	Luas Wilayah	Skor Akreditasi	Lintang	Bujur			
4	0	0	4	1200 Watt	768 m <sup>2</sup>	77	-5.0369	119.28			
7	0	1	8	7200 Watt	3300 m <sup>2</sup>	92	-5.0498	119.33			
12	0	1	13	4900 Watt	2916 m <sup>2</sup>	86	-5.1114	119.42			
6	1	1	8	2200 Watt	1984 m <sup>2</sup>	86	-5.1118	119.43			
9	1	1	11	2200 Watt	1572 m <sup>2</sup>	86	-5.1112	119.43			
5	0	1	6	900 Watt	484 m <sup>2</sup>	85	-5.1204	119.42			
5	0	1	6	800 Watt	172 m <sup>2</sup>	84	-5.1203	119.42			
6	0	1	7	2200 Watt	1200 m <sup>2</sup>	83	-5.0805	119.32			
10	1	1	12	3600 Watt	2218 m <sup>2</sup>	91	-5.0478	119.33			
7	0	1	8	1200 Watt	893 m <sup>2</sup>	85	-5.11	119.43			
9	0	1	10	2200 Watt	1500 m <sup>2</sup>	84	-5.149	119.26			
6	0	0	6	2200 Watt	1575 m <sup>2</sup>	87	-5.0321	119.09			
6	0	1	7	800 Watt	280 m <sup>2</sup>	83	-4.9766	119.21			
6	0	0	6	1200 Watt	-	83	-5.1139	119.42			
6	0	1	7	3890 Watt	2600 m <sup>2</sup>	84	-5.1145	119.42			
6	0	1	7	2200 Watt	1600 m <sup>2</sup>	83	-5.1128	119.42			
8	0	1	9	910 Watt	400 m <sup>2</sup>	81	-5.1127	119.42			
10	1	1	12	9320 Watt	3360 m <sup>2</sup>	92	-5.1154	119.42			
7	2	1	10	810 Watt	300 m <sup>2</sup>	84	-5.1117	119.43			

Gambar 12. Gambaran salah satu dataset sekolah sebelum digabung dari kota Makassar

Setelah di gabung menghasilkan 12 kolom, 9202 baris, dan 119.605. Kolom pada saat setelah penggabungan dataset ini adalah luas wilayah (wilayah\_total), akreditas (skor\_akreditasi), daya listrik (daya\_listrik), jumlah ruangan (ruangan\_total), jumlah staf (staf\_total), jumlah guru (jumlah\_guru), jumlah siswa (siswa\_total), Lintang (lintang), dan Bujur (bujur). Diberi nama dataset baru ini dengan nama data\_gabungan\_sekolah\_sulsel.csv.

No	Nama Sekolah	BP	Status	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi	Lintang	Bujur
1	UPT SPF SD INPRES BADDOKA	SD	Negeri	337	18	3	13	1300 watt	2808 m <sup>2</sup>	81	-5.09154	119.5134
2	UPT SPF SD INPRES BAKUNG 1	SD	Negeri	243	15	4	8	900 watt	3000 m <sup>2</sup>	84	-5.07336	119.5294
3	UPT SPF SD INPRES BAKUNG 2	SD	Negeri	269	16	5	8	4400 watt	2500 m <sup>2</sup>	84	-5.07914	119.5292
4	UPT SPF SD INPRES DAYA	SD	Negeri	498	28	3	13	4640 watt	1200 m <sup>2</sup>	87	-5.1129	119.512
5	UPT SPF SD INPRES KALANG TUBUNG 2	SD	Negeri	195	15	1	8	1300 watt	1505 m <sup>2</sup>	81	-5.08089	119.5367
6	UPT SPF SD INPRES KALANGTUBUNG 1	SD	Negeri	311	16	4	8	2297 watt	1650 m <sup>2</sup>	82	-5.09389	119.5354
7	UPT SPF SD INPRES LAE-LAE 2	SD	Negeri	162	14	6	7	900 watt	1200 m <sup>2</sup>	82	-5.06841	119.473
8	UPT SPF SD INPRES LAIKANG	SD	Negeri	654	23	6	15	1300 watt	2542 m <sup>2</sup>	85	-5.09583	119.5327
9	UPT SPF SD INPRES MANDAI	SD	Negeri	339	16	4	8	5000 watt	1430 m <sup>2</sup>	86	-5.08054	119.5213
10	UPT SPF SD INPRES MANGGA TIGA	SD	Negeri	509	25	6	11	2200 watt	1925 m <sup>2</sup>	92	-5.12353	119.5312
11	UPT SPF SD INPRES MANNURUKI 1	SD	Negeri	353	21	4	9	1400 watt	3250 m <sup>2</sup>	88	-5.11544	119.5319
12	UPT SPF SD INPRES MANNURUKI 2	SD	Negeri	327	19	3	8	1300 watt	2640 m <sup>2</sup>	82	-5.16755	119.4451
13	UPT SPF SD INPRES PACCERAKKANG	SD	Negeri	686	33	2	17	4400 watt	2229 m <sup>2</sup>	91	-5.1199	119.5224
14	UPT SPF SD INPRES PAGANDONGAN 1	SD	Negeri	346	16	5	8	2200 watt	4900 m <sup>2</sup>	79	-5.07906	119.4935
15	UPT SPF SD INPRES PAGANDONGAN 2	SD	Negeri	232	11	2	7	950 watt	2800 m <sup>2</sup>	82	-5.07904	119.4934
16	UPT SPF SD INPRES PAI 1	SD	Negeri	461	24	5	9	1300 watt	1600 m <sup>2</sup>	79	-5.07902	119.4934
17	UPT SPF SD INPRES PAI 2	SD	Negeri	299	17	6	7	1400 watt	3300 m <sup>2</sup>	83	-5.06846	119.473
18	UPT SPF SD INPRES PAJAJAING	SD	Negeri	336	16	6	8	900 watt	1700 m <sup>2</sup>	88	-5.1037	119.5224
19	UPT SPF SD INPRES PAJAJAING 2	SD	Negeri	639	27	3	14	1300 watt	2580 m <sup>2</sup>	86	-5.10603	119.5347
20	UPT SPF SD INPRES SUDIANG	SD	Negeri	306	18	5	7	900 watt	1460 m <sup>2</sup>	82	-5.08047	119.5213
21	UPT SPF SD INPRES TANGKALA 1	SD	Negeri	246	16	2	11	1300 watt	1875 m <sup>2</sup>	83	-5.09217	119.5247
22	UPT SPF SD INPRES TANGKALA II	SD	Negeri	561	33	4	12	2200 watt	2100 m <sup>2</sup>	86	-5.10303	119.5352
23	UPT SPF SD NEGERI BADDOKA	SD	Negeri	506	25	4	10	1300 watt	1920 m <sup>2</sup>	85	-5.10447	119.4917
24	UPT SPF SD NEGERI BATUBAMBUNG	SD	Negeri	207	14	3	10	1200 watt	3276 m <sup>2</sup>	85	-5.16809	119.4452
25	UPT SPF SD NEGERI BULUROKENG	SD	Negeri	702	32	3	14	1300 watt	2808 m <sup>2</sup>	85	-5.0776	119.4992



Gambar 13. data setelah digabung menjadi 1 dan diberi nama data\_gabungan\_sekolah\_sulsel.csv

### 3.2.3 Pembersihan Data

Akan dilakukan pembersihan data pada dataset yang telah digabung (data\_gabungan\_sekolah\_sulsel.csv). Pembersihan ini diawali dengan mengubah data yang masih kosong atau berisi “-“ dengan angka 0 seperti pada gambar 14. Data yang memiliki nilai 0 pada tahapan DBSCAN akan hilang ketika proses clustering karena akan dibaca sebagai noise/outlier. Selain itu tahap ini akan menghilangkan satuan yang masih terdapat pada daya listrik (Watt), dan satuan luas wilayah (m<sup>2</sup>). seperti yang terlihat pada gambar 14 dan 15.

[8]: df.head()

No	Nama Sekolah	BP	Status	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi	Lintang	Bujur
0	1	UPT SPF SD INPRES BADDOKA	SD Negeri	337.0	18	3	13.0	1300 watt	2808 m2	81	-5.091536	119.513350
1	2	UPT SPF SD INPRES BAKUNG 1	SD Negeri	243.0	15	4	8.0	900 watt	3000 m2	84	-5.073361	119.529407
2	3	UPT SPF SD INPRES BAKUNG 2	SD Negeri	269.0	16	5	8.0	4400 watt	2500 m2	84	-5.079141	119.529189
3	4	UPT SPF SD INPRES DAYA	SD Negeri	498.0	28	3	13.0	4640 watt	1200 m2	87	-5.112903	119.512008
4	5	UPT SPF SD INPRES KALANG TUBUNG 2	SD Negeri	195.0	15	1	8.0	1300 watt	1505 m2	81	-5.080895	119.536749

Gambar 14. Dataset sebelum di bersihkan

Setelah dilakukan pembersihan dataset maka seluruh variabel yang akan digunakan berbentuk numerik dengan membersihkan satuan watt,m2,tanda “-“, mengisi data kosong dengan angka 0, menghilangkan tanda “.” pada dataset.

[24]: df

No	Nama Sekolah	BP	Status	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi
0	1	UPT SPF SD INPRES BADDOKA	SD Negeri	337.0	18	3	13.0	1300	2808.0	81
1	2	UPT SPF SD INPRES BAKUNG 1	SD Negeri	243.0	15	4	80.0	900	3000.0	84
2	3	UPT SPF SD INPRES BAKUNG 2	SD Negeri	269.0	16	5	80.0	4400	2500.0	84
3	4	UPT SPF SD INPRES DAYA	SD Negeri	498.0	28	3	130.0	4640	1200.0	87
4	5	UPT SPF SD INPRES KALANG TUBUNG 2	SD Negeri	195.0	15	1	80.0	1300	1505.0	81

Gambar 15. Dataset setelah dilakukan pembersihan

### 3.3 Normalisasi Data

Ditahap ini menggunakan normalisasi data atau *feature scaling*. Tahap ini dilakukan dengan menggunakan tools Jupyter Notebook pada file normalisasi\_data.ipyn. Tahap ini sangat diperlukan untuk menyamakan rentang nilai data dari masing-masing variabel. Normalisasi ini menggunakan package yang terdapat pada Jupyter Notebook yang menggunakan metode MinMax Scaler, yaitu StandardScaler. StandardScaler ini akan diimplementasikan pada 9 variabel yang digunakan, yaitu kolom jumlah siswa (siswa\_total), jumlah guru (jumlah\_guru), jumlah staf (staf\_total), jumlah ruangan (jumlah\_ruangan), daya listrik (daya\_listrik), luas wilayah (wilayah\_total), dan akreditas (skor\_akreditasi). Sebelum menormalisasi data, dilakukan import library yang dibutuhkan. Library yang dibutuhkan adalah pandas, numpy, matplotlib.pyplot, dan StandardScaler, serta mendeklarasi dan memanggil data.

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler

[3]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_gabungan_sekolah_susel.csv'
data = pd.read_csv(filelocation)

[4]: data.head()
```

No	Nama Sekolah	BP	Status	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi
0	1 UPT SPF SD INPRES BADDOKA	SD	Negeri	337	18	3	13	1300	2808	81
1	2 UPT SPF SD INPRES BAKUNG 1	SD	Negeri	243	15	4	8	900	3000	84
2	3 UPT SPF SD INPRES BAKUNG 2	SD	Negeri	269	16	5	8	4400	2500	84
3	4 UPT SPF SD INPRES DAYA	SD	Negeri	498	28	3	13	4640	1200	87
4	5 UPT SPF SD INPRES KALANG TUBING 2	SD	Negeri	195	15	1	8	1300	1505	81

A

Gambar 16. Pengimportan library yang di butuhkan dan menampilkan dataset Yang telah di bersihkan sebelumnya

Setelah itu dilakukan proses normalisasi data dengan metode *MinMax Scaling* menggunakan *library StandardScaler* pada Jupyter Notebook seperti pada Gambar 17 di bawah ini.

```
[25]: x = data.loc[:, ['siswa_total', 'jumlah_guru', 'staf_total', 'jumlah_ruangan', 'daya_listrik', 'wilayah_total', 'skor_akreditasi']]

[26]: scaler = StandardScaler()
scaler.fit (x)
data_scaled = scaler.transform(x)
```

Gambar 17. Normaliasi data dengan StandardScaler

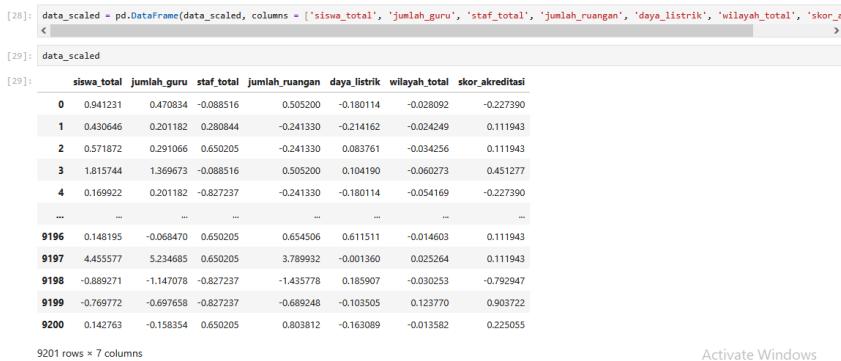
Menginisialisasi variabel (x) sebagai array data yang berisi nilai dari kolom jumlah siswa (siswa\_total), jumlah guru (jumlah\_guru), jumlah staf (staf\_total), jumlah ruangan (jumlah\_ruangan), daya listrik (daya\_listrik), luas wilayah (wilayah\_total), dan akreditas (skor\_akreditasi). Selain itu, disini juga mendeklarasikan scaler sebagai fungsi StandardScaler. Selanjutnya, data\_scaled didefinisikan sebagai array yang berisi hasil normalisasi data. Hasil normalisasi tersebut memiliki rentang nilai yang seragam, yaitu antara -1 hingga 1.

```
[8]: data_scaled
```

```
[8]: array([[ 0.9412311 ,  0.47083351, -0.08851645, ..., -0.22739045,
       0.35771901,  5.11200806],
       [ 0.430646 ,  0.20118151,  0.28084404, ...,  0.11194342,
       5.29355719,  5.11278903],
       [ 0.57187167,  0.29106551,  0.65020453, ...,  0.11194342,
       5.29980749, -0.11951391],
       ...,
       [-0.88927078, -1.14707847, -0.82723743, ..., -0.7929469 ,
       5.37159007,  5.10729636],
       [-0.76977214, -0.69765848, -0.82723743, ...,  0.90372245,
       5.33920226,  5.10733737],
       [ 0.14276292, -0.15835449,  0.65020453, ...,  0.22505471,
      -0.19288573, -0.70029893]])
```

Gambar 18. Hasil dari normalisasi data

Setelah proses Data Scaling selesai, akan dilakukan konversi kembali array data tersebut ke dalam dataframe dengan menggunakan nama kolom yang sama. Dataframe ini kemudian disimpan sebagai dataset baru dengan nama **data\_normalisasi\_sekolah\_sulsel.csv**. Langkah ini harus dilakukan untuk mempersiapkan data agar siap digunakan dalam proses berikutnya.



```
[28]: data_scaled = pd.DataFrame(data_scaled, columns = ['siswa_total', 'jumlah_guru', 'staf_total', 'jumlah_ruangan', 'daya_listrik', 'wilayah_total', 'skor_akreditasi'])
[29]: data_scaled
```

	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi
0	0.941231	0.470834	-0.088516	0.505200	-0.180114	-0.028092	-0.227390
1	0.430646	0.201182	0.280844	-0.241330	-0.214162	-0.024249	0.111943
2	0.571872	0.291066	0.650205	-0.241330	0.083761	-0.034256	0.111943
3	1.815744	1.369673	-0.088516	0.505200	0.104190	-0.060273	0.451277
4	0.169922	0.201182	-0.827237	-0.241330	-0.180114	-0.054169	-0.227390
...	...	...	...	...	...	...	...
9196	0.148195	-0.068470	0.650205	0.654506	0.611511	-0.014603	0.111943
9197	4.455577	5.234685	0.650205	3.789932	-0.001360	0.025264	0.111943
9198	-0.889271	-1.147078	-0.827237	-1.435778	0.186907	-0.030253	-0.792947
9199	-0.769772	-0.697658	-0.827237	-0.689248	-0.103505	0.123770	0.903722
9200	0.142763	-0.158354	0.650205	0.803812	-0.163089	-0.013582	0.225055

9201 rows x 7 columns

Activate Windows

Gambar 19. Mengonversi data array ke DataFrame

```
[12]: data_scaled.to_csv('data_normalisasi_sekolah_sulsel.csv', index = False)
```

Gambar 20. Download data array menjadi data\_normalisasi\_sekolah\_sulsel.csv

### 3.4 Principal Component Analysis (PCA)

Pada tahapan PCA (Principal Component Analysis), proses dimulai dengan standarisasi data agar setiap variabel memiliki skala yang sama, mengingat PCA sangat peka terhadap varians. Setelah itu, dihitung matriks kovarians untuk memahami bagaimana variabel-variabel dalam data saling berkorelasi. Dari matriks kovarians ini, eigenvalues dan eigenvectors kemudian dihitung. Eigenvalues menunjukkan seberapa besar varians yang bisa dijelaskan oleh setiap komponen utama, sedangkan eigenvectors menentukan arah dari masing-masing komponen tersebut. Selanjutnya, principal components dipilih berdasarkan nilai eigenvalues terbesar, karena komponen inilah yang mampu menjelaskan sebagian besar varians dalam data. Data asli kemudian ditransformasikan ke ruang baru yang dibentuk oleh principal components, menghasilkan dataset dengan dimensi yang lebih sedikit namun tetap mempertahankan informasi penting dari data awal. Proses ini diakhiri dengan evaluasi varians untuk menentukan berapa banyak komponen utama yang akan digunakan, biasanya dengan melihat cumulative explained variance, yaitu total proporsi varians yang dijelaskan oleh komponen-komponen utama tersebut. Hasil akhirnya adalah data yang lebih sederhana dan efisien untuk dianalisis, tanpa kehilangan terlalu banyak informasi penting.

```
[1]: #Penerapan Teknik PCA

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA

[2]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_normalisasi_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)

[3]: data.head()
```

	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi
0	0.941231	0.470834	-0.088516	0.50520	-0.180114	-0.028092	-0.227390
1	0.430646	0.201182	0.280844	-0.24133	-0.214162	-0.024249	0.111943
2	0.571872	0.291066	0.650205	-0.24133	0.083761	-0.034256	0.111943
3	1.815744	1.369673	-0.088516	0.50520	0.104190	-0.060273	0.451277
4	0.169922	0.201182	-0.827237	-0.24133	-0.180114	-0.054169	-0.227390

Gambar 21. Import library dan pemanggilan data

```
[4]: features = ['siswa_total', 'jumlah_guru', 'staf_total', 'jumlah_ruangan', 'daya_listrik', 'wilayah_total', 'skor_akreditasi']
x = data.loc[:, features].values
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)
data_pca = pd.DataFrame(data = principalComponents, columns = ['pca1', 'pca2'])

[5]: data_pca
```

	pca1	pca2
0	0.901478	-0.305677
1	0.315612	-0.143266
2	0.622986	-0.014275
3	1.906723	0.042274
4	-0.334741	-0.184505
...	...	...
9196	0.723251	0.325091
9197	7.218822	-0.627236

Gambar 22. Pengimplementasian teknik PCA

Dalam Gambar 22 dibaris pertama, penggunaan analisis PCA dengan mendefinisikan features untuk menentukan kolom-kolom yang akan dijadikan parameter dalam penelitian ini, yaitu jumlah siswa (siswa\_total), jumlah guru (jumlah\_guru), jumlah staf (staf\_total), jumlah ruangan (jumlah\_ruangan), daya listrik (daya\_listrik), luas wilayah (wilayah\_total), dan akreditasi (skor\_akreditasi). Selanjutnya, pada baris kedua, mengambil nilai dari kolom-kolom yang telah ditetapkan di features. Proses analisis PCA kemudian dijalankan menggunakan library PCA pada baris ketiga dan keempat. Terakhir, hasil analisis PCA dikonversi ke dalam bentuk DataFrame yang diberi nama data\_pca untuk mempermudah pengolahan data lebih lanjut.

[9]:	data_pca	
	pca1	pca2
0	1.628340	3.349251
1	1.781699	6.710388
2	1.373750	3.268771
3	2.638114	3.056547
4	0.434320	3.433142
...	...	...
9196	2.215221	6.915696
9197	8.545746	5.302014
9198	-0.665992	7.397538
9199	0.133514	7.232193
9200	0.548619	-0.779776

9201 rows × 2 columns

```
[11]: data_pca.to_csv('data_pca_sekolah_sulsel.csv', index = False)
```

Gambar 23. Hasil analisis menggunakan teknik PCA

Hasil analisis PCA ditampilkan pada Gambar 23, di mana proses ini menghasilkan 2 kolom dan 9201 baris. Kedua kolom tersebut akan digunakan sebagai variabel dalam penelitian ini. Agar mempermudah proses berikutnya, DataFrame **data\_pca** disimpan dalam format Comma Separated Value (CSV) dengan nama **data\_pca\_sekolah\_sulsel.csv**.

### 3.5 DBSCAN Clustering

Setelah menyelesaikan preprocessing data, langkah berikutnya adalah mengimplementasikan algoritma **DBSCAN** pada data hasil preprocessing, yaitu **data\_pca\_sekolah\_sulsel.csv**. Algoritma **DBSCAN** digunakan untuk menentukan jumlah klaster atau kelompok yang optimal dalam data penelitian ini, sekaligus mengidentifikasi dan menghilangkan data noise atau outlier. Untuk memastikan konsistensi jumlah klaster yang dihasilkan oleh **DBSCAN**, digunakan **Silhouette Score** sebagai metrik untuk mengevaluasi kinerja algoritma tersebut. Proses implementasi ini dilakukan menggunakan **Jupyter Notebook** pada file **dbscan.ipynb**

```
: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import DBSCAN
from sklearn.neighbors import NearestNeighbors
from itertools import product

: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_pca_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)
```

Gambar 24. Import library dan pemanggilan data PCA

Kembali mengimport library yang di butuhkan dan melakukan pemanggilan data setelah di lakukan teknik PCA yaitu dengan nama `data_pca_sekolah_sulsel.csv`

```
: features = ['pca1', 'pca2']
x1 = data.loc[:100, features].values
```

Gambar 25. Mendeklarasikan features dan nilai x1

### 3.5.1 DBSCAN Clustering With KNN

Selanjutnya mendefinisikan fitur (features) untuk menentukan kolom mana yang akan digunakan dalam proses ini, sebagaimana ditunjukkan pada Gambar 26. Kolom yang dipilih adalah pca1 dan pca2. Setelah itu, mendeklarasikan x1 untuk mengambil nilai dari kolom yang telah ditentukan dalam features dan baris yang dipilih, yaitu baris 0-100. Nilai dalam x1 ini akan digunakan pada NearestNeighbors menggunakan data dari baris 0-100. Data tersebut merupakan sampel yang bertujuan menyederhanakan visualisasi grafik, sehingga memudahkan dalam menentukan nilai Eps yang akan digunakan.

```
11]: nn = NearestNeighbors(n_neighbors=2)
nbrs = nn.fit(x1)
distances, indices=nbrs.kneighbors(x1)
```

Gambar 26. Penggunaan algoritma Nearest Neighbors

Selanjutnya, mendefinisikan nn sebagai NearestNeighbors untuk mengukur jarak antara dua titik. Setelah itu, mengimplementasikan NearestNeighbors dengan memanfaatkan data dari x1.

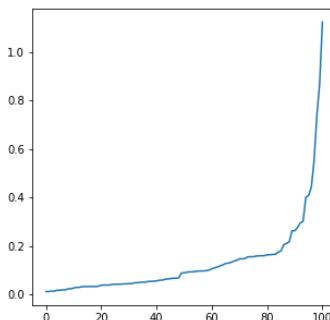
### 3.5.2 Penentuan Nilai Epsilon, Min\_Samples, Silhouette Score

Epsilon	Min_Samples	Silhouette_Score
0.1	2	0.067161
0.1	3	0.23845
0.1	4	0.243368
0.1	5	0.234241
0.1	6	0.437928
0.1	7	0.436649
0.1	8	0.432287
0.1	9	0.434173

Epsilon	Min_Samples	Silhouette_Score
0.1	10	0.083537
0.3	2	0.610818
0.3	3	0.624495
0.3	4	0.625176
0.3	5	0.625209
0.3	6	0.609087
0.3	7	0.637635
0.3	8	0.66017
0.3	9	0.665094
0.3	10	0.665912
0.5	2	0.646299
0.5	3	0.652694
0.5	4	0.657141
0.5	5	0.638071
0.5	6	0.637235
0.5	7	0.639574
0.5	8	0.642367
0.5	9	0.643838
0.5	10	0.664387
0.7	2	0.435984
0.7	3	0.65734
0.7	4	0.667537
0.7	5	0.667605
0.7	6	0.667551
0.7	7	0.667757
0.7	8	0.664972

Tabel 4. Tabel nilai epsilon, min\_samples, dan Silhouette Score

```
[8]: distances = np.sort(distances, axis=0)
distances = distances[:,1]
plt.rcParams['figure.figsize'] = (5,5)
plt.plot(distances)
plt.show()
```



Gambar 27. Grafik visualisasi *Silhouette Score*

Berdasarkan hasil visualisasi *Silhouette Score* dan table nilai *Epsilon*, *min\_samples*, dan *Silhouette Score* titik jarak optimal berada dalam rentang 0.4 hingga 0.7. Oleh karena itu, diputuskan untuk menggunakan nilai *Eps* sebesar 0.4-0.7, dengan *Min\_Sample* masing-masing 4, 5, 6, 7, dan 8.

```
[69]: x = data.loc[:, features].values
dbscan = DBSCAN(eps=0.7, min_samples=7).fit(x)
labels = dbscan.labels_
```

Gambar 28. DBSCAN Clustering

Setelah memperoleh nilai *Eps* yang optimal yaitu 0.7, saatnya mengaplikasikan algoritma DBSCAN Clustering. Selanjutnya, mendeklarasikan variabel *x* sebagai array yang berisi seluruh data dari kolom yang telah ditentukan dalam *features*. Implementasi dilakukan menggunakan data array *x* dengan *Eps* sebesar 0.7 dan *Min\_Samples* sebesar 7 serta jumlah klaster yang dihasilkan oleh algoritma DBSCAN adalah 4 seperti pada Gambar 29. Selanjutnya, mendefinisikan array hasil DBSCAN dengan nama *labels*.

```
[70]: print("klaster total:{}".format(len(set(labels[np.where(labels != -1)]))))
print(labels)

klaster total:4
[0 1 0 ... 1 1 2]
```

Gambar 29. Klaster total

Tahapan ini yaitu mendefinisikan hasil klasterisasi dari penerapan algoritma DBSCAN ke dalam sebuah DataFrame dengan label “*dbscan\_kluster*”, seperti yang ditampilkan pada Gambar 30. DataFrame ini kemudian akan dikombinasikan

dengan data `data_gabungan_sekolah_sulsel.csv` dan `data_pca_sekolah_sulsel.csv` untuk mengintegrasikan label-label yang telah ada sebelumnya.

[27]:	<code>dbscanonly_kluster</code>
0	0
1	1
2	0
3	0
4	0
...	...
9196	1
9197	3
9198	1
9199	1
9200	2

9201 rows × 1 columns

Gambar 30. Membuat DataFrame `dbscan_kluster`

### 3.5.3 Pencarian Data Noise/Outlier

```
[56]: klaster[klaster['dbscanonly_kluster'] == -1].count()
[56]: dbscanonly_kluster    51
      dtype: int64
```

Gambar 31. Total data yang dianggap sebagai noise/outlier

Sebanyak **51** data terdeteksi sebagai *noise* atau *outlier*. Menentukan jumlah data *noise/outlier* dengan menghitung jumlah data yang memiliki label bernilai -1 pada kolom “`dbscan_kluster`”. Data yang terdeteksi sebagai *noise* atau *outlier* perlu kesampingkan karena titik-titik tersebut berada jauh dari kelompok data lainnya, yang dapat memengaruhi keakuratan pembentukan klaster. Selain itu, *noise* atau *outlier* sebaiknya tidak dimasukkan ke dalam klaster dan baiknya di lakukan penyederhanaan/dikesampingkan karena posisinya yang terlalu jauh dari data lain, sehingga tidak merepresentasikan pola klaster secara valid. Ditahap ini akan dilakukan penghapusan nilai klaster -1 dan ada 3 data yang null/kosong sehingga perlu di hapus.

```
[8]: if 'dbSCAN_kluster' in data.columns:

    data_simplification = data[data['dbSCAN_kluster'] != -1]

    print("Data setelah penyederhanaan:")
    print(data_simplification.head())
else:
    print("Kolom 'dbSCAN_kluster' tidak ditemukan dalam data.")

Data setelah penyederhanaan:
   No          Nama Sekolah  BP  Status siswa_total \
0  1          UPT SPF SD INPRES BADDOKA  SD Negeri      337
1  2          UPT SPF SD INPRES BAKUNG 1  SD Negeri      243
2  3          UPT SPF SD INPRES BAKUNG 2  SD Negeri      269
3  4          UPT SPF SD INPRES DAYA  SD Negeri      498
4  5          UPT SPF SD INPRES KALANG TUBUNG 2  SD Negeri      195

   jumlah_guru  staf_total jumlah_ruangan daya_listrik wilayah_total \
0           18            3           13        1300        2808
1           15            4            8         900        3000
2           16            5            8        4400        2500
3           28            3           13        4640        1200
4           15            1            8        1300        1505

   skor_akreditasi  Lintang     Bujur      pca1      pca2  dbSCAN_kluster
0             81  509153567  1195133502  1.628340  3.349251          0
1             84  5073361105  1195294071  1.781699  6.710388          0
2             84  5079140801  119529189  1.373750  3.268771          0
3             87  5112902708  119512008  2.638114  3.056547          0
4             81  5080894528  119536749  0.434320  3.433142          0
```

Gambar 32. Penyederhanaan data noise/outlier

### 3.5.4 Uji Validasi *Silhouette Score* DBSCAN

Validasi algoritma DBSCAN dilakukan menggunakan *Silhouette Score*. Nilai *Silhouette Score* berkisar antara -1 hingga +1, di mana nilai yang lebih tinggi menunjukkan bahwa suatu titik data memiliki kemiripan yang kuat dengan klasternya sendiri dan berbeda secara signifikan dari klaster lainnya. Dalam proses ini, harus mengimport *Silhouette\_Samples* dan *Silhouette\_Score* sebagai pustaka yang digunakan. Selanjutnya, mendefinisikan *SI* sebagai array yang berisi hasil perhitungan *Silhouette Score*. Pengujian *Silhouette* ini menggunakan dataset *data\_pca\_sekolah\_sulse.csv* dan *labels* yang merepresentasikan hasil klasterisasi DBSCAN dalam bentuk array.

```
[73]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(data, labels)
print(SI)
```

Gambar 33. Hasil uji validasi *Silhouette Score* DBSCAN

Berdasarkan hasil uji validasi algoritma DBSCAN menghasilkan skor **“0.667756759373751”**.

### 3.5.5 Penggabungan Dataset

Tahap berikutnya dalam implementasi *DBSCAN Clustering* adalah proses penggabungan data. Pada tahap ini, akan menggabungkan dataset awal ***data\_gabungan\_sekolah\_sulsel.csv*** dengan ***data\_pca\_sekolah\_sulsel.csv*** serta hasil klasterisasi *DBSCAN*, yaitu klaster yang telah didefinisikan dalam bentuk array bernama *klaster*. Tujuan dari penggabungan ini adalah untuk menempatkan hasil klasterisasi ke dalam dataset awal, sekaligus mempermudah proses pada tahap selanjutnya. Sebelum menggabungkan data, harus mengimpor *data\_gabungan\_sekolah\_sulsel.csv* dan mendefinisikannya sebagai *datasekolah*. Sementara itu, *data\_pca\_sekolah\_sulsel.csv* tidak diimpor ulang karena sebelumnya sudah didefinisikan dengan nama *data*.

```
[61]: datasekolah = pd.read_csv('C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_gabungan_sekolah_sulsel.csv')

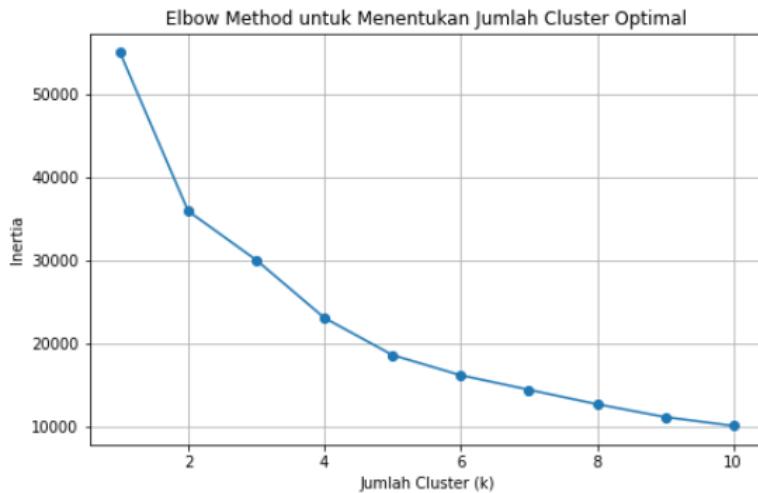
[62]: hasil1 = pd.concat([data, pd.DataFrame(klaster)],axis=1)
       hasil2 = pd.concat([datasekolah, pd.DataFrame(hasil1)],axis=1)
       hasil2.to_csv('data_dari_dbSCAN.csv', index = False)
       hasil2
```

Gambar 34. Penggabungan data hasil DBSCAN dengan PCA dan data awal

Tahap ini dilakukan proses penggabungan data dalam dua tahap. Tahap pertama adalah menggabungkan *data\_pca\_sekolah\_sulsel.csv* dengan hasil klasterisasi *DBSCAN*, yang kemudian didefinisikan sebagai *hasil1*. Tahap kedua melibatkan penggabungan *data\_gabungan\_sekolah\_sulsel.csv* dengan *hasil1*, dan hasilnya disimpan dalam *result2*. Setelah itu, mengekspor *DataFrame hasil2* ke dalam format *Comma Separated Value (CSV)* dengan nama file *data\_dari\_dbSCAN.csv*.

### 3.5.6 Grafik Jumlah Klaster Optimal

Pada Gambar 35, **Sumbu X (Jumlah klaster ( $k$ ))** Menunjukkan jumlah klaster yang diuji, dari 1 hingga 10, **Sumbu Y (Inertia)** menunjukkan total jarak kuadrat antara data dan pusat klasternya. Semakin kecil nilai ini, semakin baik data dikelompokkan, dan Titik Siku (Elbow Point), titik di mana penurunan inertia mulai melambat secara signifikan. Ini adalah indikasi jumlah klaster optimal. Titik siku pada grafik Gambar 35 terlihat jelas sebagai perubahan dari penurunan tajam ke penurunan yang lebih landai. Jika grafik menunjukkan penurunan yang signifikan sampai sekitar  $k=2$  atau  $k=4$  sebelum melambat, maka jumlah klaster optimal kemungkinan berada di sekitar angka tersebut sehingga didalam penelitian ini akan dilakukan pengujian dari klaster optimal  $k=2, k=3$ , dan  $k=4$  untuk peningkatan *DBSCAN* dan *K-Means Only*. Implementasi algoritma *DBSCAN* dalam proses klasterisasi akan menghasilkan jumlah klaster yang optimal, anggota dari masing-masing klaster, serta data *noise* atau *outlier*. Jumlah klaster yang diperoleh akan digunakan sebagai acuan untuk menentukan jumlah klaster dalam proses selanjutnya yaitu peningkatan *DBSCAN* dengan *K-Means Clustering*.



Gambar 35. Grafik jumlah klaster optimal

### 3.6 Peningkatan DBSCAN Dengan K-Means Clustering

Peningkatan DBSCAN ini adalah gabungan dari beberapa algoritma sehingga menciptakan algoritma untuk pengelompokan yang baik dalam mengelompokkan sekolah rekomendasi yang membutuhkan bantuan BOP sesuai dengan kriteria yang ada. Dengan bantuan dari beberapa algoritma seperti KNN dan K-Means maka algoritma peningkatan DBSCAN ini bisa memberikan hasil terbaik dalam pengelompokan rekomendasi sekolah dalam penyaluran BOP untuk penelitian ini.

Setelah menentukan jumlah klaster yang optimal dan menyederhanakan/mengesampingkan data dari noise atau outlier pada tahap DBSCAN Clustering, dilanjutkan proses klasterisasi menggunakan algoritma K-means Clustering untuk peningkatan DBSCAN. Dataset yang digunakan merupakan hasil dari DBSCAN Clustering, yaitu **data\_dari\_dbSCAN.csv**. Hasil klasterisasi dengan bantuan K-means Clustering ini menjadi output utama dalam algoritma peningkatan DBSCAN ini. Langkah awal dalam proses peningkatan DBSCAN adalah mengimpor library yang diperlukan, yaitu KMeans. Selanjutnya, memuat dataset **data\_dari\_dbSCAN.csv** dan mendefinisikannya dengan nama data.

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

[2]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_dari_dbSCAN.csv'
data = pd.read_csv(filelocation)

[3]: data.head()
```

BP	Status	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi	Lintang	Bujur	pca1	pca2	dbSCAN_kluster
SD	Negeri	337	18	3	13	1300	2808	81	509153567	1195133502	1.628340	3.349251	0
SD	Negeri	243	15	4	8	900	3000	84	5073361105	1195294071	1.781699	6.710388	0
SD	Negeri	269	16	5	8	4400	2500	84	5079140801	119529189	1.373750	3.268771	0

Gambar 36. Import library dan KMeans dan pemanggilan dataset *data\_dari\_dbSCAN.csv*

Pada Gambar 37, data yang terdapat pada label atau kolom ditentukan sebagai parameter untuk proses *Peningkatan DBSCAN*. Tahap ini menggunakan data hasil analisis PCA dari dataset *data\_dari\_dbSCAN.csv*, yaitu kolom *pca1* dan *pca2*. Kedua kolom tersebut didefinisikan dalam *fitur*. Selanjutnya, mendeklarasikan array *x* yang berisi nilai-nilai dari kolom *pca1* dan *pca2*.

```
[4]: fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values
```

Gambar 37. Penentuan kolom untuk peningkatan DBSCAN

### 3.6.1 Pengujian n\_cluster (*k*) = 3

Selanjutnya, mengimplementasikan K-Means Clustering menggunakan pustaka KMeans. Mendefinisikan *kmeans* untuk menjalankan algoritma K-Means Clustering dan melakukan prediksi pola pengelompokan klaster dengan menetapkan *n\_clusters* = 3. Untuk menghindari jebakan centroid, digunakan metode *k-means++*. Setelah itu, membuat objek *y\_kmeans* dengan menerapkan metode *fit\_predict* pada variabel *x* yang telah didefinisikan, guna memperoleh indeks klaster.

```
[9]: dbSCAN_kmeans = KMeans(n_clusters=3, init = 'k-means++')
y_dbSCAN_kmeans = dbSCAN_kmeans.fit_predict(x)

[10]: y_dbSCAN_kmeans

[10]: array([1, 1, 1, ..., 1, 1, 0])
```

Gambar 38. Pengimplementasian algoritma K-Means pada peningkatan DBSCAN

Objek `y_dbSCAN_kmeans` pada Gambar 38 merupakan hasil prediksi dari algoritma *K-Means Clustering* yang disajikan dalam bentuk array. Hasil klasterisasi tersebut kemudian di definisikan ke dalam sebuah **DataFrame** dengan label “peningkatanDBSCAN\_3klaster”, seperti yang ditampilkan pada Gambar 39 dengan nama objek `klaster_dbSCAN_kmeans`. Pada Gambar 39, juga terlihat *DataFrame* `klaster` yang memuat hasil klasterisasi.

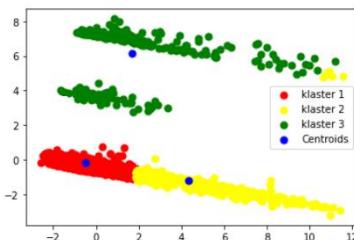
```
[25]: klaster_dbSCAN_kmeans = pd.DataFrame(data = y_dbSCAN_kmeans, columns = ['peningkatanDBSCAN_3klaster'])

[25]:   peningkatanDBSCAN_3klaster
      0              2
      1              2
      2              2
      3              2
      4              2
      ...
      9180           2
      9181           2
      9182           2
      9183           2
      9184           0
      9185 rows x 1 columns
```

Gambar 39. DataFrame dari hasil K-Means Clustering pada peningkatan DBSCAN

Selanjutnya, pada Gambar 40, memvisualisasikan hasil klasterisasi K-Means menggunakan scatter plot. Parameter dalam `plt.scatter` terdiri dari sumbu x dan sumbu y, di mana sumbu x merepresentasikan kolom `pca1` dan sumbu y merepresentasikan kolom `pca2`. Kedua sumbu ini diambil dari objek `x` yang telah didefinisikan sebelumnya. Untuk memvisualisasikan klaster 1, sumbu x ditulis sebagai `x[y_dbSCAN_kmeans == 0, 0]`, yang berarti data poin berasal dari objek x. Pernyataan `y_dbSCAN_kmeans == 0` digunakan untuk memilih baris-baris dalam `x` yang termasuk dalam klaster 1. Angka 0 setelah koma menunjukkan kolom pertama, yaitu `pca1`. Sementara itu, sumbu y ditulis sebagai `x[y_dbSCAN_kmeans == 0, 1]`, di mana baris yang dipilih adalah data dengan label klaster 0, dan angka 1 setelah koma merepresentasikan kolom kedua, yaitu `pca2`. Untuk menampilkan titik-titik klaster lainnya, dilakukan pengubahan nilai `y_dbSCAN_kmeans` pada baris berikutnya. Visualisasi hasil K-Means Clustering ini kemudian dapat dilihat pada Gambar 40.

```
[282]: plt.scatter(x[y_dbSCAN_kmeans==0,0], x[y_dbSCAN_kmeans==0,1], s=50, c='red', label = 'klaster 1')
plt.scatter(x[y_dbSCAN_kmeans==1,0], x[y_dbSCAN_kmeans==1,1], s=50, c='yellow', label = 'klaster 2')
plt.scatter(x[y_dbSCAN_kmeans==2,0], x[y_dbSCAN_kmeans==2,1], s=50, c='green', label = 'klaster 3')
plt.scatter(dbSCAN_kmeans.cluster_centers_[:, 0], dbSCAN_kmeans.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()
```



Gambar 40. Grafik hasil K-Means Clustering pada peningkatan DBSCAN

Proses selanjutnya adalah menggabungkan dataset awal data\_dari\_dbSCAN.csv dengan DataFrame hasil implementasi K-Means Clustering, yaitu klaster. Hasil dari penggabungan ini kemudian didefinisikan sebagai hasil. Setelah itu, menyimpan DataFrame hasil dalam format Comma Separated Value (CSV) dengan nama data\_hasil\_peningkatan\_dbSCAN\_3cluster.csv, seperti yang ditampilkan pada Gambar 41. Dataset hasil penggabungan tersebut menjadi keluaran akhir dari implementasi K-Means Clustering pada dalam DBSCAN dan tahap selanjutnya akan menggunakan dataset tersebut untuk pemetaan dan pengelompokan sekolah yang membutuhkan rekomendasi bantuan BOP jika memiliki skor silhouette paling tinggi di akhir. Tahap akhir dari metode *Peningkatan DBSCAN* adalah melakukan uji validasi terhadap hasil pengembangan *K-Means Clustering* dengan *DBSCAN Clustering*. Digunakan *Silhouette Score* untuk mengukur validitas klaster, dengan rentang nilai antara -1 hingga 1. Hasil dari pengujian n\_cluster = 3 adalah “**0.7801306473707272**”

```
[71]: hasil = pd.concat([data, pd.DataFrame(klaster_dbSCAN_kmeans)], axis=1)
hasil.to_csv('data_hasil_peningkatan_dbSCAN_3cluster.csv', index = False)

[69]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_dbSCAN_kmeans)
print(SI)

0.7801306473707272
```

Gambar 41. Penggabungan dataset data\_dari\_dbSCAN.csv dengan DataFrame klaster serta uji validasi n\_cluster = 3

### 3.6.2 Pengujian n\_cluster ( $k$ ) = 2 dan n\_cluster ( $k$ ) = 4

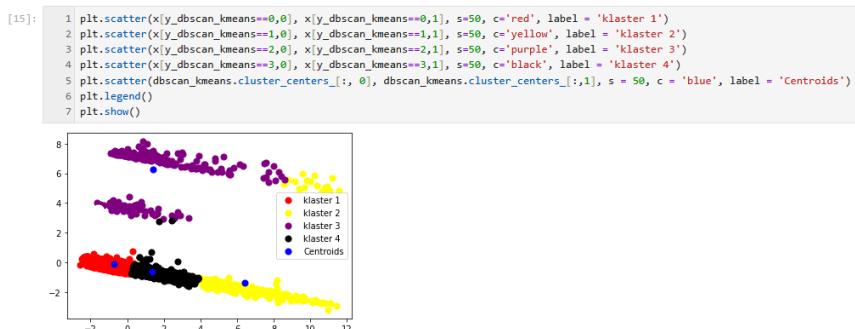
Tahap ini akan dilakukan pengujian terhadap n\_cluster = 2 dan n cluster = 4 untuk membandingkan skor silhouette yang paling tinggi agar model dari hasil tersebut dapat digunakan untuk pembentukan pengelompokan dan pemetaan sekolah untuk rekomendasi bantuan BOP. Pada pengujian n\_cluster = 2 di dapatkan skor silhouette

sebesar “**0.7505380325803438**” dan pada pengujian n\_cluster 4 di dapati skor silhouette sebesar “**0.6582168177473653**”.

```
[7]: klaster_dbSCAN_kmeans = pd.DataFrame(data = y_dbSCAN_kmeans, columns = ['peningkatanDBSCAN_4klaster'])

[7]:   peningkatanDBSCAN_4klaster
    0           3
    1           3
    2           3
    3           3
    4           3
    ...
9180          3
9181          2
9182          3
9183          3
9184          0
9185 rows × 1 columns
```

Gambar 42. Pengimplementasian algoritma K-Means Clustering dalam peningkatan DBSCAN dan hasil DataFrame pada n\_cluster = 4

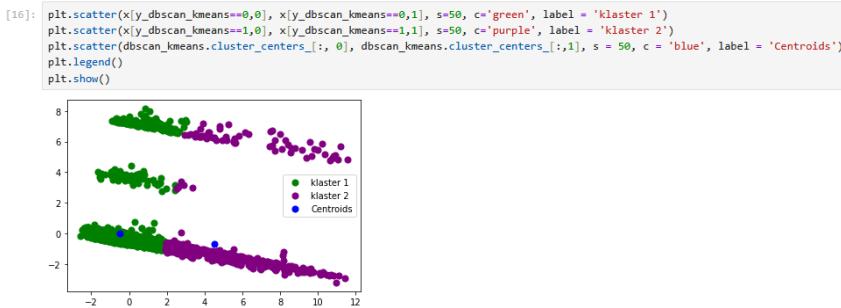


Gambar 43. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n\_cluster = 4

```
[9]: klaster_dbSCAN_kmeans = pd.DataFrame(data = y_dbSCAN_kmeans, columns = ['peningkatanDBSCAN_2klaster'])

[9]:   peningkatanDBSCAN_2klaster
    0           0
    1           0
    2           0
    3           1
    4           0
    ...
9180          0
9181          1
9182          0
9183          0
9184          0
9185 rows × 1 columns
```

Gambar 44. Pengimplementasian algoritma K-Means Clustering dalam peningkatan DBSCAN dan hasil DataFrame pada n\_cluster = 2



Gambar 45. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n\_cluster = 2

```
[33]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_dbSCAN_kmeans)
print(SI)
```

0.7505380325803438

Gambar 46. Hasil uji validasi Silhouette Score n\_cluster = 2

```
[32]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_dbSCAN_kmeans)
print(SI)
```

0.6582168177473653

Gambar 47. Hasil uji validasi Silhouette Score n\_cluster = 4

### 3.6.3 Hasil Uji Validasi Peningkatan DBSCAN

Pada pengujian dengan nilai  $k = 2$ , metode peningkatan DBSCAN menghasilkan dua klaster, yaitu klaster 1 dengan 6747 anggota dan klaster 2 dengan 2404 anggota. Selanjutnya, untuk pengujian dengan  $k = 3$ , metode ini membentuk tiga klaster, di mana klaster 1 memiliki 2290 anggota, klaster 2 sebanyak 3731 anggota, dan klaster 3 berjumlah 3130 anggota. Sementara itu, pada pengujian terakhir dengan  $k = 4$ , Peningkatan DBSCAN menghasilkan empat klaster, terdiri dari klaster 1 dengan 1308 anggota, klaster 2 sebanyak 6149 anggota, klaster 3 dengan 1006 anggota, dan klaster 4 berjumlah 688 anggota.

Jumlah Klaster	Silhouette Score
2	0.7505380325803438
3	0.7801306473707272
4	0.6582168177473653

Tabel 5. Hasil uji validasi peningkatan DBSCAN

Penggabungan metode K-Means Clustering dengan DBSCAN Clustering sehingga tercipta peningkatan DBSCAN menggunakan data Sekolah

SD,SMP,SMA,dan SMK yang ada di Sulawesi Selatan menghasilkan *Silhouette Score* tertinggi sebesar 0.78, seperti yang ditampilkan pada tabel 5, dengan pembentukan 3 klaster. Nilai ini menunjukkan bahwa proses pembentukan 3 klaster melalui metode gabungan tersebut memiliki struktur klaster yang kuat, karena skor tersebut berada dalam rentang 0.7 hingga 1.

Berdasarkan semua grafik hasil klasterisasi pada Gambar 40, Gambar 43, dan Gambar 45 menggunakan metode Peningkatan DBSCAN, terlihat bahwa penyebaran titik pada setiap klaster tampak jelas dan teratur. Tidak ditemukan titik-titik data yang berada jauh dari pusat klaster atau centroid. Hal ini disebabkan oleh penyederhanaan data noise atau outlier sebelumnya. Hasil klasterisasi dari pengembangan metode ini menunjukkan bahwa titik-titik data dalam setiap klaster memiliki kemiripan yang erat dengan titik centroid.

### 3.7 Metode K-Means Only

Implementasi ini adalah dengan menerapkan metode K-means Clustering tanpa integrasi dengan metode DBSCAN Clustering. Tujuan penerapan K-means ini adalah untuk membandingkan hasil klasterisasi antara metode yang dikembangkan dan metode standar. Pada tahap ini, menggunakan data hasil preprocessing, yaitu data\_pca\_sekolah\_sulsel.csv. Selain itu, ditahap ini juga menentukan jumlah klaster yang akan dibentuk. Dalam hal ini, jumlah klaster yang digunakan mengikuti jumlah klaster pada proses sebelumnya, yakni 2, 3, dan 4, sesuai dengan hasil peningkatan DBSCAN sebelumnya. Sama seperti sebelumnya, langkah pertama dalam penerapan K-means Only adalah mengimpor library yang diperlukan, yaitu KMeans. Selain itu, ditahap ini juga memuat dataset data\_pca\_sekolah\_sulsel.csv dan mendefinisikannya sebagai data. Selanjutnya, seperti yang ditunjukkan pada Gambar 48, memilih data dari label atau kolom tertentu yang akan dijadikan parameter dalam K-means Clustering. Data yang digunakan merupakan hasil analisis PCA yang terdapat dalam data\_pca\_sekolah\_sulsel.csv, yakni kolom pca1 dan pca2. Kedua kolom tersebut kemudian didefinisikan dalam variabel fitur. Setelah itu, mendeklarasikan array x untuk merepresentasikan nilai dari kolom pca1 dan pca2.

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_pca_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)

data.head()

    pca1      pca2
0  1.628340  3.349251
1  1.781699  6.710388
2  1.373750  3.268771
3  2.638114  3.056547
4  0.434320  3.433142

fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values

```

Gambar 48. Import library,pemanggilan data, dan penentuan variabel K-Means Only

### 3.7.1 Pengujian n\_cluster ( $k$ ) = 3

Ditahap ini sama seperti saat metode peningkatan DBSCAN dengan menerapkan K-Means Clustering menggunakan library KMeans serta mendefinisikan *kmeans\_only* untuk menjalankan algoritma K-Means Clustering. Proses pengelompokan klaster dilakukan dengan menggunakan *n\_clusters* = 3, serta dilakukan pengujian tambahan dengan *n\_clusters* bernilai 2 dan 4 untuk membandingkan hasil dan mencari konfigurasi terbaik. Dalam proses ini digunakan *k-means++* untuk menghindari jebakan centroid. Setelah itu, objek *y\_kmeans\_only* didefinisikan menggunakan metode *fit\_predict* pada variabel *x* yang sebelumnya telah ditetapkan, guna memperoleh indeks klaster.

```

[1294]: kmeans_only = KMeans(n_clusters=3, init = 'k-means++')
y_kmeans_only = kmeans_only.fit_predict(x)

[1295]: y_kmeans_only

[1295]: array([0, 2, 0, ..., 2, 2, 0])

```

Gambar 49. Implementasi K-Means Clustering pada K-Means Only

Objek *y\_kmeans\_only* pada Gambar 49 merupakan hasil prediksi dari algoritma K-Means Clustering dalam bentuk *array*. Hasil klasterisasi yang diperoleh melalui implementasi algoritma K-Means ini kemudian didefinisikan dalam bentuk

*DataFrame* dengan label "kmeans\_only\_klaster". Seperti yang ditampilkan pada Gambar 50, *DataFrame* tersebut diberi nama objek *klaster\_kmeans\_only*.

```
[20]: klaster_kmeans_only = pd.DataFrame(data = y_kmeans_only, columns = ['kmeans_only_3klaster'])

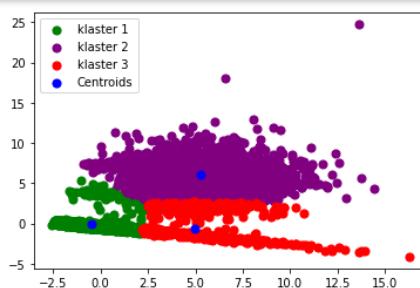
[20]: kmeans_only_3klaster
      0
      1
      2
      3
      4
      ...
9196  2
9197  2
9198  2
9199  2
9200  0

9201 rows × 1 columns
```

Gambar 50. Pembuatan DataFrame hasil K-Means Only

Memvisualisasikan hasil klasterisasi K-Means menggunakan scatter plot. Parameter dalam fungsi plt.scatter terdiri dari sumbu x diikuti oleh sumbu y, di mana sumbu x merepresentasikan kolom pca1 dan sumbu y merepresentasikan kolom pca2. Kedua sumbu ini diambil dari objek x yang telah didefinisikan sebelumnya. Selain itu tahap ini juga memvisualisasikan anggota dari tiga klaster yang terbentuk dengan warna hijau, ungu, dan merah, serta menampilkan titik centroid masing-masing klaster dengan warna biru, sebagaimana terlihat pada Gambar 51.

```
plt.scatter(x[y_kmeans_only==0,0], x[y_kmeans_only==0,1], s=50, c='green', label = 'klaster 1')
plt.scatter(x[y_kmeans_only==1,0], x[y_kmeans_only==1,1], s=50, c='purple', label = 'klaster 2')
plt.scatter(x[y_kmeans_only==2,0], x[y_kmeans_only==2,1], s=50, c='red', label = 'klaster 3')
plt.scatter(kmeans_only.cluster_centers_[:, 0], kmeans_only.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()
```



Gambar 51. Grafik hasil K-Means only dengan n\_cluster = 3

Penggabungan dataset awal, yaitu **data\_pca\_sekolah\_sulsel.csv**, dengan *DataFrame* hasil klasterisasi K-Means yang telah didefinisikan sebagai *klaster\_kmeans\_only*. Hasil dari proses penggabungan ini kemudian didefinisikan

sebagai **hasil**. Selanjutnya, menyimpan DataFrame **hasil** dalam format *Comma Separated Value* (CSV) dengan nama **data\_hasil\_kmeans\_only.csv**, seperti yang ditunjukkan pada Gambar 52. Dataset hasil penggabungan inilah yang menjadi keluaran akhir dari implementasi K-Means Only. Selain itu ditahap ini dilakukan uji validasi dengan menerapkan Uji *Silhouette* untuk menilai kualitas pembentukan klaster. Hasil uji *Silhouette Score* K-Means only dengan  $n\_cluster = 3$  adalah “**0.6939256280849587**”.

```
[1320]: hasil = pd.concat([data, pd.DataFrame(klaster_kmeans_only)], axis=1)
hasil.to_csv('data_hasil_kmeans_only.csv', index = False)

[1319]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_kmeans_only)
print(SI)

0.6939256280849587
```

Gambar 52. Penggabungan dataset data\_pca\_sekolah\_sulsel dan DataFrame klaster\_kmeans\_only

### 3.7.2 Pengujian $n\_cluster (k) = 2$ dan $n\_cluster (k) = 4$

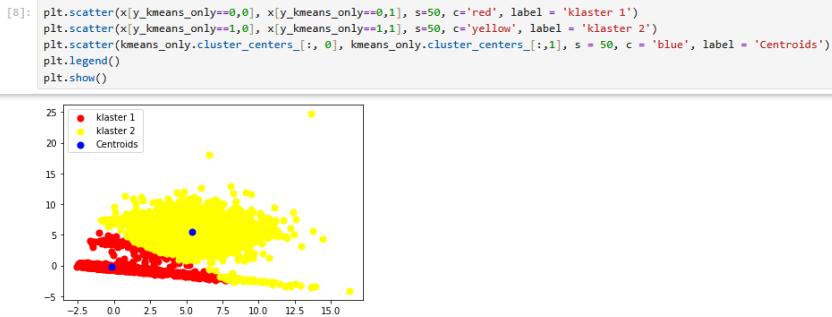
Ditahap akan dilakukan pula pengujian terhadap  $n\_cluster = 2$  dan  $n\_cluster = 4$  untuk membandingkan *skor silhouette* yang paling tinggi dan jika skornya mengalahkan metode peningkatan DBSCAN maka model data dari K-Means Only yang akan digunakan untuk melakukan pemetaan rekomendasi sekolah yang membutuhkan bantuan BOP. Pada pengujian  $n\_cluster = 2$  di dapat *skor silhouette* sebesar “**0.7331125971428345**” dan pada pengujian  $n\_cluster 4$  di dapat *skor silhouette* sebesar “**0.6582168177473653**”.

```
[7]: klaster_kmeans_only = pd.DataFrame(data = y_kmeans_only, columns = ['kmeans_only_2klaster'])
klaster_kmeans_only
```

	kmeans_only_2klaster
0	0
1	1
2	0
3	1
4	0
...	...
9196	1
9197	1
9198	1
9199	1
9200	0

9201 rows × 1 columns

Gambar 53. Pengimplementasian algoritma K-Means Clustering pada K-Means Only dan hasil DataFrame pada  $n\_cluster = 2$



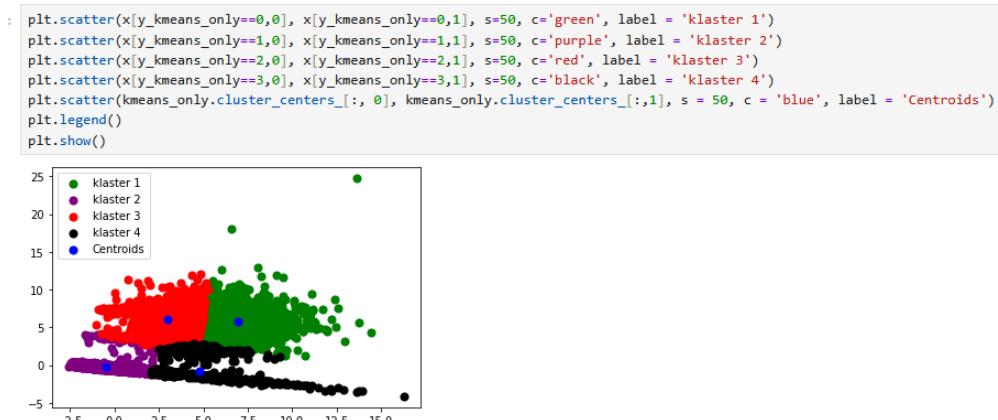
Gambar 54. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n\_cluster = 2

```
[7]: klaster_kmeans_only = pd.DataFrame(data = y_kmeans_only, columns = ['kmeans_only_4klaster'])
klaster_kmeans_only
```

	kmeans_only_4klaster
0	1
1	1
2	1
3	1
4	2
...	...
9196	1
9197	3
9198	1
9199	1
9200	2

9201 rows × 1 columns

Gambar 55. Pengimplementasian algoritma K-Means Clustering pada K-Means Only dan hasil DataFrame pada n\_cluster = 4



Gambar 56. Grafik hasil K-Means Clustering pada peningkatan DBSCAN n\_cluster = 4

```
from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_kmeans_only)
print(SI)
```

0.7331125971428345

Gambar 57. Hasil uji validasi *Silhouette Score* n\_cluster = 2

```
from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_kmeans_only)
print(SI)
```

0.6634546357693659

Gambar 58. Hasil uji validasi *Silhouette Score* n\_cluster = 4

### 3.7.3 Hasil Uji Validasi K-Means Only

Hasil klasterisasi yang dilakukan menggunakan metode **K-Means Only** dengan tujuan klasterisasi ini adalah untuk membandingkan hasilnya dengan metode yang telah dikembangkan (Peningkatan DBSCAN). Penentuan jumlah klaster disesuaikan dengan kebutuhan. Dataset yang digunakan berasal dari hasil *preprocessing*, yaitu `data_pca_sekolah_sulsel.csv`. Sama seperti pada klasterisasi dengan **Peningkatan DBSCAN**, pengujian dilakukan dengan mencoba berbagai jumlah klaster, yaitu **2**, **3**, dan **4 klaster** sama dengan pengujian peningkatan DBSCAN. Pada pengujian dengan nilai  $k = 2$ , metode peningkatan DBSCAN menghasilkan dua klaster, yaitu klaster 1 dengan 2291 anggota dan klaster 2 dengan 6910 anggota. Selanjutnya, untuk pengujian dengan  $k = 3$ , metode ini membentuk tiga klaster, di mana klaster 1 memiliki 3251 anggota, klaster 2 sebanyak 3731 anggota, dan klaster 3 berjumlah 2220 anggota. Sementara itu, pada pengujian terakhir dengan  $k = 4$ , K-Means Only menghasilkan empat klaster, terdiri dari klaster 1 dengan 2180 anggota, klaster 2 sebanyak 4066 anggota, klaster 3 dengan 1582 anggota, dan klaster 4 berjumlah 1374 anggota.

Jumlah Klaster	Silhouette Score
2	0.7331125971428345
3	0.6939256280849587
4	0.6634546357693659

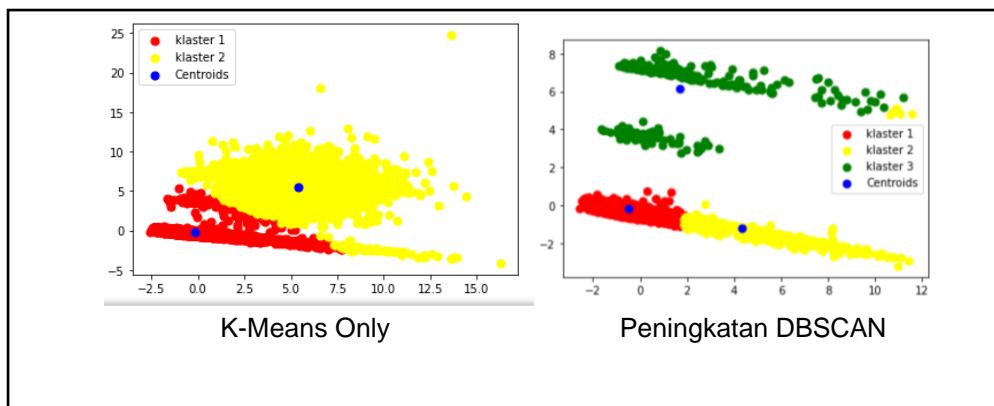
Tabel 6. Hasil uji validasi K-Means Only

Klasterisasi menggunakan metode K-Means Only pada data sekolah SD,SMA,SMP,dan SMA yang ada di Sulawesi Selatan memperoleh *Silhouette Score* tertinggi sebesar “**0.7331125971428345**”, seperti yang ditampilkan pada tabel 6, dengan pembentukan 2 klaster. Nilai ini menunjukkan bahwa proses

pembentukan 2 klaster menggunakan metode K-Means Clustering memiliki struktur yang kuat, karena berada dalam rentang 0.73 hingga 1.

Seperti yang ditampilkan grafik pada **Gambar 51**, **Gambar 54**, dan **Gambar 56**, **klaster 1** diwakili oleh **warna hijau**, **klaster 2** dengan **warna ungu**, **klaster 3** menggunakan **warna merah**, dan **klaster 4** menggunakan **warna hitam**. Berdasarkan grafik hasil klasterisasi menggunakan **metode K-Means Only**, terlihat bahwa setiap klaster memiliki anggota dengan titik-titik yang tersebar cukup jauh dari pusat kerumunan klasternya, bahkan ada yang berjarak signifikan dari **centroid**. Khusus pada **klaster 2**, penyebaran titik antar anggota klaster tampak lebar, mengindikasikan rendahnya tingkat kemiripan antar anggota dalam klaster tersebut.

### 3.8 Perbandingan Hasil Grafik Algoritma Peningkatan DBSCAN dan K-Means Only



Gambar 59. Grafik perbandingan hasil terbaik K-Means Only dan peningkatan DBSCAN

Berdasarkan hasil K-means Only dan peningkatan DBSCAN, akan dilakukan perbandingan untuk mengetahui metode mana yang lebih efisien agar modelnya dapat di pakai untuk pemetaan kelompok sekolah yang membutuhkan rekomendasi bantuan BOP. Perbandingan ini dilihat dari grafik hasil klasterisasi. Berikut perbandingan grafik dengan hasil terbaik, yaitu 2 klaster pada K-means Only dan 3 klaster pada peningkatan DBSCAN.

Dari hasil perbandingan grafik pada Gambar 59, terlihat perbedaan penyebaran data yang dihasilkan. Pada grafik K-means Only, terdapat beberapa data yang dalam metode peningkatan DBSCAN dianggap sebagai noise/outlier, namun justru membentuk klaster tersendiri. Klaster tersebut sebenarnya terdiri dari data noise/outlier yang seharusnya dihapus atau dikeluarkan dari proses klasterisasi. Hal ini terjadi karena K-means Only membentuk klaster hanya berdasarkan kedekatan titik data dengan centroid, tanpa mempertimbangkan syarat

lain. Setiap centroid akan membentuk klaster, dan titik data secara otomatis masuk ke klaster dengan centroid terdekat meskipun titik tersebut jauh dan tidak memiliki kemiripan dengan anggota klaster lainnya bahkan grafik K-Means Only (3 klaster) sangat berantakan jika diuji dengan 3 klaster yang dimana score silhouettenya semakin rendah karena tambahan centroid yang membentuk klaster dan data yang jauh dari klaster 3 akan otomatis masuk ke centroid baru yang terbentuk (terdekat) sehingga score silhouettenya akan semakin rendah. Sementara itu, pada hasil peningkatan DBSCAN, penyebaran data terlihat lebih rapi dan terstruktur. Pembagian klaster lebih jelas, dan setiap klaster terdiri dari data yang memiliki kemiripan tinggi, tanpa adanya atau mengesampingkan noise/outlier yang mengganggu hasil klasterisasi.

Berdasarkan grafik hasil terbaik, hasil ini menunjukkan bahwa metode **Peningkatan DBSCAN** lebih layak digunakan dalam proses klasterisasi. Sebaliknya, metode **K-Means Only** masih memproses dataset yang mengandung noise/outlier, sehingga hasil klasterisasi menjadi kurang optimal karena data yang seharusnya tidak termasuk dalam klaster tetap dimasukkan.

### 3.9 Perbandingan Hasil Uji Validasi DBSCAN Only, Peningkatan DBSCAN, dan K-Means Only

Jumlah Cluster	Silhouette Score		Jumlah Cluster	Silhouette Score
2	0.7505		2	0.7331
3	0.7801		3	0.6939
4	0.6582		4	0.6634

Hasil uji validasi DBSCAN Only

Tabel hasil uji validasi  
peningkatan DBSCAN

Tabel hasil uji validasi  
K-Means Only

Gambar 60. Perbandingan tabel hasil uji validasi *Silhouette Score* DBSCAN Only, peningkatan DBSCAN, dan K-Means Only

Berdasarkan Gambar 60 perbandingan score dari DBSCAN Only, Peningkatan DBSCAN, dan K-Means Only dapat diketahui bahwa hasil uji validasi menggunakan score yang baik di manangkan oleh metode peningkatan DBSCAN

dengan skor tertinggi yaitu 0.7801 dengan jumlah klaster 3 dimana skor tersebut sangat tinggi karena mendekati 1 yaitu pada rentang 0.78-1 mengalahkan hasil uji validasi dari K-Means Only. Hasil *Silhouette Score* dari K-Means Only yang tertinggi adalah 0.7331 dengan jumlah klaster 2 dan seluruh jumlah klaster 2,3,dan 4 dari K-Means Only tidak ada yang mengalahkan hasil uji validasi dengan skor tertinggi dari peningkatan DBSCAN. Sementara itu hasil *Silhouette Score* dari DBSCAN Only adalah 0.6677 dengan menghasilkan 4 jumlah klaster masih kalah dengan hasil uji validasi tertinggi dari K-Means Only dan peningkatan DBSCAN. Berdasarkan hasil keseluruhan algoritma peningkatan DBSCAN lah yang memiliki pemodelan algoritma yang paling baik dan akurat untuk dapat menentukan untuk membuat pemetaan/pengelompokan rekomendasi sekolah SD,SMP,SMA,dan SMK yang ada di Sulawesi Selatan untuk bantuan BOP sehingga didalam penelitian ini pemodelan dari algoritma peningkatan DBSCAN lah yang akan digunakan untuk pemetaan/pengelompokan sekolah rekomendasi untuk mendapatkan bantuan BOP dari Dinas Pendidikan Sulawesi Selatan berdasarkan kriteria yang sesuai untuk mendapatkan bantuan BOP.

#### 4.10 Uji Validasi Pembobotan Menggunakan Klaster Optimal Peningkatan

##### DBSCAN (3 klaster)

Pada tahap ini pemodelan yang akan digunakan adalah pemodelan dari peningkatan DBSCAN karena telah terbukti memiliki pemodelan yang paling baik yaitu pemodelan dengan peningkatan DBSCAN 3 klaster memiliki nilai *silhouette score* paling tinggi. Pertama-tama di tahap pemetaan ini dilakukan import library dan pemanggilan dataset data\_hasil\_peningkatan\_dbSCAN\_3cluster.

```
[32]: import pandas as pd
import folium
from folium.plugins import MarkerCluster, Search
from folium import FeatureGroup
from folium.plugins import MarkerCluster

[33]: file_path = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_hasil_peningkatan_dbSCAN_3cluster.csv'
df = pd.read_csv(file_path)
```

Gambar 61. Import library dan pemanggilan dataset  
data\_hasil\_peningkatan\_dbSCAN\_3cluster

Selanjutnya, pada Gambar 62 bobot untuk proses perangkingan ditentukan dalam bentuk dictionary,di mana masing-masing kriteria diberi bobot tertentu berdasarkan kepentingannya. Skor akreditasi diberi bobot tertinggi sebesar 29, karena semakin rendah skor akreditasi, sekolah dianggap semakin memerlukan bantuan. Jumlah siswa memiliki bobot 23 karena semakin banyak siswa, semakin besar kebutuhan fasilitas. Jumlah ruangan memiliki bobot 19, luas wilayah 14, daya listrik, jumlah guru, dan staf total masing-masing memiliki bobot 5, karena meski penting, mereka relatif memiliki pengaruh yang lebih kecil dibandingkan faktor-faktor utama lainnya. Seluruh pembobotan berdasarkan kriteria untuk sekolah

penerima bantuan BOP yang di jelaskan oleh Dinas Pendidikan Sulawesi Selatan yang tercantum pada Tabel 1.

```
[3]: # Bobot perangkingan
bobot = {
    'skor_akreditasi': 29,
    'siswa_total': 23,
    'jumlah_ruangan': 19,
    'wilayah_total': 14,
    'daya_listrik': 5,
    'jumlah_guru': 5,
    'staf_total': 5
}
```

Gambar 62. Bobot perangkingan

Setelah bobot ditetapkan, proses normalisasi dilakukan pada Gambar 63 untuk menghindari ketimpangan akibat skala data yang berbeda. Normalisasi menggunakan rumus Min-Max Scaling, di mana rentang data setiap kolom dikonversi ke skala antara 0 dan 1. Kolom jumlah siswa dinormalisasi secara langsung, karena semakin besar jumlah siswa, semakin baik untuk skor perangkingan. Namun, kolom lainnya seperti skor akreditasi, jumlah ruangan, luas wilayah, daya listrik, jumlah guru, dan staf total dinormalisasi secara terbalik. Artinya, semakin kecil nilainya, semakin baik posisinya dalam perangkingan, sehingga digunakan rumus  $1 - (\text{nilai} - \text{min}) / (\text{max} - \text{min})$ .

```
[4]: # Normalisasi data
for col in bobot.keys():
    if col == 'siswa_total':
        df[f'{col}_n'] = (df[col] - df[col].min()) / (df[col].max() - df[col].min())
    else:
        df[f'{col}_n'] = 1 - (df[col] - df[col].min()) / (df[col].max() - df[col].min())
```

Gambar 63. Normalisasi data menggunakan Min-Max Scaling

Skor akhir dihitung dengan mengalikan setiap kolom yang sudah dinormalisasi dengan bobotnya masing-masing, lalu menjumlahkan semua hasil perkalian. Ini menciptakan satu skor komposit yang merefleksikan performa total sekolah berdasarkan semua kriteria. Namun, karena skor ini bisa memiliki rentang yang tidak terstandar, dilakukan normalisasi ulang pada skor akhir tersebut, agar rentangnya menjadi antara 0 hingga 100. Hal ini dilakukan pada Gambar 64 agar skor lebih mudah dipahami dan dibandingkan.

```
# Hitung skor akhir
df['skor_rekomendasi'] = (
    df['skor_akreditasi_n'] * bobot['skor_akreditasi'] +
    df['siswa_total_n'] * bobot['siswa_total'] +
    df['jumlah_ruangan_n'] * bobot['jumlah_ruangan'] +
    df['wilayah_total_n'] * bobot['wilayah_total'] +
    df['daya_listrik_n'] * bobot['daya_listrik'] +
    df['jumlah_guru_n'] * bobot['jumlah_guru'] +
    df['staf_total_n'] * bobot['staf_total']
)

# Normalisasi skor akhir ke skala 100
df['skor_rekomendasi'] = (df['skor_rekomendasi'] - df['skor_rekomendasi'].min()) / (df['skor_rekomendasi'].max() - df['skor_rekomendasi'].min()) * 100
```

Gambar 64. Menjumlahkan hasil perkalian total performa sekolah dan normalisasi ulang agar rentangnya menjadi 0 hingga 100

Data kemudian diurutkan berdasarkan skor akhir secara descending, sehingga sekolah dengan skor tertinggi muncul di peringkat pertama. Peringkat diberikan secara berurutan, mulai dari 1 untuk skor tertinggi hingga seterusnya.

```
7]: # Peringkatan
df = df.sort_values(by='skor_rekomendasi', ascending=False).reset_index(drop=True)
df['peringkat'] = df.index + 1
```

Gambar 65. Pengurutan dan penentuan peringkat dengan skor tertinggi

Data skor akhir kemudian digabungkan dengan semua kolom asli lalu menampilkan 10 data teratas.

```
[30]: # Tampilkan semua kolom asli ditambah skor akhir dan peringkat
hasil_rekomendasi = df.copy()

[9]: # Tampilkan hasil perangkingan
print(hasil_rekomendasi.head(10))

      No           Nama Sekolah   BP Status siswa_total \
0  5260          UPT SDN 72 BONTOLE  SD Negeri    294
1  2097  SD INPRES PENGENTUNGANG UTARA  SD Negeri    406
2  2242          SD INPRES BORONGKALUKU  SD Negeri    334
3  4072          UPT SD NEGERI 149 PINRANG  SD Negeri    240
4  2227          SD NEGERI ANASSAPPU  SD Negeri    201
5  4475          SD NEGERI 4 ARAWA  SD Negeri    257
6  7163  UPTD SMPN 31 SATAP LALANG TEDONG  SMP Negeri    150
7  3076          SD KRISTEN BUNGIN PASANG  SD Swasta    146
8  2318          SD INPRES BIRING PANTING  SD Negeri    176
9  2671          UPT SD NEGERI 11 BANGKALA BARAT  SD Negeri    187

  jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total ...
0         8          2            7       5500        1965 ...
1        21          2            7       900        2594 ...
2         10          2            7       900        4126 ...
3         10          3            7       5500        602 ...
4         10          0            8       1300        1547 ...
5         12          2            10      5500        2237 ...
6         9          2            4       3500        578 ...
7         4          1            6       900        1626 ...
8         8          1            7       3500        2623 ...
9         9          2            7       900        3703 ...

  kmeans_klaster skor_akreditasi_n siswa_total_n jumlah_ruangan_n \
0             0           1.000000     0.184789      0.920455
1             1           0.999999     0.184789      0.920455
2             2           0.999999     0.184789      0.920455
3             3           0.999999     0.184789      0.920455
4             4           0.999999     0.184789      0.920455
5             5           0.999999     0.184789      0.920455
6             6           0.999999     0.184789      0.920455
7             7           0.999999     0.184789      0.920455
8             8           0.999999     0.184789      0.920455
9             9           0.999999     0.184789      0.920455

  wilayah_total_n daya_listrik_n jumlah_guru_n staf_total_n \
0  0.997940      0.980442      0.936      0.93750
1  0.997269      0.997164      0.832      0.93750
2  0.9995634     0.997164      0.920      0.93750
3  0.9993935     0.980442      0.920      0.90625
4  0.998386      0.995710      0.920      1.00000
5  0.997650      0.980442      0.904      0.93750
6  0.999420      0.987713      0.928      0.93750
7  0.998302      0.997164      0.968      0.96875
8  0.997238      0.987713      0.936      0.96875
9  0.996085      0.997164      0.928      0.93750

  skor_rekomendasi peringkat
0  100.000000      1
1  99.799970      2
2  98.374842      3
3  97.723149      4
4  97.154843      5
5  96.923130      6
6  96.762670      7
7  96.530093      8
8  96.520095      9
9  96.506138     10

[10 rows x 26 columns]
```

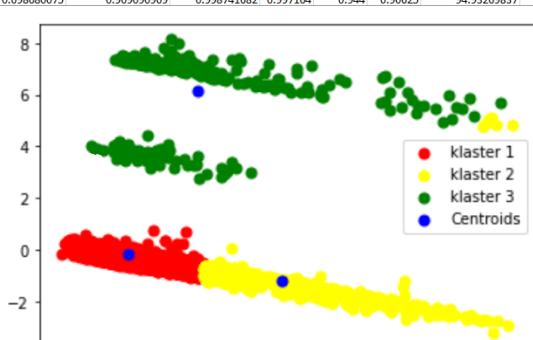
Gambar 66. Penggabungan data skor akhir dan kolom asli serta menampilkan 10 data teratas hasil perangkingan dari model peningkatan DBSCAN 3 klaster

Dalam format CSV maka tampilannya akan seperti Gambar 68 yang didalamnya ada total 9151 baris dengan peringkat 1-9151 yang paling membutuhkan bantuan BOP. Data kemudian disimpan dengan nama *Sekolah\_Rekomendasi\_Bantuan\_BOP\_PeningkatanDSBCAN3Cluster.csv*.

```
[34]: # Simpan hasil ke file CSV
hasil_rekomendasi.to_csv('Sekolah_Rekomendasi_Bantuan_BOP_PeningkatanDSBCAN3Cluster.csv', index=False)
```

Gambar 67. Menyimpan dataset dengan nama  
Sekolah\_Rekomendasi\_Bantuan\_BOP\_PeningkatanDSBCAN3Cluster.csv

1	No	Nama Sekolah	BP	Status	siswa_tot	jumlah_guru	staf_total	jumlah_ruangan	jumlah_wilayah	total	ru_daya	listr	wilayah_t	skor_ake	Lintang	Bujur	pca1	pca2	dbscan0	Peningkat	skor_ake:					
2	5260	UPT SDN 72 BONTOLOE	SD	Negeri	294	8	2	7	5500	1965	70	5.44719	118.828	-0.43201	-0.12744	0	0	1								
3	2097	SD INPRES PENGENTUNGAN	SD	Negeri	406	21	2	7	900	2594	71	5.34926	119.786	-0.43507	-0.43036	1	0	0.956522								
4	2242	SD INPRES BORONGKALUKU	SD	Negeri	334	10	2	7	900	4126	71	5.02884	119.913	-0.26239	-0.25791	0	0	0.956522								
5	4072	UPT SD NEGERI 149 PINRANG	SD	Negeri	240	10	3	7	5500	602	70	3.55179	120.0475	-0.32856	-0.171	0	0	1								
6	2227	SD NEGERI ANASAPPU	SD	Negeri	201	10	0	8	1300	1547	70	5.03842	120.0675	-0.87068	-0.10704	0	0	1								
7	4475	SD NEGERI 4 ARAWA	SD	Negeri	257	12	2	10	5500	2237	70	-3.8753	119.8707	-0.19646	-0.54236	1	0	1								
8	7163	UPTD SMPN 31 SATAP LALANG SMP	SMP	Negeri	150	9	2	4	3500	578	70	5.00296	119.5899	-1.01631	-0.05998	0	0	1								
9	3076	SD KRISTEN BUNGIN PASANG	SD	Swasta	146	4	1	6	900	1626	70	-2.65442	120.4369	-1.36077	-0.37961	1	0	1								
10	2316	SD INPRES BIRING PANTING	SD	Negeri	176	8	1	7	3500	2623	70	-5.07165	120.0287	-0.92534	-0.05317	1	0	1								
11	2671	UPT SD NEGERI 11B BANGKALA	SD	Negeri	187	9	2	7	900	3703	70	5.56328	119.6768	-0.78726	-0.51428	1	0	1								
12	3887	SD NEGERI 24 KALIBONE	SD	Negeri	194	8	2	8	1300	1040	70	-5.04391	119.9195	-0.66637	-0.17699	1	0	1								
13	2758	UPT SD NEGERI 11 TAROWANG	SD	Negeri	173	10	1	7	900	2171	70	-5.68204	119.7447	-0.86613	-0.13055	1	0	1								
14	2264	SD INPRES KARAMPUANG	SD	Negeri	187	7	4	7	2200	2381	70	-5.37958	119.9648	-0.55216	-0.56166	1	0	1								
15	2820	SD NEGERI 57 PADANG SAPPA	SD	Negeri	285	14	3	13	3500	4264	70	-3.12477	120.3359	0.045607	-0.36394	0	0	1								
16	2571	UPT SD NEGERI 16 BANGKALA	SD	Negeri	151	8	1	7	1300	2535	70	-5.57959	119.4099	-1.01347	-0.08689	1	0	1								
17	5338	UPT SD 183 INPRES KAMPUNG	SD	Negeri	189	10	4	7	3500	664	70	-5.45779	119.4886	-0.32752	-0.23519	1	0	1								
18	5128	UPT SD NEGERI 11B BONTOSAN	SD	Negeri	174	10	3	7	5500	1116	70	-5.45679	119.4994	-0.05062	-0.13321	1	0	1								
19	2614	UPT SD NEGERI 11B RUMBA	SD	Negeri	147	9	2	7	2200	1356	70	-5.68162	119.4866	-0.81276	-0.12753	1	0	1								
20	3118	SD NEGERI 268 TOWUTI	SD	Negeri	323	20	3	10	2200	4298	71	-2.58273	121.2079	0.560258	-0.42936	1	0	0.956522								
21	2772	UPT SD NEGERI 6 ARUNGKEKE	SD	Negeri	141	11	1	7	2200	1044	70	-5.79037	119.3422	-0.89525	-0.10183	1	0	1								
22	2154	SD INPRES WATUWATU	SD	Negeri	139	6	2	7	2200	3041	70	-5.25509	119.5789	-0.97098	-0.08338	1	0	1								
23	275	SD ANAK BANGSA	SD	Swasta	0	3	0	0	1342	250	70	-5.14171	119.424	-2.32874	0.199964	1	0	1								
24	1776	SD NEGERI 186 MATTIROWALE	SD	Negeri	160	8	4	7	900	917	70	-5.68097	120.3938	-0.5198	-0.24229	1	0	1								
25	2375	SD INPRES KAMPUNG BERU	SD	Negeri	124	7	1	7	2200	2390	70	-5.39968	119.5112	-1.19514	-0.38615	1	0	1								
26	3867	SD NEGERI 24 ALEKRAJAE	SD	Negeri	157	7	3	8	900	1214	70	-5.19737	119.8933	-0.7291	-0.54372	0	0	1								
		siswa_total_n			jumlah_ruangan_n				wilayah_total_n		daya		listr		jumlah_guru		staf_total		sko		rekomendasi		peringkat			
		0.184789441			0.920454545				0.997940158		0.980442		0.936		0.9375				100		1					
		0.255185418			0.920454545				0.997268841		0.997164		0.832		0.9375				99.999701		2					
		0.209930861			0.920454545				0.995633775		0.997164		0.92		0.9375				98.37484196		3					
		0.150848523			0.920454545				0.999394855		0.980442		0.92		0.90625				97.72314865		4					
		0.126335638			0.909090909				0.998386279		0.99571		0.92		1				97.15484347		5					
		0.161533627			0.886363636				0.997649859		0.980442		0.904		0.9375				96.92312976		6					
		0.094280327			0.954545455				0.999420469		0.987713		0.928		0.9375				96.76266953		7					
		0.051918615			0.931818182				0.998301964		0.997164		0.968		0.96875				96.53009265		8					
		0.110622225			0.920454545				0.997237389		0.987713		0.936		0.96875				96.52009532		9					
		0.117536141			0.920454545				0.996085233		0.997164		0.928		0.9375				96.50613811		10					
		0.121935889			0.909090909				0.998927388		0.99571		0.936		0.9375				96.40973068		11					
		0.108736644			0.920454545				0.997720299		0.997164		0.92		0.96875				96.31630383		12					
		0.117536141			0.920454545				0.997496171		0.992439		0.944		0.875				95.96660775		13					
		0.179132621			0.852272727				0.995486491		0.987713		0.888		0.90625				95.84211671		14					
		0.094908862			0.920454545				0.997331811		0.99571		0.936		0.96875				95.78872331		15					
		0.118793212			0.920454545				0.999328684		0.987713		0.92		0.875				95.76258049		16					
		0.109365179			0.920454545				0.998846275		0.980442		0.92		0.90625				95.52564424		17					
		0.09239472			0.920454545				0.998590129		0.992439		0.928		0.9375				95.21108792		18					
		0.20301697			0.886363636				0.9954550204		0.992439		0.84		0.90625				95.20081999		19					
		0.088623507			0.920454545				0.998923119		0.992439		0.912		0.96875				95.19777665		20					
		0.087938404			0.920454545				0.996791769		0.992439		0.952		0.9375				95.1635013		21					
		0.089680075			0.909090909				0.998741682		0.997164		0.944		0.90625				94.93269837		25					



Gambar 68. Tampilan dataset hasil rekomendasi sekolah untuk bantuan BOP dan Grafik (Peningkatan DBSCAN 3 Cluster)

Berdasarkan data yang dihasilkan terlihat bahwa ketika hasil perangkingan setelah pembobotan berhasil maka kolom PeningkatanDSBCAN\_3klaster pada Gambar 66 berurut dengan teratur dimana ini menggambarkan bahwa **nilai 0** di kolom *PeningkatanDSBCAN\_3klaster* menyatakan kelompok klaster 1 dalam grafik (warna merah) yang paling membutuhkan bantuan BOP, sedangkan **nilai 1** pada kolom *PeningkatanDSBCAN\_3klaster* menunjukkan kelompok klaster 2 dalam grafik (warna kuning) yang tidak terlalu membutuhkan bantuan BOP, dan yang terakhir **nilai 2** pada kolom *PeningkatanDSBCAN\_3klaster* menyatakan kelompok klaster 3 dalam grafik (warna hijau) yang tidak membutuhkan bantuan BOP. Dari hasil perangkingan yang terbentuk menjelaskan bahwa sekolah yang membutuhkan bantuan BOP berada pada klaster yang sama yaitu klaster 1 pada grafik peningkatan DBSCAN (3 klaster) dengan nilai silhouette score tertinggi begitupun dengan sekolah yang tidak terlalu memerlukan bantuan BOP dan sekolah yang tidak membutuhkan bantuan BOP berada pada klaster 2 dan 3 pada grafik peningkatan DBSCAN (3 klaster).

#### **4.11 Uji Validasi Pembobotan Menggunakan Klaster Optimal K-Means Only (2 klaster)**

Selanjutnya akan dilakukan uji validasi pembobotan pada K-Means Only menggunakan klaster terbaik dari K-Means Only dengan *silhouette score* tertinggi yaitu (2 klaster), sama seperti uji validasi pembobotan menggunakan peningkatan DBSCAN. Pada tahap ini dilakukan import library dan menggunakan data dari kmeans\_only\_2cluster.csv

```
[1]: import pandas as pd
import folium
from folium.plugins import MarkerCluster, Search
from folium import FeatureGroup
from folium.plugins import MarkerCluster

[2]: file_path = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_hasil_kmeans_2cluster.csv'
df = pd.read_csv(file_path)
```

Gambar 69. Import library dan pemanggilan dataset kmeans\_only\_2cluster.csv

Selanjutnya, pada Gambar 70 bobot untuk proses perangkingan ditentukan dalam bentuk dictionary, di mana masing-masing kriteria diberi bobot tertentu berdasarkan ketentuan yang telah ditetapkan

```
[3]: # Bobot perangkingan
bobot = {
    'skor_akreditasi': 29,
    'siswa_total': 23,
    'jumlah_ruangan': 19,
    'wilayah_total': 14,
    'daya_listrik': 5,
    'jumlah_guru': 5,
    'staf_total': 5
}
```

Gambar 70. Bobot Perangkingan

Setelah bobot ditetapkan, proses normalisasi dilakukan pada Gambar 71 untuk menghindari ketimpangan akibat skala data yang berbeda. Normalisasi menggunakan rumus Min-Max Scaling, di mana rentang data setiap kolom dikonversi ke skala antara 0 dan 1. Skor akhir dihitung dengan mengalikan setiap kolom yang sudah dinormalisasi dengan bobotnya masing-masing, lalu menjumlahkan semua hasil perkalian. Ini menciptakan satu skor komposit yang merefleksikan performa total sekolah berdasarkan semua kriteria. Lalu sekolah akan diurutkan dari 1-9151.

```
[4]: # Normalisasi data
for col in bobot.keys():
    if col == 'siswa_total':
        df[f'{col}_n'] = (df[col] - df[col].min()) / (df[col].max() - df[col].min())
    else:
        df[f'{col}_n'] = 1 - (df[col] - df[col].min()) / (df[col].max() - df[col].min())

[5]: # Hitung skor akhir
df['skor_rekomendasi'] = (
    df['skor_akreditasi_n'] * bobot['skor_akreditasi'] +
    df['siswa_total_n'] * bobot['siswa_total'] +
    df['jumlah_ruangan_n'] * bobot['jumlah_ruangan'] +
    df['wilayah_total_n'] * bobot['wilayah_total'] +
    df['daya_listrik_n'] * bobot['daya_listrik'] +
    df['jumlah_guru_n'] * bobot['jumlah_guru'] +
    df['staf_total_n'] * bobot['staf_total']
)

[6]: # Normalisasi skor akhir ke skala 100
df['skor_rekomendasi'] = (df['skor_rekomendasi'] - df['skor_rekomendasi'].min()) / (df['skor_rekomendasi'].max() - df['skor_rekomendasi'].min()) * 100

[7]: # Perangkingan
df = df.sort_values(by='skor_rekomendasi', ascending=False).reset_index(drop=True)
df['peringkat'] = df.index + 1
```

Gambar 71. Normalisasi Data dan Pengurutan

Data skor akhir kemudian digabungkan dengan semua kolom asli lalu menampilkan 10 data teratas.

```
# Tampilkan hasil perangkingan K-Means Only
print(hasil_rekomendasi.head(10))

      No           Nama Sekolah   BP Status siswa_total \
0  5260.0        UPT SDN 72 BONTOLOE SD Negeri    294.0
1  2097.0        SD INPRES PENGENTUNGANG UTARA SD Negeri    406.0
2  2242.0        SD INPRES BORONGKALUKU SD Negeri    334.0
3  4072.0        UPT SD NEGERI 149 PINRANG SD Negeri    240.0
4  2227.0        SD NEGERI ANASSAPPU SD Negeri    201.0
5  4475.0        SD NEGERI 4 ARAWA SD Negeri    257.0
6  7163.0        UPTD SMPN 31 SATAP LALANG TEDONG SMP Negeri    150.0
7  3076.0        SD KRISTEN BUNGIN PASANG SD Swasta    146.0
8  2318.0        SD INPRES BIRING PANTING SD Negeri    176.0
9  2671.0        UPT SD NEGERI 11 BANGKALA BARAT SD Negeri    167.0

  jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total ... \
0     8.0       2.0          7.0      5500.0      1965.0 ...
1    21.0       2.0          7.0      900.0      2594.0 ...
2    10.0       2.0          7.0      900.0      4126.0 ...
3    10.0       3.0          7.0      5500.0      602.0 ...
4    10.0       0.0          8.0      1300.0      1547.0 ...
5    12.0       2.0          10.0      5500.0      2237.0 ...
6     9.0       2.0          4.0      3500.0      578.0 ...
7     4.0       1.0          6.0      900.0      1626.0 ...
8     8.0       1.0          7.0      3500.0      2623.0 ...
9     9.0       2.0          7.0      900.0      3703.0 ...

  kmeans_klaster skor_akreditasi_n siswa_total_n jumlah_ruangan_n \
0             0       1.000000  0.184789  0.920455
1             0       0.956522  0.255185  0.920455
2             0       0.956522  0.209931  0.920455
3             0       1.000000  0.150849  0.920455
4             0       1.000000  0.126336  0.900091
5             0       1.000000  0.161534  0.886364
6             0       1.000000  0.094280  0.954545
7             0       1.000000  0.091766  0.931818
8             0       1.000000  0.110622  0.920455
9             0       1.000000  0.117536  0.920455

  wilayah_total_n daya_listrik_n jumlah_guru_n staf_total_n \
0      0.997940  0.980442  0.936  0.93750
1      0.997269  0.997164  0.832  0.93750
2      0.995634  0.997164  0.920  0.93750
3      0.999395  0.980442  0.920  0.98625
4      0.998386  0.995710  0.920  1.00000
5      0.997650  0.980442  0.904  0.93750
6      0.999420  0.987713  0.928  0.93750
7      0.998302  0.997164  0.968  0.96875
8      0.997238  0.987713  0.936  0.96875
9      0.996085  0.997164  0.928  0.93750

  skor_rekomendasi peringkat
0      100.000000      1
1      99.799970      2
2      98.374842      3
3      97.723149      4
4      97.154843      5
5      96.923130      6
6      96.762670      7
```

Gambar 72. Menampilkan 10 Data Teratas Hasil Perangkingan

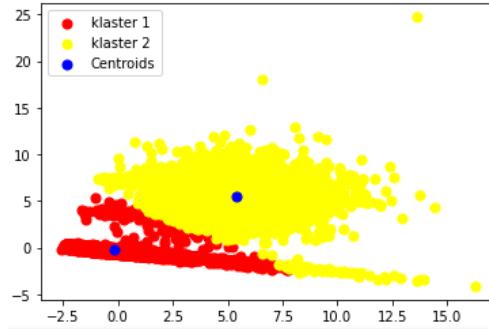
Dalam format CSV maka tampilannya akan seperti Gambar 74 yang didalamnya ada total 9151 baris dengan peringkat 1-9151 yang paling membutuhkan bantuan BOP. Data kemudian disimpan dengan nama Sekolah\_Rekomendasi\_Bantuan\_BOP\_KMeansOnly2Cluster.csv

```
[10]: # Simpan hasil ke file CSV
hasil_rekomendasi.to_csv('Sekolah_Rekomendasi_Bantuan_BOP_KMeansOnly2Cluster.csv', index=False)
```

Gambar 73. Menyimpan dataset dengan nama Sekolah\_Rekomendasi\_Bantuan\_BOP\_KMeansOnly2Cluster.csv

	No	Nama Sekolah	BP	Status	siswa_tot;jumlah_gi	staf_total	jumlah_ru	daya_listr	wilayah_t	skor_akre	Lintang	Bujur	pca1	pca2	kmeans_only_2kластер	
2	5260	UPT SDN 72 BONTOLOE	SD	Negeri	294	8	2	7	5500	1965	70	-5.44719	119.4828	-0.43201	-0.12744	0
3	2097	SD INPRES PENGENTUNGANG UTARA	SD	Negeri	406	21	2	7	900	2594	71	-5.34926	119.7366	0.43507	-0.43036	0
4	224	SD INPRES BORONGKALUKU	SD	Negeri	334	10	2	7	900	4126	71	-5.02884	119.913	-0.26239	-0.25791	0
5	4072	UPT SD NEGERI 149 PINRANG	SD	Negeri	240	10	3	7	5500	602	70	-3.55179	120.0475	-0.32856	-0.171	0
6	2227	SD NEGERI ANASSAPPU	SD	Negeri	201	10	0	8	1300	1547	70	-5.03842	120.0675	-0.87068	-0.10704	0
7	4475	SD NEGERI 4 ARAWA	SD	Negeri	257	12	2	10	5500	2237	70	-3.8753	119.8707	-0.19648	-0.54236	0
8	7163	UPT SMPN 31 SATAP LALANG TEDONG	SMP	Negeri	150	9	2	4	3500	576	70	-5.00296	119.5893	-1.01631	-0.05998	0
9	3076	SD KRISTEN BUNGIN PASANG	SD	Swasta	146	4	1	6	900	1626	70	-2.65442	120.4369	-1.36077	-0.37961	0
10	2318	SD INPRES BIRING PANTING	SD	Negeri	176	8	1	7	3500	2623	70	-5.07165	120.0287	-0.92534	-0.05317	0
11	2671	UPT SD NEGERI 11 BANGKALA BARAT	SD	Negeri	187	9	2	7	900	3703	70	-5.56328	119.6796	-0.87728	-0.51428	0
12	3887	SD NEGERI 24 KALIBONE	SD	Negeri	194	8	2	8	1300	1040	70	-5.04391	119.4915	-0.66637	-0.17699	0
13	2759	UPT SD NEGERI 11 TAROWANG	SD	Negeri	173	10	1	7	900	2171	70	-5.68204	119.7447	-0.86613	-0.13055	0
14	2266	SD INPRES KARAMPUANG	SD	Negeri	187	7	4	7	2200	2381	70	-5.37958	119.9648	-0.55216	-0.56166	0
15	2820	SD NEGERI 57 PADANG SAPPA	SD	Negeri	285	14	3	13	3500	4264	70	-3.12477	120.3359	0.045607	-0.36394	0
16	2571	UPT SD NEGERI 16 BANGKALA	SD	Negeri	151	8	1	7	1300	2535	70	-5.57959	119.4093	-1.01342	-0.08689	0
17	5338	UPT SD 183 INPRES KAMPUNG PARANG	SD	Negeri	189	10	4	7	3500	664	70	-5.5477	119.4886	-0.32752	-0.23519	0
18	5128	UPT SD NEGERI 11 BONTOSANRA	SD	Negeri	174	10	3	7	5500	1116	70	-5.45679	119.4994	-0.50602	-0.13321	0
19	2614	UPT SD NGEN 13 RUMBIA	SD	Negeri	147	9	2	7	2200	1356	70	-5.68162	119.4868	-0.81279	-0.12735	0
20	3118	SD NEGERI 268 TOWUTI	SD	Negeri	323	20	3	10	2200	4298	71	-2.58273	121.2074	0.560258	-0.42936	0
21	2727	UPT SD NEGERI 6 ARUNGKEKE	SD	Negeri	141	11	1	7	2200	1044	70	-5.79037	119.3422	-0.89525	-0.10183	0
22	2154	SD INPRES WATUWATU	SD	Negeri	139	6	2	7	2200	3041	70	-5.25059	119.5789	-0.97099	-0.08338	0
23	275	SD ANAN BANGSA	SD	Swasta	0	3	0	0	1342	250	70	-5.14171	119.424	-2.32876	0.199964	0
24	1776	SD NEGERI 186 MATTIROWALIE	SD	Negeri	160	8	4	7	900	917	70	-5.68097	120.3934	-0.5198	-0.24229	0
25	2374	SD INPRES KAMPUNG BERU	SD	Negeri	124	7	1	7	2200	2390	70	-5.39968	119.5112	-1.19514	-0.38615	0
26	3867	SD NEGERI 24 ALEKARAJAE	SD	Negeri	157	7	3	8	900	1214	70	-5.19737	119.3933	-0.7291	-0.54372	0

skor_akre siswa_tot;jumlah_ru:wilayah_t:daya_listr:jumlah_gi:staf_total:skor_reko peringkat
1 0.184789 0.920455 0.99794 0.980442 0.936 0.9375 100 1
0.956522 0.255185 0.920455 0.997269 0.997164 0.832 0.9375 99.79997 2
0.956522 0.209931 0.920455 0.995634 0.997164 0.92 0.9375 98.37484 3
1 0.150849 0.920455 0.999395 0.980442 0.92 0.90625 97.72315 4
1 0.126336 0.909091 0.998266 0.99571 0.92 1 97.15848 5
1 0.161534 0.886364 0.99765 0.980442 0.904 0.9375 96.92313 6
1 0.09428 0.954545 0.99942 0.987713 0.928 0.9375 96.76267 7
0.091766 0.9181818 0.998302 0.997164 0.968 0.96875 96.53009 8
1 0.110622 0.920455 0.997238 0.987713 0.936 0.96875 96.5201 9
1 0.117536 0.920455 0.996085 0.997164 0.928 0.9375 96.50614 10
1 0.121936 0.909091 0.998927 0.99571 0.936 0.9375 96.40973 11
1 0.108737 0.920455 0.99772 0.997164 0.92 0.96875 96.36163 12
1 0.117536 0.920455 0.997496 0.992439 0.944 0.875 95.96607 13
1 0.179133 0.852273 0.995486 0.987713 0.888 0.90625 95.84212 14
0.094909 0.920455 0.997332 0.99571 0.936 0.96875 95.78872 15
1 0.118793 0.920455 0.999329 0.987713 0.92 0.875 95.76258 16
1 0.109365 0.920455 0.998846 0.980442 0.92 0.90625 95.52564 17
0.092395 0.920455 0.99859 0.992439 0.928 0.9375 95.21109 18
0.956522 0.203017 0.886364 0.99545 0.992439 0.84 0.90625 95.20082 19
1 0.088624 0.920455 0.998923 0.992439 0.912 0.96875 95.19778 20
1 0.087366 0.920455 0.996792 0.992439 0.952 0.9375 95.1635 21
1 0 1 0.999771 0.995558 0.976 1 95.14461 22
1 0.100566 0.920455 0.999059 0.997164 0.936 0.875 95.08684 23
1 0.077938 0.920455 0.997487 0.992439 0.944 0.96875 94.95589 24
1 0.098668 0.909091 0.998742 0.997164 0.944 0.90625 94.9327 25



Gambar 74. Tampilan dataset hasil rekomendasi sekolah untuk bantuan BOP dan Grafik (K-Means Only 2 klaster)

Berdasarkan data yang dihasilkan dari pemodelan data kmeans\_only\_2cluster terlihat bahwa ketika hasil perangkingan setelah pembobotan berhasil maka kolom kmeans pada Gambar 74 menggambarkan bahwa **nilai 0** di kolom *kmeans\_only\_2klaster* menyatakan kelompok klaster 1 dalam grafik (warna merah) yang paling membutuhkan bantuan BOP, sedangkan **nilai 1** pada kolom *kmeans\_only\_2klaster* menunjukkan kelompok klaster 2 dalam

grafik (warna kuning) yang tidak membutuhkan bantuan BOP. Dari hasil perangkingan yang terbentuk menjelaskan bahwa sekolah yang membutuhkan bantuan BOP berada pada klaster yang sama yaitu klaster 1 pada grafik K-Means Only (2 klaster) dengan nilai silhouette score tertinggi begitupun dengan sekolah yang tidak memerlukan bantuan BOP dan sekolah yang tidak membutuhkan bantuan BOP berada pada klaster 2.

## 4.12 Perbandingan Uji Validasi Pembobotan Klaster Optimal Peningkatan

### DBSCAN (3 klaster) dan K-Means Only (2 klaster)

Dari hasil perbandingan uji validasi ditemukan beberapa perbandingan antara uji validasi pembobotan antara peningkatan DBSCAN (3 klaster) dan K-Means Only (2 klaster). Pada data awal semuanya terlihat sama dengan menunjukkan perangkingan sekolah yang sama namun ketika di teliti lebih lanjut dari hasil perbandingan data antara

*Sekolah\_Rekomendasi\_Bantuan\_BOP\_PeningkatanDSBCAN3Cluster.csv* dan *kmeans\_only\_2klaster* ditemukan di tengah-tengah hasil pembobotan perbedaan dari uji validasi pembobotan pada 2 algoritma ini dimana pada hasil dari pembobotan K-Means Only memasukkan sekolah yang tidak terlalu membutuhkan BOP sebanyak 178 sekolah dan sekolah yang tidak membutuhkan BOP sebanyak 32 sekolah ke dalam sekolah yang membutuhkan bantuan BOP. Selain itu klaster hasil pembobotan K-Means Only masih tidak tersusun rapuh dan hal tersebut berbeda dengan hasil dari pembobotan peningkatan DBSCAN (3 klaster) lebih rapih serta terstruktur sehingga hasil pembobotan peningkatan DBSCAN akan digunakan untuk pembuatan peta interaktif karena hasil datanya lebih valid. **Kolom P** pada K-Means Only (2 klaster) menunjukkan nilai dari klaster yang terbentuk sedangkan **kolom Q** menunjukkan nilai dari klaster yang terbentuk pada peningkatan DBSCAN (3 klaster). Pada Gambar 75 terlihat bahwa perbedaan perangkingan sekolah pada kolom 2294-2472 pada hasil pembobotan.

K-Means Only 2 klaster															Formula Bar	N	O	P
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P			
2291	2123 SD ISLAM TERPADU AZZAHRAH	SD	Swasta	182	18	3	13	2200	2456	77	-5.16534	119.7082	0.378823	-0.36116	0			
2292	3843 SD NEGERI 3B BUTTIE	SD	Negeri	39	3	1	2	900	1426	78	-5.28222	119.4698	-1.82943	0.122192	0			
2293	7404 SMP NEGERI 2 WALENRANG	SMP	Negeri	65	10	17	11	1300	3985	73	-2.5995	120.3372	1.70401	-0.85678	1			
2294	582 SD NEGERI 18 EREMERASA	SD	Negeri	58	9	3	9	3500	3874	75	-5.5193	119.9622	-0.67509	-0.10262	1			
2295	9031 SMKN 8 JENEPONTO	SMK	Negeri	454	42	8	26	1300	3582	75	-5.675	119.7334	3.397537	-1.23993	1			
2296	2091 SD INGRES MANGASAG	SD	Negeri	315	14	3	8	2200	2295	78	-5.1719	119.6246	0.066556	-0.35126	1			
2297	2458 SD INGRES JENEMADINGING	SD	Negeri	110	8	3	7	3500	2012	76	-5.04278	119.5105	-0.81166	-0.1234	1			
2298	2745 UPT SD NEGERI 22 KELARA	SD	Negeri	114	9	2	8	1300	2056	76	-5.60399	119.4117	-0.85804	-0.14751	1			
2299	4193 UPT SD NEGERI 126 PINRANG	SD	Negeri	118	10	3	7	3500	4769	76	-3.22016	119.6635	-0.71907	-0.14713	1			
2300	1182 SD NEGERI 184 POLEWALI	SD	Negeri	80	6	1	7	1300	2346	76	-4.84701	119.5857	-1.34199	-0.03773	1			
2301	2296 SD INGRES JONGGOWA	SD	Negeri	69	6	0	7	2200	1116	76	-5.23716	119.8537	-1.5215	0.031521	1			
2302	5992 SD NEGERI 269 MANNAGAE	SD	Negeri	87	7	1	7	3500	3835	76	-4.0704	119.6247	-1.14277	0.026351	1			
2303	2651 UPT SD NEGERI 20 TAMILATEA	SD	Negeri	114	9	3	7	3500	1628	76	-5.58722	119.535	-0.75866	-0.09958	1			
2304	7998 SMP ALMUTTAQIEM	SMP	Swasta	41	3	1	11	1080	345	75	-5.1319	119.4047	-1.26708	-0.40425	1			
2305	7210 SMP IT TIGO PUTERA	SMP	Swasta	43	3	1	5	1300	4369	76	-5.01049	119.5781	-1.61287	0.085889	1			
2306	618 SD NEGERI 24 PANDANGPANDANG	SD	Negeri	82	11	5	9	5500	703	75	-5.518	119.9498	-0.18644	-0.20058	1			
2307	470 UPT SPF SD NEGERI MELAYU	SD	Negeri	257	15	3	10	2200	1056	77	-5.123	119.4165	0.139615	-0.71058	1			
2308	1798 SD NEGERI TERPENCIL 350 KAHAYA	SD	Negeri	111	9	3	7	2200	3633	76	-5.4839	120.332	-0.79568	-0.15821	1			
2309	5674 SD NEGERI 8 SESEAN	SD	Negeri	147	11	3	9	3500	3032	76	-2.96616	119.8917	-0.33329	-0.17654	1			
2310	6444 UPTD SD NEGERI 3 PAREPARE	SD	Negeri	584	32	5	21	5500	4372	78	-3.97514	119.6416	3.05678	-0.88336	1			
2311	2870 SD NEGERI 437 KARIAKO	SD	Negeri	111	9	2	8	1300	1523	76	-3.06849	121.2383	-0.77446	-0.11733	1			
2312	7704 SMP NEGERI 5 ALLA	SMP	Negeri	167	22	3	14	5500	3251	75	-3.56776	119.7654	0.604354	-0.30503	1			
2313	2620 UPT SD NEGERI 19 RUMBIA	SD	Negeri	80	11	1	6	5500	854	76	-5.75536	119.5537	-1.12602	0.009155	1			
2314	16 UPT SPF SD INGRES PAI 1	SD	Negeri	461	24	5	9	1300	1600	79	-5.07902	119.4934	2.3960197	6.413879	1			
2315	6257 UPT SD KAMPUNG TANGNGA NO 108 KEPULSD	SD	Negeri	111	12	2	7	3500	6057	76	-6.36899	120.9824	-0.68791	-0.07683	1			
2316	6057 SD NEGERI 365 AJURAJA	SD	Negeri	91	7	2	7	900	3377	76	-3.94754	119.8252	-1.06954	-0.42023	1			

Peningkatan DBSCAN 3 Klaster														
			H	I	J	K	L	M	N	O	P	Q	R	S
2293	1111 SD NEGERI 49 LAPPO ASE	SD	Negeri	69	8	3	7	5500	2704	77	-4.41356	120.3601	-0.79337	-0.02223
2294	2823 SD NEGERI 61 MARIO	SD	Negeri	285	18	4	13	5500	3400	78	-3.00012	121.0058	0.856948	-0.38806
2295	2946 SD NEGERI 352 TOBEMBA	SD	Negeri	146	6	3	7	3500	799	78	-2.56674	121.4763	-0.81693	-0.11922
2296	2132 SD INPRER BELAKA	SD	Negeri	393	18	2	10	5500	3645	80	-5.04018	119.6916	0.631352	-0.30531
2297	1692 SD NEGERI 101 KAJANG II	SD	Negeri	152	8	3	7	3500	1385	78	-5.34698	120.3034	-0.78817	-0.52715
2298	1467 SD NEGERI 147 BULU ALLAPPO SD	SD	Negeri	65	7	3	7	5500	2154	77	-4.45324	119.5633	-0.85684	-0.01252
2299	4779 SD Neg No 93 Timbasoang	SD	Negeri	63	8	3	7	1300	2801	77	-5.11513	120.2624	-0.85437	-0.10575
2300	7881 SMPN SATAPP4 LAMURU	SMP	Negeri	25	9	2	5	900	3219	77	-5.46464	120.2125	-1.22096	-0.07279
2301	6860 UPT SMP NEGERI 6 POLONGBA SMP	SMP	Negeri	40	10	3	5	2200	3123	77	-5.46064	119.4917	-0.96525	-0.066901
2302	7524 UPT SMP NEGERI 2 ARUNGKEK SMP	SMP	Negeri	179	19	4	12	2200	1982	77	-5.67628	119.7432	0.499844	-0.404
2303	2202 SD Negeri Lonrong	SD	Negeri	217	7	3	6	900	1187	79	-5.42657	119.9349	-0.72769	-0.55034
2304	7664 SMPN SATAPP5 PONRE	SMP	Negeri	31	9	4	4	2200	1534	77	-5.45251	120.2175	-0.95253	-0.08735
2305	3784 SD ISLAM TERPADU ABU BAKR SD	Swasta	Swasta	658	30	2	26	2200	1858	80	-4.84014	119.7598	2.326145	-0.85138
2306	6340 SD ISLAM TERPADU DARUSSALAM SD	Swasta	Swasta	636	36	2	29	2200	6021	79	-2.91039	120.2556	3.484013	-0.99624
2307	6869 SMP NEGERI 3 WATANOPPEN SMP	SMP	Negeri	288	23	9	21	1300	840	76	-4.35809	119.9001	2.41469	-0.90485
2308	6032 SD NEGERI 88 WALENNAE	SD	Negeri	30	7	0	7	3500	3439	77	-4.08656	119.9494	-1.44203	0.110663
2309	1844 SD NEGERI 197 SAPOLOH	SD	Negeri	227	14	4	10	5500	3358	78	-5.22459	120.2875	0.29401	-0.26565
2310	7439 SMP NEGERI 4 BASTEM	SMP	Negeri	71	9	2	8	2200	4057	77	-2.60099	120.3266	-0.9336	-0.41573
2311	6428 UPTD SD NEGERI 84 PAREPARE SD	SD	Negeri	191	11	3	9	3500	4032	78	-3.99335	119.6571	-0.19147	-0.19161
2312	5389 UPT SDN 12 MAKALE	SD	Negeri	135	11	2	12	3500	1415	77	-3.06214	119.8393	-0.28975	-0.16915
2313	3994 SD NEGERI 10 BONTO	SD	Negeri	70	7	4	7	900	4133	77	-5.30341	119.8332	-0.86355	-0.19864
2314	9149 SMKS KEPERAWATAN HARAPAN SMK	Swasta	Swasta	17	0	1	7	2199	1444	77	-5.16268	119.4542	-0.11734	7.226452
2315	1787 SD NEGERI 47 ANNIHUA	SD	Negeri	141	8	2	7	3500	3647	78	-5.2488	120.1356	-0.77077	-0.05479
2316	134 SD MUHAMMADIAH 2 MAMA SD	Swasta	Swasta	70	7	3	8	900	520	77	-5.1589	119.4176	-0.89192	-0.51265
2317	6331 SD MUHAMMADIAH II	SD	Swasta	161	9	4	7	2200	1373	78	-2.88308	120.2526	-0.36831	-0.2025
2318	4178 UPT SD NEGERI 47 PINRANG	SD	Negeri	94	11	4	8	1300	2588	77	-3.31338	120.0154	-0.40173	-0.22234

Gambar 75. Perbandingan hasil pembobotan K-Means Only (2 klaster) & Peningkatan DBSCAN (3 klaster)

#### 4.13 Pembuatan Peta Interaktif Menggunakan Pustaka Folium

Setelah melakukan uji validasi perangkingan sekolah untuk rekomendasi BOP, maka selanjutnya dilakukan pembuatan peta Folium pada hasil data pembobotan peningkatan DBSCAN 3 klaster karena mempunyai data yang terstruktur baik dan valid serta skor silhouette yang tinggi. Peta ini difokuskan ke wilayah Indonesia (koordinat -2.5489, 118.0149) dan memiliki zoom awal 8, agar bisa melihat seluruh peta secara luas. Untuk menghindari kepadatan marker, fitur MarkerCluster ditambahkan, yang mengelompokkan sekolah-sekolah terdekat menjadi satu titik saat peta masih dalam tampilan jauh. Ini membantu mengurangi kecacuan visual.

```
# Buat peta menggunakan Folium
map = folium.Map(location=[-2.5489, 118.0149], zoom_start=8)
marker_cluster = MarkerCluster().add_to(map)
```

Gambar 76. Pembuatan peta interaktif menggunakan pustaka Folium

Setiap sekolah diwakili oleh sebuah marker yang diletakkan berdasarkan kolom Lintang dan Bujur. Klik pada marker akan memunculkan pop-up yang berisi detail sekolah, termasuk nama sekolah, peringkat, skor akreditasi, jumlah siswa, luas wilayah, jumlah ruangan, total staf, total guru, dan daya listrik. Marker diberi warna merah untuk sekolah yang paling membutuhkan BOP, kuning untuk sekolah yang tidak terlalu membutuhkan BOP, sedangkan sisanya berwarna hijau tidak membutuhkan rekomendasi BOP, sehingga sekolah yang paling direkomendasikan terlihat jelas di peta.

```
[59]: for idx, row in hasil_rekomendasi.iterrows():
    # Lewati jika Lintang atau Bujur kosong
    if pd.isna(row['Lintang']) or pd.isna(row['Bujur']):
        continue

    popup_info = (
        f"Nama Sekolah: {row['Nama Sekolah']}<br>"
        f"Peringkat Rekomendasi: {row['peringkat']}<br>"
        f"Skor Akreditasi: {row['skor_akreditasi']}<br>"
        f"Jumlah Siswa: {row['siswa_total']}<br>"
        f"Luas Wilayah: {row['wilayah_total']}<br>"
        f"Jumlah Ruangan: {row['jumlah_ruangan']}<br>"
        f"Total Staf: {row['staf_total']}<br>"
        f"Total Guru: {row['jumlah_guru']}<br>"
        f"Daya Listrik: {row['daya_listrik']}<br>"
    )

    # Tentukan warna marker berdasarkan index
    if idx <= 2290:
        warna = 'red'
    elif idx <= 6021:
        warna = 'orange'
    else:
        warna = 'green'

    marker = folium.Marker(
        location=[row['Lintang'], row['Bujur']],
        popup=popup_info,
        icon=folium.Icon(color=warna)
    )
    jenjang = row['BP'] if row['BP'] in jenjang_groups else 'SD'
    marker.add_to(jenjang_groups[jenjang])
```

Gambar 77. Memasukkan marker sekolah ke peta

Selanjutnya, fitur pencarian ditambahkan menggunakan fungsi Search dari folium.plugins. Ini memungkinkan pengguna mengetik nama sekolah tertentu dan peringkat sekolahnya, dan peta secara otomatis akan melompat ke lokasi marker sekolah tersebut.

```
[58]: # Tambahkan fitur pencarian Nama Sekolah
Search(
    layer=marker_cluster,
    search_label='Nama Sekolah',
    placeholder='Cari sekolah...',
    collapsed=False
).add_to(map)

[58]: <folium.plugins.search.Search at 0x1b13a16d220>

[59]: Search(
    layer=marker_cluster,
    search_label='peringkat',
    placeholder='Masukkan Peringkat Sekolah',
    collapsed=False
).add_to(map)

[59]: <folium.plugins.search.Search at 0x1b13da5fb20>
```

Gambar 78. Fitur pencarian sekolah dan pencarian peringkat sekolah

Selain pencarian, peta juga memiliki fitur filter jenjang pendidikan. Sekolah dibagi menjadi empat layer berdasarkan jenjangnya: SD, SMP, SMA, dan SMK. Masing-masing layer dibuat menggunakan FeatureGroup dan ditambahkan ke peta. Dengan demikian, pengguna bisa memilih untuk menampilkan hanya satu jenjang atau mengombinasikan beberapa jenjang sekaligus, sesuai kebutuhan mereka. Tahap akhir peta disimpan sebagai file HTML dengan nama Peta\_FIX\_Rekomendasi\_Sekolah\_Sulsel.html. File ini bisa langsung dibuka di browser untuk melihat peta interaktif yang sudah mencakup semua fitur dari pencarian, klaster marker, filter jenjang, hingga tampilan pop-up informasi sekolah. Pada Gambar 80 terlihat tampilan hasil peta yang dimana hasilnya berdasarkan hasil pembobotan perangkingan peningkatan DBSCAN 3 klaster sehingga warna petanya ada 3 sesuai dengan grafik dari algoritma peningkatan DSBCAN 3 klaster yaitu merah untuk klaster 1 (yang butuh bantuan BOP), kuning (tidak terlalu membutuhkan bantuan BOP), dan hijau (tidak membutuhkan bantuan BOP).

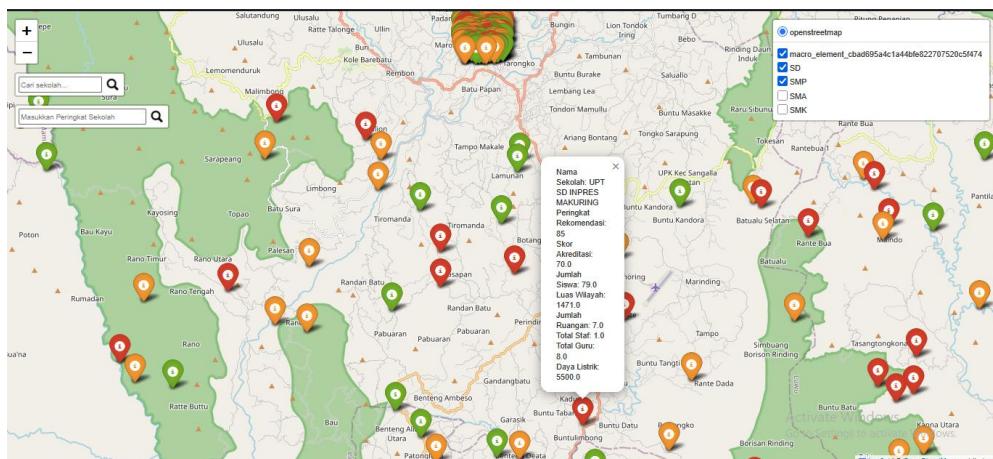
```
[22]: # Tambahkan kontrol Layer untuk filter jenjang
folium.LayerControl().add_to(map)

[22]: <folium.map.LayerControl at 0xleab96d5fd0>

[23]: # Simpan peta ke file HTML
map.save('Peta_FIX_Rekomendasi_Sekolah_Sulsel.html')
print("Peta Rekomendasi telah disimpan sebagai 'Peta_FIX_Rekomendasi_Sekolah_Sulsel.html'")

Peta Rekomendasi telah disimpan sebagai 'Peta FIX Rekomendasi Sekolah Sulsel.html'
```

Gambar 79. Menampilkan fitur filter berbentuk layer dengan 4 layer yaitu SD,SMP,SMA,dan SMK



Gambar 80. Tampilan hasil peta interaktif menggunakan pustaka Folium

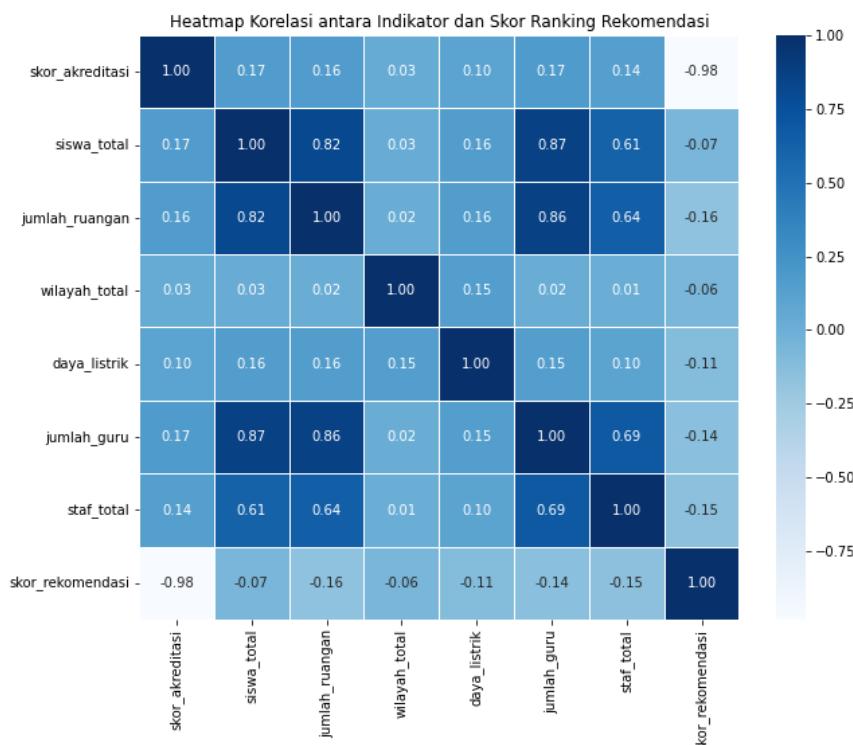
Pada Gambar 81 terlihat pada heatmap bahwa ketika memperhatikan heatmap korelasinya, terlihat bagaimana setiap angka memiliki makna tersendiri. Skor akreditasi tampak memiliki hubungan negatif yang kuat dengan skor perangkingan. Artinya, semakin rendah skor akreditasi suatu sekolah, semakin besar peluang mereka memperoleh skor akhir yang tinggi hal ini sejalan dengan prinsip bahwa sekolah-sekolah yang tertinggal umumnya menjadi prioritas untuk menerima bantuan.

Di sisi lain, jumlah siswa justru menunjukkan korelasi positif yang cukup signifikan. Ini menggambarkan bahwa semakin banyak jumlah siswa dalam sebuah sekolah, semakin besar dorongan untuk meningkatkan peringkatnya. Wajar saja, sekolah dengan populasi siswa yang besar namun sumber daya terbatas tentu menjadi perhatian utama.

Jumlah ruangan dan luas wilayah total juga memperlihatkan korelasi negatif, meskipun tidak sekuat akreditasi. Sekolah dengan ruangan yang terbatas dan lahan yang kecil cenderung memiliki skor perangkingan yang lebih tinggi, menegaskan bahwa keterbatasan fasilitas menjadi salah satu faktor penting dalam penentuan prioritas.

Menariknya, daya listrik, jumlah guru, dan staf total memperlihatkan korelasi negatif yang lebih lemah. Meski bobotnya kecil, sekolah dengan tenaga pengajar dan staf yang terbatas tetap memiliki pengaruh terhadap skor akhir mereka.

Setelah dilakukan perangkingan, **UPT SDN 72 BONTOLOE** keluar sebagai peringkat 1 untuk menerima rekomendasi bantuan, dapat disimpulkan bahwa perpaduan antara skor akreditasi yang rendah, jumlah siswa yang tinggi, serta minimnya fasilitas dan sumber daya secara konsisten mendorong mereka ke peringkat teratas. Heatmap ini, meskipun berupa data numerik, secara tidak langsung mencerminkan kondisi nyata di lapangan bagaimana sekolah-sekolah ini bertahan dan berjuang di tengah segala keterbatasan mereka.



Gambar 81. Heatmap korelasi antara indikator dan skor rangking rekomendasi

## BAB 4

### KESIMPULAN DAN SARAN

#### 4.1 Kesimpulan

Berdasarkan rumusan masalah yang di analisis pada penelitian ini dapat disimpulkan bahwa :

1. Penerapan algoritma K-Means Only mampu mengelompokkan sekolah di Sulawesi Selatan berdasarkan parameter luas wilayah, akreditasi, daya listrik, jumlah ruangan, jumlah staf, jumlah guru, jumlah siswa, serta koordinat geografis (lintang dan bujur). Sedangkan pada algoritma peningkatan DBSCAN clustering, yang menggunakan bantuan K-Means dan K-Nearest Neighbors untuk dapat menentukan parameter eps (epsilon), terbukti sangat baik dalam mengidentifikasi klaster dengan beragam berbentuk sehingga sangat efektif serta baik dalam menemukan outlier/noise yang ada dalam dataset sehingga skornya lebih baik dari algoritma K-Means Only.
2. Hasil dari kedua algoritma ini berjalan dengan baik pada penelitian ini dan dapat memberikan implikasi nyata bagi proses penyaluran BOP. Terdapat beberapa kelompok sekolah yang memiliki akreditasi rendah, jumlah siswa yang banyak, jumlah ruangan yang sedikit, dan berada di wilayah kecil. Untuk mendapatkan rekomendasi bantuan BOP, kelompok ini menjadi prioritas utama. Dinas Pendidikan Sulawesi Selatan memiliki dasar yang lebih kuat untuk menentukan target sekolah yang membutuhkan bantuan, yang memungkinkan alokasi dana BOP dilakukan secara lebih merata dan tepat sasaran. Hasil pemodelan dari algoritma peningkatan DBSCAN lebih baik dari K-Means Only sehingga data yang digunakan berdasarkan pemodelan dari peningkatan DBSCAN karena tingkat akurasi untuk mengelompokkan sekolah rekomendasi untuk mendapatkan bantuan BOP di Sulawesi Selatan lebih tinggi.
3. Hasil akhir evaluasi dan uji validasi menggunakan *Silhouette Score* menunjukkan hasil perbandingan yang signifikan bahwa dibandingkan dengan K-Means Only, metode peningkatan DBSCAN mendeteksi klaster yang tidak berbentuk (non-spherical) dan menangani noise dengan lebih baik. Namun, algoritma K-Means dapat membuat kinerja lebih cepat khususnya dalam proses pengelompokan. Perbedaan signifikan terlihat dari hasil pengelompokan kedua algoritma ini, di mana DBSCAN mampu mengeliminasi data sekolah yang memiliki karakteristik jauh berbeda dari kriteria BOP, sedangkan K-Means cenderung memasukkan seluruh data ke dalam klaster meskipun memiliki jarak yang jauh dari centroid. Sekolah dengan skor akreditasi rendah, jumlah siswa yang banyak, ruang kelas sedikit, luas wilayah yang kecil, jumlah guru yang sedikit, daya listrik rendah, dan staf yang sedikit berada di klaster yang sama dan cenderung terkumpul menjadi satu. Kelompok klaster ini yang menjadi prioritas utama untuk rekomendasi sekolah penerima bantuan BOP.

4. Performa algoritma K-Means dan peningkatan DBSCAN dalam mengelompokkan data sekolah di Sulawesi Selatan untuk penyaluran Bantuan Operasional Pendidikan (BOP) menunjukkan perbedaan yang cukup signifikan. K-Means bekerja dengan membagi sekolah ke dalam kelompok berdasarkan kedekatan data terhadap titik pusat atau centroid yang telah ditentukan. Metode ini cukup efisien dalam membentuk klaster pada data yang memiliki struktur teratur dan berbentuk bulat. Namun, K-Means kurang optimal dalam menangkap pola non-linear dan sangat sensitif terhadap outlier. Hal ini menyebabkan kemungkinan beberapa sekolah yang sebenarnya sangat membutuhkan bantuan tidak masuk dalam kelompok prioritas, sehingga distribusi BOP menjadi kurang tepat sasaran. Sementara itu, peningkatan DBSCAN mampu mengelompokkan sekolah dengan lebih fleksibel karena algoritma ini bekerja berdasarkan kepadatan data. Peningkatan DBSCAN tidak memerlukan jumlah klaster yang ditentukan sebelumnya dan dapat menangani data dengan distribusi yang tidak beraturan. Selain itu, algoritma ini memiliki keunggulan dalam mendeteksi outlier, yang berarti sekolah-sekolah yang tidak memenuhi kriteria penerima BOP dapat diidentifikasi dan dipisahkan dengan lebih akurat. Evaluasi menggunakan *Silhouette Score* menunjukkan bahwa peningkatan DBSCAN memiliki skor yang lebih tinggi dibandingkan K-Means, yang menandakan kualitas pengelompokan yang lebih baik.

Dari hasil penelitian, dapat disimpulkan bahwa peningkatan DBSCAN lebih direkomendasikan dalam konteks penyaluran BOP karena kemampuannya dalam menangani variasi kepadatan data serta mendeteksi sekolah yang benar-benar membutuhkan bantuan. K-Means memang memiliki keunggulan dalam efisiensi komputasi, tetapi kurang cocok untuk data sekolah yang memiliki distribusi tidak merata. Dengan menggunakan DBSCAN yang telah ditingkatkan, proses pengelompokan sekolah dapat dilakukan dengan lebih akurat.

## 4.2 Saran

Berikut beberapa saran untuk pengembangan penelitian selanjutnya:

1. Mengkaji ataupun mengkombinasikan algoritma lain untuk dapat meningkatkan performa pengelompokan.
2. Mengimplementasikan hasil pengelompokan/clustering langsung ke sistem informasi Dinas Pendidikan Sulawesi Selatan dalam mendukung keputusan penyaluran bantuan BOP.
3. Membangun dan mengembangkan pemodelan prediksi kebutuhan bantuan BOP berbasis machine learning sehingga penyaluran dananya lebih efektif dan proaktif.

## DAFTAR PUSTAKA

- Wei. L, Tang. W, Huang. H, Chen. T. (2021). Research and Application of Intersection Clustering Algorithm Based on PCA Feature Extraction and K-Means. IOP Publishing.  
[https://www.researchgate.net/publication/350774378\\_Research\\_and\\_Application\\_of\\_Intersection\\_Clustering\\_Algorithm\\_Based\\_on\\_PCA\\_Feature\\_Extraction\\_and\\_K-Means](https://www.researchgate.net/publication/350774378_Research_and_Application_of_Intersection_Clustering_Algorithm_Based_on_PCA_Feature_Extraction_and_K-Means).
- Adhikari, A dam Sen, S. (2023). RECENT TRENDS OF CLUSTER ANALYSIS IN EDUCATION. International Research Journal of Modernization in Engineering Technology and Science IRJMETS. Vol. 5, 1858-1861. Jain, A. K. (2020). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 131, 1-16.
- Tandy, J., & Assegaff, S. (2019). Analisis Dan Perancangan Clustering Siswa Baru Menggunakan Metode K-Means. Manajemen Sistem Informasi, 4(4), 389–399.
- Alawi, S. J. S., Shaharanee, I. N. M., & Jamil, J. M. (2023). Clustering student performance data using k-means algorithms. Journal of Computational Innovation and Analytics, 2(1), 41-55.  
<https://doi.org/10.32890/jcia2023.2.1.3>
- Sutramiani. P, Teguh. M , Fane'a. P, Aurelia. S, Fauzi, M , Agus. W. (2024). The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover.
- Lloyd, S. P. (2021). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129-137.
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (2020). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 226-231.
- Zhu, X., Li, M., & Gao, Q. (2022). Clustering-based analysis of gene expression data for identifying disease-related genes. Bioinformatics, 38(3), 456-463.
- Hassan, M., Karim, A., & Ali, S. (2022). Community detection in social networks using clustering algorithms: A survey. International Journal of Computer Applications, 178(10), 1-6.
- Wang, T., & Xie, L. (2023). Clustering for face recognition: A review. Journal of Visual Communication and Image Representation, 83, 103196.
- Arora, A., Kumar, P., & Rastogi, R. (2021). A study of scalable clustering algorithms for big data. Springer.

- Schubert, E., & Rousseeuw, P. J. (2021). Understanding DBSCAN and its applications. *Data Mining and Knowledge Discovery*, 35(2), 467-485.
- R. Kumar & A. Sharma, "An Efficient Survey of K-Means Clustering Algorithm and Its Applications," *Journal of Machine Learning and Data Science*, vol. 7, no. 4, pp. 56-62, 2021.
- H. Lee & S. Park, "Improved K-Means Clustering for High-Dimensional Data: A Review," *Journal of Data Analysis and Applications*, vol. 9, no. 3, pp. 211-220, 2022.
- M. Gupta & R. Desai, "Evaluating K-Means Clustering Performance in Multi-Class Data Clustering," *International Journal of Computational Intelligence*, vol. 15, no. 2, pp. 135-141, 2023.
- L. Wang & Z. Yang, "A Hybrid Approach to K-Means Clustering for Complex Data," *Advanced Machine Learning and Data Mining Applications*, vol. 11, no. 1, pp. 102-110, 2024.
- A. Sharma & S. Raj, "Efficient Outlier Detection Using DBSCAN for Complex Data Structures," *International Journal of Advanced Data Analytics*, vol. 10, no. 1, pp. 55-64, 2023.
- Han, J., Kamber, M., & Pei, J. (2021). *Data Mining: Concepts and Techniques* (4th ed.). Elsevier.
- Indahsari, G., Fatah, Z., & Saintek, G. (2023). Penerapan Data Mining Clustering K-Means Dalam Pengelompokan Penduduk Penyandang Disabilitas di Jawa Timur. *GJMI (Global Journal of Management and Information Technology)*, 13(1), 1-8.
- Tarigan, P. M. S., Hardinata, J. T., Qurniawan, H., Safii, M., & Winanjaya, R. (2022). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang: Studi Kasus: Toko Sinar Harahap. *Jurnal Janitra Informatika dan Sistem Informasi*, 2(1), 9-19. <https://jurnal.umj.ac.id/index.php/just-it/article/download/11375/7258>
- Rias, C. N., Riandianto, R., & Purwono, P. (2023). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer dan Informatika*, 8(2), 168-180.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang. *Indonesian Journal on Computer and Information Technology*, 5(2), 127-136.
- Fahmi, M. N. (2023). Implementasi Machine Learning menggunakan Python Library: Scikit-Learn (Supervised dan Unsupervised Learning). *Sains Data: Jurnal Studi Matematika dan Teknologi*, 1(2), 87-96.

- Jolliffe, I. T., & Cadima, J. (2019). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 377(2130), 20190010. <https://doi.org/10.1098/rsta.2019.0010>
- Zhang, W., & Zhang, Y. (2022). An improved PCA-based method for feature extraction in high-dimensional data analysis. *Applied Sciences*, 12(1), 98. <https://doi.org/10.3390/app12010098>
- Khairati, A. F., Adlina, A. A., Hertono, G. F., & Handari, B. D. (2019). Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means Maximum Minimum Criterion Algorithm. *PRISMA*, 2, 161-170.
- Cutting, V., & Stephen, N. (2021). A Review on Using Python as a Preferred Programming Language for Beginners. *International Research Journal of Engineering and Technology (IRJET)*, 8(8), 4258-4262.
- Barlian, U. C., Erawan, R. D. T., Ardiana, L., & Yuniawati. (2023). *Analisis Kebijakan Bantuan Operasional Pendidikan Daerah (BOPD) Pemerintah Provinsi Jawa Barat dalam Meningkatkan Mutu Pendidikan di SMA Negeri 1 Purwakarta*. *ARMADA: Jurnal Penelitian Multidisiplin*, 1(8), 832-837.

## LAMPIRAN

Lampiran 1 Surat permohonan pengambilan data 5 sekolah yang belum masuk di dataset



**PEMERINTAH PROVINSI SULAWESI SELATAN  
DINAS PENDIDIKAN**  
Jl. Perintis Kemerdekaan Km. 10 Tamalanrea Telepon 586083, Fax 585257  
**MAKASSAR 90245**

Makassar, 18 Februari 2025

Nomor : 000.9.2/413/DISDIK  
Sifat : Biasa  
Lampiran : -  
Hal : Persetujuan Pengambilan Data

Yth . Kepala UPT SMA/SMK se-Sulawesi Selatan

di  
Tempat

Menindaklanjuti surat Wakil Dekan Bidang Akademik, Riset dan Inovasi Fakultas Teknik Universitas Hasanuddin nomor 205/UN4.8.7.1/PT.01.02/2025 tanggal 10 Februari 2025 perihal Pengambilan Data/Penelitian Mahasiswa. Sehubungan dengan hal tersebut, Dinas Pendidikan Provinsi Sulawesi Selatan menyetujui Mahasiswa di bawah ini :

Nama : Muhammad Zulfikar Aris  
NIM : D121191064  
Judul : Analisis Perbandingan Performa Algoritma K-Means dan Peningkatan DBSCAN pada Clustering data Pengelompokan Sekolah di Sulawesi Selatan Untuk Penyaluran Merata BOP (Bantuan Operasional Sekolah)  
(Studi Kasus : Dinas Pendidikan Sulawesi Selatan)

Untuk melakukan pengambilan data dalam rangka untuk penyelesaian tugas akhir Mahasiswa Program Studi Teknik Informatika, Fakultas Teknik Universitas Hasanuddin, sepanjang tidak bertentangan dengan ketentuan dan peraturan Perundang-undangan yang berlaku.

Demikian disampaikan atas perhatian dan kerjasamanya diucapkan terima kasih.

KEPALA DINAS,



**IQBAL NADJAMUDDIN**  
Pangkat/Gol : Pembina Utama Muda/IV/c  
NIP. 197407012002121003

Tembusan :

1. Kepala Cabang Dinas Pendidikan Wilayah I sd XII di Tempat:
2. Dekan Fakultas Teknik Universitas Hasanuddin di Makassar:

## Lampiran 2 Source code program

### a. Data Preprocessing

```
[1]: #Pembersihan Data
import pandas as pd

[7]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Data/data_gabungan_sekolah_sulsel.csv'
df = pd.read_csv(filelocation)

[8]: df.head()

[8]:   No      Nama Sekolah  BP Status siswa_total jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total skor_akreditasi
0  1    UPT SPF SD INPRES BADDOKA  SD Negeri     337.0       18       3        13.0  1300 watt     2808 m2          81
1  2    UPT SPF SD INPRES BAKUNG 1  SD Negeri     243.0       15       4        8.0   900 watt     3000 m2          84
2  3    UPT SPF SD INPRES BAKUNG 2  SD Negeri     269.0       16       5        8.0  4400 watt     2500 m2          84
3  4    UPT SPF SD INPRES DAYA  SD Negeri     498.0       28       3        13.0  4640 watt     1200 m2          87
4  5    UPT SPF SD INPRES KALANG TUBUNG 2  SD Negeri     195.0       15       1        8.0  1300 watt     1505 m2          81
```

---

```
[25]: columns_to_clean = ['jumlah_ruangan', 'daya_listrik', 'wilayah_total', 'Lintang', 'Bujur']

def clean_data(value):
    if pd.isnull(value):
        return value
    value = str(value).lower()
    value = value.replace('watt', '').replace('m2', '').replace('-', '').replace('.', '')
    try:
        return pd.to_numeric(value.strip())
    except ValueError:
        return pd.NA

df[columns_to_clean] = df[columns_to_clean].applymap(clean_data)

[26]: df.head()

[26]:   No      Nama Sekolah  BP Status siswa_total jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total skor_akreditasi
0  1    UPT SPF SD INPRES BADDOKA  SD Negeri     337.0       18       3        1300.0  1300     28080.0          81
1  2    UPT SPF SD INPRES BAKUNG 1  SD Negeri     243.0       15       4        800.0   900     30000.0          84
2  3    UPT SPF SD INPRES BAKUNG 2  SD Negeri     269.0       16       5        800.0  4400     25000.0          84
3  4    UPT SPF SD INPRES DAYA  SD Negeri     498.0       28       3        1300.0  4640     12000.0          87
4  5    UPT SPF SD INPRES KALANG TUBUNG 2  SD Negeri     195.0       15       1        800.0  1300     15050.0          81
```

### b. Seleksi Fitur (Normalisasi Data & Penerapan Teknik PCA)

```
[1]: #Normalisasi Data (Standar Scaler)
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler

[3]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_gabungan_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)

[4]: data.head()

[4]:
```

No	Nama Sekolah	BP	Status	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi
0	UPT SPF SD INPRES BADDOKA	SD	Negeri	337	18	3	13	1300	2808	81
1	UPT SPF SD INPRES BAKUNG 1	SD	Negeri	243	15	4	8	900	3000	84
2	UPT SPF SD INPRES BAKUNG 2	SD	Negeri	269	16	5	8	4400	2500	84
3	UPT SPF SD INPRES DAYA	SD	Negeri	498	28	3	13	4640	1200	87
4	UPT SPF SD INPRES KALANG TUBUNG 2	SD	Negeri	195	15	1	8	1300	1505	81

```
[6]: x = data.loc[:, ['siswa_total', 'jumlah_guru', 'staf_total', 'jumlah_ruangan', 'daya_listrik', 'wilayah_total', 'skor_akreditasi']]

[7]: scaler = StandardScaler()
scaler.fit(x)
data_scaled = scaler.transform(x)

[8]: data_scaled
```

```
[8]: array([[ 0.9412311 ,  0.47083351, -0.08851645, ..., -0.22739045,
       0.35771901,  5.11200806],
       [ 0.430646 ,  0.20118151,  0.28084404, ...,  0.11194342,
       5.29355719,  5.11278903],
       [ 0.57187167,  0.29106551,  0.65020453, ...,  0.11194342,
       5.29980749, -0.11951391],
       ...,
       [-0.88927078, -1.14707847, -0.82723743, ..., -0.7929469 ,
       5.37159007,  5.10729636],
       [-0.76977214, -0.69765848, -0.82723743, ...,  0.90372245,
       5.35920226,  5.10733737],
       [ 0.14276292, -0.15835449,  0.65020453, ...,  0.22505471,
       -0.19288573, -0.70029933]])
```

```
[10]: data_scaled = pd.DataFrame(data_scaled, columns = ['siswa_total', 'jumlah_guru', 'staf_total'],
 <...>

[11]: data_scaled
```

```
[11]:
```

	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi
0	0.941231	0.470834	-0.088516	0.505200	-0.180114	-0.028092	-0.227390
1	0.430646	0.201182	0.280844	-0.241330	-0.214162	-0.024249	0.111943
2	0.571872	0.291066	0.650205	-0.241330	0.083761	-0.034256	0.111943
3	1.815744	1.369673	-0.088516	0.505200	0.104190	-0.060273	0.451277
4	0.169922	0.201182	-0.827237	-0.241330	-0.180114	-0.054169	-0.227390
...	...	...	...	...	...	...	...
9196	0.148195	-0.068470	0.650205	0.654506	0.611511	-0.014603	0.111943
9197	4.455577	5.234695	0.650205	3.789932	-0.001360	0.025264	0.111943
9198	-0.889271	-1.147078	-0.827237	-1.435778	0.185907	-0.030253	-0.792947
9199	-0.769772	-0.697658	-0.827237	-0.689248	-0.103505	0.123770	0.903722
9200	0.142763	-0.158354	0.650205	0.803812	-0.163089	-0.013582	0.225055

```
[12]: data_scaled.to_csv('data_normalisasi_sekolah_sulsel.csv', index = False)
```

```
[1]: #Penerapan Teknik PCA
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
[2]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_normalisasi_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)
[3]: data.head()
   siswa_total jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total skor_akreditasi
0  0.941231  0.470834 -0.088516  0.50520 -0.180114 -0.028092 -0.227390
1  0.430646  0.201182  0.280844 -0.24133 -0.214162 -0.024249  0.111943
2  0.571872  0.291066  0.650205 -0.24133  0.083761 -0.034256  0.111943
3  1.815744  1.369673 -0.088516  0.50520  0.104190 -0.060273  0.451277
4  0.169922  0.201182 -0.827237 -0.24133 -0.180114 -0.054169 -0.227390
[4]: features = ['siswa_total', 'jumlah_guru', 'staf_total', 'jumlah_ruangan', 'daya_listrik', 'wilayah_total', 'skor_akreditasi']
x = data.loc[:, features].values
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)
data_pca = pd.DataFrame(data = principalComponents, columns = ['pca1', 'pca2'])
Ac
Go
[5]: data_pca
   pca1      pca2
0  1.628340  3.349251
1  1.781699  6.710388
2  1.373750  3.268771
3  2.638114  3.056547
4  0.434320  3.433142
...
9196  2.215221  6.915696
9197  8.545746  5.302014
9198 -0.665992  7.397538
9199  0.133514  7.232193
9200  0.548619 -0.779776
9201 rows × 2 columns
[11]: data_pca.to_csv('data_pca_sekolah_sulsel.csv', index = False)
```

### c. Algoritma DBSCAN Only

```
[1]: #Penerapan Algoritma DBSCAN Only

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import DBSCAN
from sklearn.neighbors import NearestNeighbors
from itertools import product

[64]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/pca_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)

[65]: data.head()

[65]:   pca1    pca2
0  1.628340  3.349251
1  1.781699  6.710388
2  1.373750  3.268771
3  2.638114  3.056547
4  0.434320  3.433142

[66]: features = ['pca1', 'pca2']
x1 = data.loc[:100, features]

[67]: nn = NearestNeighbors(n_neighbors=2)
nbrs = nn.fit(x1)
distances,indices=nbrs.kneighbors(x1)

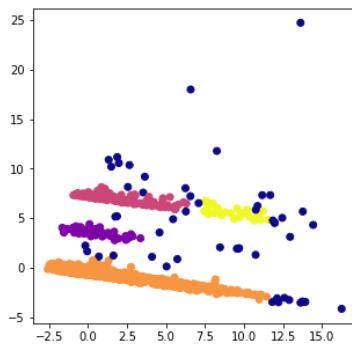
[68]: distances = np.sort(distances, axis=0)
distances = distances[:,1]
plt.rcParams['figure.figsize'] = (5,5)
plt.plot(distances)
plt.show()
```

```
[69]: x = data.loc[:, features].values
dbscan = DBSCAN(eps=0.7, min_samples=7).fit(x)
labels = dbscan.labels_
```

```
[70]: print("klaster total:{}".format(len(set(labels[np.where(labels != -1)]))))
print(labels)
```

```
klaster total:4
[0 1 0 ... 1 1 2]
```

```
[71]: plt.scatter(x[:,0], x[:,1], c=labels, cmap="plasma")
plt.show()
```



```
[72]: klaster = pd.DataFrame(data = labels, columns = ['dbscan_kluster'])
klaster
```

	dbscan_kluster
0	0
1	1
2	0
3	0
4	0
...	...
9196	1
9197	3
9198	1
9199	1
9200	2

9201 rows × 1 columns

```
[73]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(data, labels)
print(SI)
```

---

0.667756759373751

---

```
[73]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(data, labels)
print(SI)

0.667756759373751

[60]: klaster[klaster['dbscan_kluster'] == -1].count()

[60]: dbSCAN_kluster    16
      dtype: int64

[61]: datasekolah = pd.read_csv('C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_gabungan_sekolah_susel.csv')

[62]: hasil1 = pd.concat([data, pd.DataFrame(klaster)],axis=1)
hasil2 = pd.concat([datasekolah, pd.DataFrame(hasil1)],axis=1)
hasil2.to_csv('data_dari_dbSCAN.csv', index = False)
hasil2

[62]:   No  Nama Sekolah  BP  Status siswa_total jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total skor_akreditasi Lintang Bujur
  0  1  UPT SPF SD  INPRES BADDOKA SD  Negeri     337       18       3        13     1300      2808          81  509153567  1195133502
  1  2  UPT SPF SD  INPRES BAKUNG 1 SD  Negeri     243       15       4        8      900      3000          84  5073361105  1195294071
  2  3  UPT SPF SD  INPRES BAKUNG 2 SD  Negeri     269       16       5        8      4400      2500          84  5079140801  119529189
  3  4  UPT SPF SD  SD  Negeri     498       28       3        13     4640      1200          87  5112902708  119512008
```

#### d. Algoritma Peningkatan DBSCAN

```
[1]: #Algoritma Peningkatan DBSCAN dengan Pengujian n_cluster == 2
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

[25]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/data_dari_dbSCAN.csv'
data = pd.read_csv(filelocation)

[26]: data.head()

[26]:   No      Nama Sekolah  BP Status siswa_total jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total skor_akreditasi Lintang Bujur pca1  f
0    1      UPT SPF SD INPRES SD Negeri      337       18      3        13     1300      2808          81 509153567 1195133502 1.628340 3.346
1    2      UPT SPF SD BAKUNG SD Negeri      243       15      4        8      900      3000          84 5073361105 1195294071 1.781699 6.710
2    3      UPT SPF SD BAKUNG SD Negeri      269       16      5        8     4400      2500          84 507914080 119529189 1.373750 3.266
[5]: fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values

[6]: dbSCAN_kmeans = KMeans(n_clusters=2, init = 'k-means++')
y_dbSCAN_kmeans = dbSCAN_kmeans.fit_predict(x)

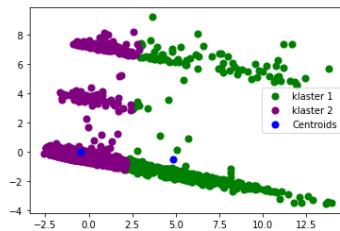
[7]: y_dbSCAN_kmeans

[7]: array([0, 0, 0, ..., 0, 0, 0])

[9]: klaster_dbSCAN_kmeans = pd.DataFrame(data = y_dbSCAN_kmeans, columns = ['peningkatanDBSCAN_2klaster'])
klaster_dbSCAN_kmeans

[9]:   peningkatanDBSCAN_2klaster
0                 0
1                 0
2                 0
3                 1
4                 0
...
9180               0
9181               1
9182               0
9183               0
```

```
[31]: plt.scatter(x[y_dbSCAN_kmeans==0,0], x[y_dbSCAN_kmeans==0,1], s=50, c='green', label = 'klaster 1')
plt.scatter(x[y_dbSCAN_kmeans==1,0], x[y_dbSCAN_kmeans==1,1], s=50, c='purple', label = 'klaster 2')
plt.scatter(dbSCAN_kmeans.cluster_centers_[:, 0], dbSCAN_kmeans.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()
```



```
[32]: hasil = pd.concat([data, pd.DataFrame(klaster_dbSCAN_kmeans)], axis=1)
hasil.to_csv('2 cluster.csv', index = False)
```

```
[33]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_dbSCAN_kmeans)
print(SI)
```

0.7505380325803438

```
[1]: ##Algoritma Peningkatan DBSCAN dengan Pengujian n_cluster = 3
```

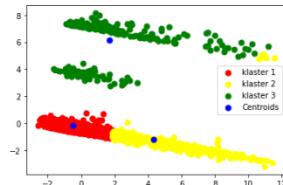
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

```
[59]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_dari_dbSCAN.csv'
data = pd.read_csv(filelocation)
```

```
[60]: data.head()
```

	No	Nama Sekolah	BP	Status	siswa_total	jumlah_guru	staf_total	jumlah_ruangan	daya_listrik	wilayah_total	skor_akreditasi	Lintang	Bujur	pca1	F
0	1	UPT SPF IN PRES BADDOKA	SD	Negeri	337	18	3	13	1300	2808	81	509153567	1195133502	1.628340	3.345
1	2	UPT SPF IN PRES BAKUNG 1	SD	Negeri	243	15	4	8	900	3000	84	5073361105	1195294071	1.781699	6.710
2	3	UPT SPF IN PRES BAKUNG 2	SD	Negeri	269	16	5	8	4400	2500	84	5079140801	1195291891	1.373750	3.268

```
[282]: plt.scatter(x[y_dbSCAN_kmeans==0,0], x[y_dbSCAN_kmeans==0,1], s=50, c='red', label = 'klaster 1')
plt.scatter(x[y_dbSCAN_kmeans==1,0], x[y_dbSCAN_kmeans==1,1], s=50, c='yellow', label = 'klaster 2')
plt.scatter(x[y_dbSCAN_kmeans==2,0], x[y_dbSCAN_kmeans==2,1], s=50, c='green', label = 'klaster 3')
plt.scatter(dbSCAN_kmeans.cluster_centers_[:, 0], dbSCAN_kmeans.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()
```



```
[10]: hasil = pd.concat([data, pd.DataFrame(klaster_dbSCAN_kmeans)], axis=1)
hasil.to_csv('data_hasil_peningkatan_dbSCAN_3clusterrrr.csv', index = False)
```

```
[12]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_dbSCAN_kmeans)
print(SI)
```

0.7745940801495571

Activate Windows  
Go to Settings to activate Wind

```
[8]: #Algoritma Peningkatan DBSCAN dengan Pengujian n_cluster = 4
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

[24]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/data_dari_dbSCAN.csv'
data = pd.read_csv(filelocation)

[25]: data.head()

[25]:
  No Nama Sekolah BP Status siswa_total jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total skor_akreditasi Lintang Bujur pca1 F
  0 UPT SPF SD INPRES Negeri 337 18 3 13 1300 2808 81 509153567 1195133502 1.628340 3.345
  1 UPT SPF SD INPRES Negeri 243 15 4 8 900 3000 84 5073361105 1195294071 1.781699 6.710
  2 UPT SPF SD INPRES Negeri 269 16 5 8 4400 2500 84 5079140801 119529189 1.373750 3.266

[33]: fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values

[34]: dbSCAN_kmeans = KMeans(n_clusters=4, init = 'k-means++')
y_dbSCAN_kmeans = dbSCAN_kmeans.fit_predict(x)

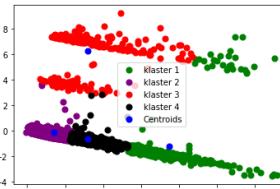
[35]: y_dbSCAN_kmeans

[35]: array([2, 2, 2, ..., 2, 2, 3])

[36]: klaster_dbSCAN_kmeans = pd.DataFrame(data = y_dbSCAN_kmeans, columns = ['kmeans_klaster'])
klaster_dbSCAN_kmeans

[36]:
  kmeans_klaster
  0 2
  1 2
  2 2
  3 2
  4 2
  ...
  9180 2
  9181 0
  9182 2
  9183 2

[37]: plt.scatter(x[y_dbSCAN_kmeans==0,0], x[y_dbSCAN_kmeans==0,1], s=50, c='green', label = 'klaster 1')
plt.scatter(x[y_dbSCAN_kmeans==1,0], x[y_dbSCAN_kmeans==1,1], s=50, c='purple', label = 'klaster 2')
plt.scatter(x[y_dbSCAN_kmeans==2,0], x[y_dbSCAN_kmeans==2,1], s=50, c='red', label = 'klaster 3')
plt.scatter(x[y_dbSCAN_kmeans==3,0], x[y_dbSCAN_kmeans==3,1], s=50, c='black', label = 'klaster 4')
plt.scatter(dbSCAN_kmeans.cluster_centers_[:, 0], dbSCAN_kmeans.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()


[38]: hasil = pd.concat([data, pd.DataFrame(klaster_dbSCAN_kmeans)], axis=1)
hasil.to_csv('data_hasil_peningkatan_dbSCAN_4cluster.csv', index = False)

[39]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_dbSCAN_kmeans)
print(SI)
```

```
[1]: # Import ulang pustaka yang diperlukan
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans

# Reload dataset
filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/Mentahan data gabungan/data_gabungan_sekolah_sulsel.csv'
df = pd.read_csv(filelocation)

# Pilih kembali kolom numerik yang relevan untuk clustering
numerical_cols = ["siswa_total", "jumlah_guru", "staf_total", "jumlah_ruangan", "daya_listrik", "wilayah_total"]
df_clean = df[numerical_cols].dropna()

# Normalisasi data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df_clean)

# Tentukan jumlah cluster optimal menggunakan metode Elbow
inertia = []
K_range = range(1, 11)

for k in K_range:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(X_scaled)
    inertia.append(kmeans.inertia_)

# Plot Elbow Method
plt.figure(figsize=(8, 5))
plt.plot(K_range, inertia, marker="o", linestyle="-")
plt.xlabel("Jumlah Cluster (k)")
plt.ylabel("Inertia")
plt.title("Elbow Method untuk Menentukan Jumlah Cluster Optimal")
plt.grid()
plt.show()
```

### e. Algoritma K-Means Only

```
[1]: #Algoritma K-Means Only dengan pengujian n_cluster = 2

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

[2]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_pca_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)

[3]: data.head()

[3]:   pca1    pca2
0  1.628340  3.349251
1  1.781699  6.710388
2  1.373750  3.268771
3  2.638114  3.056547
4  0.434320  3.433142

[4]: fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values

[15]: kmeans_only = KMeans(n_clusters=2, init = 'k-means++')
y_kmeans_only = kmeans_only.fit_predict(x)

[12]: fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values

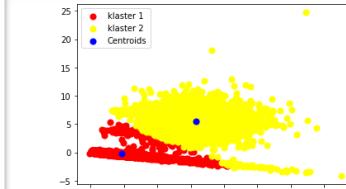
[13]: kmeans_only = KMeans(n_clusters=2, init = 'k-means++')
y_kmeans_only = kmeans_only.fit_predict(x)

[14]: y_kmeans_only
array([0, 1, 0, ..., 1, 1, 0])

[15]: klaster_kmeans_only = pd.DataFrame(data = y_kmeans_only, columns = ['kmeans_only_2klaster'])
klaster_kmeans_only

[15]:   kmeans_only_2klaster
0                 0
1                 1
2                 0
3                 1
4                 0
...
9196              1
9197              1
9198              1
9199              1

[16]: plt.scatter(x[y_kmeans_only==0], x[y_kmeans_only==0], s=50, c='red', label = 'klaster 1')
plt.scatter(x[y_kmeans_only==1], x[y_kmeans_only==1], s=50, c='yellow', label = 'klaster 2')
plt.scatter(kmeans_only.cluster_centers_[:, 0], kmeans_only.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()

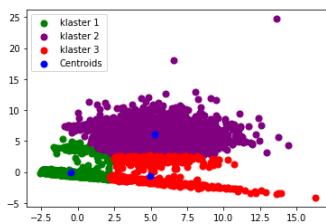

[17]: hasil = pd.concat([data, pd.DataFrame(klaster_kmeans_only)], axis=1)
hasil.to_csv('data_hasil_kmeans_2cluster.csv', index = False)

[18]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_kmeans_only)
print(SI)
```

Acti  
Go to

```
[1]: #Algoritma K-Means Only dengan pengujian n_cluster = 3
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
[2]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_pca_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)
[3]: data.head()
[3]:
   pca1      pca2
0  1.628340  3.349251
1  1.781699  6.710388
2  1.373750  3.268771
3  2.638114  3.056547
4  0.434320  3.433142
[4]: fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values
[5]: kmeans_only = KMeans(n_clusters=3, init = 'k-means++')
y_kmeans_only = kmeans_only.fit_predict(x)
[19]: y_kmeans_only
[19]: array([0, 2, 0, ..., 2, 2, 0])
[20]: klaster_kmeans_only = pd.DataFrame(data = y_kmeans_only, columns = ['kmeans_only_3klaster'])
klaster_kmeans_only
[20]:
   kmeans_only_3klaster
0                      0
1                      2
2                      0
3                      2
4                      0
...
9196                  2
9197                  2
9198                  2
9199                  2
9200                  0
9201 rows × 1 columns
```

```
[8]: plt.scatter(x[y_kmeans_only==0,0], x[y_kmeans_only==0,1], s=50, c='green', label = 'klaster 1')
plt.scatter(x[y_kmeans_only==1,0], x[y_kmeans_only==1,1], s=50, c='purple', label = 'klaster 2')
plt.scatter(x[y_kmeans_only==2,0], x[y_kmeans_only==2,1], s=50, c='red', label = 'klaster 3')
plt.scatter(kmeans_only.cluster_centers_[:, 0], kmeans_only.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()
```



```
[9]: hasil = pd.concat([data, pd.DataFrame(klaster_kmeans_only)], axis=1)
hasil.to_csv('data_hasil_kmeans_3cluster.csv', index = False)
```

```
[10]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_kmeans_only)
print(SI)
```

Act  
Go t

```
[1]: #Algoritma K-Means Only dengan pengujian n_cluster = 3
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

```
[1340]: filelocation = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_pca_sekolah_sulsel.csv'
data = pd.read_csv(filelocation)
```

```
[1341]: data.head()
```

	pca1	pca2
0	1.628340	3.349251
1	1.781699	6.710388
2	1.373750	3.268771
3	2.638114	3.056547
4	0.434320	3.433142

```
[1342]: fitur = ['pca1', 'pca2']
x = data.loc[:, fitur].values
```

```
[1343]: kmeans_only = KMeans(n_clusters=4, init = 'k-means++')
y_kmeans_only = kmeans_only.fit_predict(x)
```

```
[4]: futur = ['pca1', 'pca2']
x = data.loc[:, futur].values

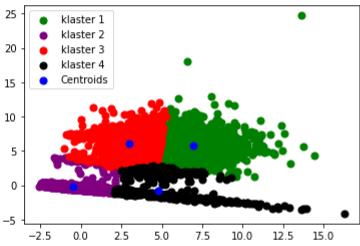
[5]: kmeans_only = KMeans(n_clusters=4, init = 'k-means++')
y_kmeans_only = kmeans_only.fit_predict(x)

[6]: y_kmeans_only
array([1, 1, 1, ..., 1, 1, 2])

[7]: klaster_kmeans_only = pd.DataFrame(data = y_kmeans_only, columns = ['kmeans_only_4klaster'])
klaster_kmeans_only

[7]: kmeans_only_4klaster
   0   1
  1  1
  2  1
  3  1
  4  2
 ...
9196  1
9197  3
9198  1
9199  1

353]: plt.scatter(x[y_kmeans_only==0,0], x[y_kmeans_only==0,1], s=50, c='green', label = 'klaster 1')
plt.scatter(x[y_kmeans_only==1,0], x[y_kmeans_only==1,1], s=50, c='purple', label = 'klaster 2')
plt.scatter(x[y_kmeans_only==2,0], x[y_kmeans_only==2,1], s=50, c='red', label = 'klaster 3')
plt.scatter(x[y_kmeans_only==3,0], x[y_kmeans_only==3,1], s=50, c='black', label = 'klaster 4')
plt.scatter(kmeans_only.cluster_centers_[:, 0], kmeans_only.cluster_centers_[:,1], s = 50, c = 'blue', label = 'Centroids')
plt.legend()
plt.show()


351]: hasil = pd.concat([data, pd.DataFrame(klaster_kmeans_only)], axis=1)
hasil.to_csv('data_hasil_kmeans_4cluster.csv', index = False)

348]: from sklearn.metrics import silhouette_samples, silhouette_score
SI = silhouette_score(x, y_kmeans_only)
print(SI)
0.6634546357693659
```

## f. Pemprosesan pembobotan perangkingan dan pemetaan

```
[54]: import pandas as pd
import folium
from folium.plugins import MarkerCluster, Search
from folium import FeatureGroup
from folium.plugins import MarkerCluster

[55]: file_path = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/Sekolah_Rekomendasi_BOP_PeningkatanDSBCAN3Cluster.csv'
df = pd.read_csv(file_path)

[57]: # Bobot perangkingan
bobot = [
    'skor_akreditasi': 29,
    'siswa_total': 23,
    'jumlah_ruangan': 19,
    'wilayah_total': 14,
    'daya_listrik': 5,
    'jumlah_guru': 5,
    'staf_total': 5
]

[58]: # Normalisasi data
for col in bobot.keys():
    if col == 'siswa_total':
        df['f'+col+'_n'] = (df[col] - df[col].min()) / (df[col].max() - df[col].min())
    else:
        df['f'+col+'_n'] = 1 * (df[col] - df[col].min()) / (df[col].max() - df[col].min())

```

Activate W  
Go to Settings

```
[1]: import pandas as pd
import folium
from folium.plugins import MarkerCluster, Search
from folium import FeatureGroup
from folium.plugins import MarkerCluster

[2]: file_path = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/data_hasil_kmeans_2cluster.csv'
df = pd.read_csv(file_path)

[5]: # Hitung skor akhir
df['skor_rekomendasi'] = (
    df['skor_akreditasi_n'] * bobot['skor_akreditasi'] +
    df['siswa_total_n'] * bobot['siswa_total'] +
    df['jumlah_ruangan_n'] * bobot['jumlah_ruangan'] +
    df['wilayah_total_n'] * bobot['wilayah_total'] +
    df['daya_listrik_n'] * bobot['daya_listrik'] +
    df['jumlah_guru_n'] * bobot['jumlah_guru'] +
    df['staf_total_n'] * bobot['staf_total']
)

[6]: # Normalisasi skor akhir ke skala 100
df['skor_rekomendasi'] = (df['skor_rekomendasi'] - df['skor_rekomendasi'].min()) / (df['skor_rekomendasi'].max() - df['skor_rekomendasi'].min()) * 100

[7]: # Perangkingan
df = df.sort_values(by='skor_rekomendasi', ascending=False).reset_index(drop=True)
df['peringkat'] = df.index + 1

[8]: # Tampilkan semua kolom asli ditambah skor akhir dan peringkat
hasil_rekomendasi = df.copy()

[9]: # Tampilkan hasil perangkingan
print(hasil_rekomendasi.head(10))

```

No	Nama Sekolah	BP	Status	siswa_total
0	UPT SDN 72 BONTOLE	SD	Negeri	294
1	SD INGRES PENGENTINGANG UTARA	SD	Negeri	406
2	SD INGRES BORONGKALUKU	SD	Negeri	334
3	UPT SD NEGERI 149 PINRANG	SD	Negeri	240
4	SD NEGERI ANASSAPPU	SD	Negeri	201
5	SD NEGERI 4 ARAMA	SD	Negeri	257

Activate Windows  
Go to Settings to activate Wi

```

4  4475          SD NEGERI ANASSAPPU  SD Negeri      201
5  4475          SD NEGERI 4 ARAWA  SD Negeri      257
6  7163 UPTD SMPN 31 SATAP LALANG TEDONG  SMP Negeri    150
7  3076          SD KRISTEN BUNGIN PASANG  SD Swasta    146
8  2318          SD INPRES BIRING PANTING  SD Negeri    176
9  2671 UPT SD NEGERI 11 BANGKALA BARAT  SD Negeri    187

jumlah_guru staf_total jumlah_ruangan daya_listrik wilayah_total ...
0      8        2        7      5500     1965 ...
1     21        2        7      900      2594 ...
2     10        2        7      900      4126 ...
3     10        3        7      5500      602 ...
4     10        0        8     1300     1547 ...
5     12        2       10      5500     2237 ...
6      9        2        4     3500      578 ...
7      4        1        6      900     1626 ...
8      8        1        7     3500     2623 ...
9      9        2        7     900     3703 ...

kmeans_klaster skor_akreditasi_n siswa_total_n jumlah_ruangan_n \
0      0      1.000000  0.184789  0.920455
1      0      0.956522  0.255185  0.920455
2      0      0.956522  0.209931  0.920455
3      0      1.000000  0.150849  0.920455
4      0      1.000000  0.126336  0.909091
5      0      1.000000  0.161534  0.886364
6      0      1.000000  0.094280  0.954545
7      0      1.000000  0.091766  0.931818
8      0      1.000000  0.110622  0.920455
9      0      1.000000  0.117536  0.920455

wilayah_total_n daya_listrik_n jumlah_guru_n staf_total_n \
0      0.997940      0.980442      0.936      0.93750

wilayah_total_n daya_listrik_n jumlah_guru_n staf_total_n \
0      0.997940      0.980442      0.936      0.93750
1      0.997269      0.997164      0.832      0.93750
2      0.995634      0.997164      0.920      0.93750
3      0.999395      0.980442      0.920      0.90625
4      0.998386      0.995710      0.920      1.00000
5      0.997650      0.980442      0.904      0.93750
6      0.999420      0.987713      0.928      0.93750
7      0.998302      0.997164      0.968      0.96875
8      0.997238      0.987713      0.936      0.96875
9      0.996085      0.997164      0.928      0.93750

skor_rekomendasi peringkat
0      100.000000      1
1      99.799970      2
2      98.374842      3
3      97.723149      4
4      97.154843      5
5      96.923130      6
6      96.762670      7
7      96.530093      8
8      96.520095      9
9      96.506138      10

[10 rows x 26 columns]

:# Simpan hasil ke file CSV
hasil_rekomendasi.to_csv('Sekolah_Rekomendasi_Bantuan_BOP_PeningkatanDSBCAN3Cluster.csv', index=False)

:# Buat peta menggunakan Folium
map = folium.Map(location=[-2.5489, 118.0149], zoom_start=8)
marker_cluster = MarkerCluster().add_to(map)

```

2	69.571546	3	
3	68.659188	4	
4	68.562265	5	
5	66.793444	6	
6	66.350382	7	
7	65.799840	8	
8	65.651895	9	
9	65.614784	10	

```

[10 rows x 25 columns]

:# Simpan hasil ke file CSV
hasil_rekomendasi.to_csv('Sekolah_Rekomendasi_Bantuan_BOP_KMeansOnly2Cluster.csv', index=False)

```

```
[57]: # Filter jenjang pendidikan
jenjang_groups = {
    'SD': folium.FeatureGroup(name='SD').add_to(map),
    'SMP': folium.FeatureGroup(name='SMP').add_to(map),
    'SMA': folium.FeatureGroup(name='SMA').add_to(map),
    'SMK': folium.FeatureGroup(name='SMK').add_to(map)
}
```

```
[59]: for idx, row in hasil_rekomendasi.iterrows():
    # Lewati jika Lintang atau Bujur kosong
    if pd.isna(row['Lintang']) or pd.isna(row['Bujur']):
        continue

    popup_info = (
        f"Nama Sekolah: {row['Nama Sekolah']}<br>"
        f"Peringkat Rekomendasi: {row['peringkat']}<br>"
        f"Skor Akreditasi: {row['skor_akreditasi']}<br>"
        f"Jumlah Siswa: {row['siswa_total']}<br>"
        f"Luas Wilayah: {row['wilayah_total']}<br>"
        f"Jumlah Ruangan: {row['jumlah_ruangan']}<br>"
        f"Total Staf: {row['staf_total']}<br>"
        f"Total Guru: {row['jumlah_guru']}<br>"
        f"Daya Listrik: {row['daya_listrik']}<br>"
    )

    # Tentukan warna marker berdasarkan index
    if idx <= 2290:
        warna = 'red'
    elif idx <= 6021:
        warna = 'orange'
    else:
        warna = 'green'

    marker = folium.Marker(
        location=[row['Lintang'], row['Bujur']],
        popup=popup_info,
        icon=folium.Icon(color=warna)
    )
```

```
jenjang = row['BP'] if row['BP'] in jenjang_groups else 'SD'
marker.add_to(jenjang_groups[jenjang])
```

```
[60]: # Tambahkan fitur pencarian Nama Sekolah
Search(
    layer=marker_cluster,
    search_label='Nama Sekolah',
    placeholder='Cari sekolah...',
    collapsed=False
).add_to(map)
```

```
[60]: <folium.plugins.search.Search at 0x1cb401ec820>
```

```
[61]: # Tambahkan fitur pencarian Nama Sekolah
Search(
    layer=marker_cluster,
    search_label='peringkat',
    placeholder='Masukkan Peringkat Sekolah',
    collapsed=False
).add_to(map)
```

```
[61]: <folium.plugins.search.Search at 0x1cb373c6b80>
```

```
[62]: # Tambahkan kontrol Layer untuk filter jenjang
folium.LayerControl().add_to(map)
```

```
[62]: <folium.map.LayerControl at 0xcb3bbe2460>
```

```
[63]: # Simpan peta ke file HTML
map.save('Peta_FIX_Rekomendasi_Sekolah_Sulsel.html')
print("Peta Rekomendasi telah disimpan sebagai 'Peta_FIX_Rekomendasi_Sekolah_Sulsel.html'")

Peta Rekomendasi telah disimpan sebagai 'Peta_FIX_Rekomendasi_Sekolah_Sulsel.html'
```

```
[11]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Membaca data dari file CSV
file_path = 'C:/Users/User/Documents/Berkas Fikar/Skripsi Fikar/Database/Hasil_FIX_Sekolah_Rekomendasi_Bantuan_BOP.csv'
df = pd.read_csv(file_path)

# Korelasi untuk heatmap
indicators = ['skor_akreditasi', 'siswa_total', 'jumlah_ruangan', 'wilayah_total', 'daya_listrik', 'jumlah_guru', 'staf_total']
correlation = df[indicators + ['skor_rekomendasi']].corr()

# Visualisasi heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation, annot=True, cmap='Blues', fmt=".2f", linewidths=0.5)
plt.title('Heatmap Korelasi antara Indikator dan Skor Ranking Rekomendasi')
plt.show()
```



### Lampiran 3 Source Code keseluruhan

#### Source Code

<https://github.com/Zulfikarrish/K-Means-dan-Peningkatan-DBSCAN-BOP.git>

### Lampiran 4 Gambaran dataset sari Dinas Pendidikan Sulawesi Selatan

Kab. Bantaeng.xlsx	2/21/2025 3:48 PM	Microsoft Excel W...	37 KB
Kab. Barru.xlsx	2/21/2025 4:02 PM	Microsoft Excel W...	41 KB
Kab. Bone.xlsx	2/21/2025 4:44 PM	Microsoft Excel W...	120 KB
Kab. Bulukumba.xlsx	2/21/2025 8:29 PM	Microsoft Excel W...	68 KB
Kab. ENrekang.xlsx	2/21/2025 8:42 PM	Microsoft Excel W...	47 KB
Kab. Gowa.xlsx	2/22/2025 9:06 AM	Microsoft Excel W...	90 KB
Kab. Jeneponto.xlsx	2/22/2025 9:49 AM	Microsoft Excel W...	61 KB
Kab. Luwu Timur.xlsx	2/22/2025 10:13 AM	Microsoft Excel W...	42 KB
Kab. Luwu Utara.xlsx	2/22/2025 10:29 AM	Microsoft Excel W...	55 KB
Kab. Luwu.xlsx	2/22/2025 9:57 AM	Microsoft Excel W...	68 KB
Kab. Pangkajene Kepulauan.xlsx	2/22/2025 11:24 AM	Microsoft Excel W...	68 KB
Kab. Pinrang.xlsx	2/22/2025 12:43 PM	Microsoft Excel W...	61 KB
Kab. Sinjai.xlsx	2/23/2025 7:16 AM	Microsoft Excel W...	50 KB
Kab. Soppeng.xlsx	2/23/2025 7:36 AM	Microsoft Excel W...	49 KB
Kab. Takalar.xlsx	2/23/2025 7:46 AM	Microsoft Excel W...	51 KB
Kab. Tana Toraja.xlsx	2/23/2025 7:57 AM	Microsoft Excel W...	53 KB
Kab. Toraja Utara.xlsx	2/23/2025 8:10 AM	Microsoft Excel W...	47 KB
Kab. Wajo.xlsx	2/23/2025 8:24 AM	Microsoft Excel W...	73 KB
Kab.Sidrap.xlsx	2/22/2025 1:17 PM	Microsoft Excel W...	50 KB
Kep. Selayar.xlsx	2/23/2025 9:43 AM	Microsoft Excel W...	38 KB
Kota Makassar.xlsx	2/27/2025 9:41 PM	Microsoft Excel W...	15 KB
Kota Palopo.xlsx	2/23/2025 9:52 AM	Microsoft Excel W...	28 KB
Maros.xlsx	2/22/2025 10:56 AM	Microsoft Excel W...	60 KB
Pare-pare.xlsx	2/23/2025 12:10 PM	Microsoft Excel W...	27 KB

No	Nama Sekolah	NPSN	BP	Data Sekolah Kota Makassar												
				Status	Last Sync	Jml Sync	PD	Rombel	Guru	Fewgawal	R. Kelas	R. Lab	R. Perpus	Jumlah Rl Daya Listril Luas Wila Skor Akre L		
1	UPT SPF SD INPRES BARRANG CADDI 2	40312057	SD	Negeri	15 Feb 2024 11:54:16	6	80	6	6	3	4	0	0	4 1200 Watt 768 m2	77	
2	UPT SPF SD INPRES BARRANG LOMPO	40312024	SD	Negeri	18 Apr 2024 18:46:23	6	263	12	15	2	7	0	1	8 7200 Watt 3300 m2	92	
3	UPT SPF SD INPRES BERTINGKAT TABARINGAN I	40313725	SD	Negeri	23 April 2024 09:12:49	11	250	12	16	3	12	0	1	13 4900 Watt 2916 m2	86	
4	UPT SPF SD INPRES CAMBAYA 1	40312024	SD	Negeri	27 April 2024 09:02:01	19	328	12	13	3	6	1	1	8 2200 Watt 1984 m2	86	
5	UPT SPF SD INPRES CAMBAYA 2	40312054	SD	Negeri	27 April 2024 16:25:59	21	265	11	15	6	9	1	1	11 2200 Watt 1572 m2	86	
6	UPT SPF SD INPRES TABARINGAN	40313727	SD	Negeri	18 April 2024 15:16:15	10	289	11	15	4	5	0	1	6 900 Watt 484 m2	85	
7	UPT SPF SD INPRES TABARINGAN I	40313726	SD	Negeri	26 April 2024 16:14:04	22	248	10	14	4	5	0	1	6 800 Watt 172 m2	84	
8	UPT SPF SD NEGERI BARRANG CADDI	40313875	SD	Negeri	01 May 2024 17:32:18	15	175	6	9	4	6	0	1	7 2200 Watt 1200 m2	83	
9	UPT SPF SD NEGERI BARRANG LOMPO	40312198	SD	Negeri	23 Jan 2024 09:29:19	2	343	12	18	4	10	1	1	12 3600 Watt 2238 m2	91	
10	UPT SPF SD NEGERI CAMBAYA	40312083	SD	Negeri	22 April 2024 22:38:08	29	286	12	16	5	7	0	1	8 1200 Watt 899 m2	85	
11	UPT SPF SD NEGERI KODINGARENG	40313883	SD	Negeri	30 April 2024 15:21:02	17	502	20	27	6	9	0	1	10 2200 Watt 1500 m2	84	
12	UPT SPF SD NEGERI LANGKAJ	40313755	SD	Negeri	19 April 2024 17:12:05	15	131	6	8	2	6	0	0	6 2200 Watt 1575 m2	87	
13	UPT SPF SD NEGERI LUMU-LUMU	40312094	SD	Negeri	16 Feb 2024 11:23:53	2	131	6	9	1	6	0	1	7 800 Watt 280 m2	83	
14	UPT SPF SD NEGERI PATTINGALOONG 1	40312091	SD	Negeri	21 April 2024 09:54:44	22	320	12	16	3	6	0	0	6 1200 Watt -	83	
15	UPT SPF SD NEGERI TABARINGAN 5	40313724	SD	Negeri	17 Mar 2024 20:53:39	12	178	8	9	4	6	0	1	7 3890 Watt 2600 m2	84	
16	UPT SPF SD NEGERI UJUNG TANAH 2	40312094	SD	Negeri	30 April 2024 12:49:11	19	378	12	16	6	6	0	1	7 2200 Watt 1600 m2	83	
17	UPT SPF SD NEGERI UJUNG TANAH I	40312093	SD	Negeri	06 May 2024 08:51:27	27	315	12	15	3	8	0	1	9 910 Watt 400 m2	81	
18	SD HANG TUAH MAKASSAR	40313109	SD	Swasta	04 April 2024 11:36:32	15	279	12	16	5	10	1	1	12 930 Watt 3360 m2	92	
19	SD MUHAMMADIYAH 7	40314042	SD	Swasta	04 May 2024 14:20:45	18	192	6	5	1	7	2	1	10 810 Watt 300 m2	84	
1	UPT SPF SMP NEGERI 28 MAKASSAR	40312434	SMP	Negeri	2/10/2025 9:30	11	200	8	17	8	8	2	1	11 4200 Watt 2608 m2	85	
2	UPT SPF SMP NEGERI 38 MAKASSAR	40310203	SMP	Negeri	2/2/2025 14:39	13	187	8	16	7	8	2	1	11 2200 Watt 1550 m2	83	
3	UPT SPF SMP NEGERI 39 MAKASSAR	40313805	SMP	Negeri	2/6/2025 14:40	9	86	3	8	2	3	1	1	5 1200 Watt -	76	
4	UPT SPF SMP NEGERI 42 MAKASSAR	69787564	SMP	Negeri	1/26/2025 20:41	3	20	3	10	2	3	0	1	4 800 Watt 192 m2	74	
5	UPT SPF SMP NEGERI 43 MAKASSAR	69786472	SMP	Negeri	1/25/2025 21:51	3	45	3	11	3	2	1	1	4 2200 Watt 1560 m2	78	
6	UPT SPF SMP NEGERI 7 MAKASSAR	40312922	SMP	Negeri	2/9/2025 20:13	36	1164	33	58	17	33	2	1	36 8219 Watt 6240 m2	91	
7	SMP AL-MUTTAQIEM	40319226	SMP	Swasta	2/9/2025 18:13	16	41	3	3	1	8	2	1	11 1080 Watt 345 m2	75	
8	SMP DDI ALIRSYAD RAMPEGADING	40312466	SMP	Swasta	10/24/2024 23:46	0	0	0	2	2	0	0	0	0 1200 Watt 872 m2	82	

## Lampiran 5 Riwayat Hidup

### Curriculum Vitae

#### 1. Data Pribadi

- 1. Nama : Muhammad Zulfikar Aris
- 2. Tempat, tgl. Lahir : Palopo, 15 Maret 2001
- 3. Alamat : Nusa Harapan Permai F3/15
- 4. Kewarganegaraan : Warga Negara Indonesia

#### 2. Riwayat Pendidikan

- Tamat SD tahun 2014 di SDN 54 Salupikung Palopo, Sulawesi Selatan
- Tamat SMP tahun 2016 di SMPN 1 Palopo, Sulawesi Selatan
- Tamat SMA tahun 2019 di SMAN 3 Palopo, Sulawesi Selatan