

Implementasi Machine Learning Menggunakan Algoritma K-Means Untuk Klasifikasi Sekolah Dasar

Yudistra Bagus Pratama^{1,*}, Agung Setiawan²

¹Fakultas Teknik dan Sains, Program Studi Ilmu Komputer, Universitas Muhammadiyah Bangka Belitung, Pangkal Pinang, Indonesia

²Fakultas Keguruan dan Ilmu Pendidikan, Program Studi Pendidikan Guru Sekolah Dasar, Universitas Muhammadiyah Bangka Belitung, Pangkal Pinang, Indonesia

Email: ^{1,*}yudistira.bagus@unmuhbabel.ac.id, ²agung.setiawan@unmuhbabel.ac.id

Email Penulis Korespondensi: yudistira.bagus@unmuhbabel.ac.id

Abstrak—Sebagian besar orang tua mempertimbangkan status sekolah anak mereka dengan cara tertentu. Status sekolah, total sekolah, total guru, total siswa, dan total ruang kelas adalah faktor penting bagi orang tua saat memilih sekolah. Permasalahannya adalah data terkait klasifikasi sekolah dasar di kota Pangkalpinang belum tersedia sehingga orang tua murid & dinas terkait pun belum tahu akan status & klasifikasi sekolah-sekolah yang ada di lingkungan mereka. Data sekolah kota pangkalpinang sekarang dapat diolah dengan machine learning berkat kemajuan teknologi data science. Penelitian ini menghasilkan pengklasifikasian data sekolah berdasarkan cluster status sekolah. Metode klasifikasi pada penelitian ini menggunakan model pembelajaran mesin (ML) unsupervised yaitu K-Means Clustering. Data yang digunakan sejumlah 14 record kecamatan di Pangkalpinang untuk algoritma k-means clustering diperoleh dari <https://dapo.kemdikbud.go.id/>, yang dikonfirmasi ulang oleh Dinas Pendidikan dan Kebudayaan Kota Pangkalpinang. Dengan memodelkan data, penelitian ini menentukan standar sekolah dan menggunakan algoritma tersebut untuk menganalisis ketepatan klasifikasi sekolah berdasarkan parameternya. Berdasarkan hasil kumulatif nilai sillhouette score yang didapatkan dari status sekolah dengan persentase 100%, Klaster 1 memiliki persentase 21,43%, klaster 2 memiliki persentase 28,25%, klaster 3 memiliki persentase 14,28%, klaster 4 memiliki persentase 21,42%, dan klaster 5 memiliki persentase 14,29%. Jumlah klaster terbesar terlihat dari hasil plot kumulatif dengan nilai atribut paling rendah adalah klaster 2. Dengan akumulasi jumlah klaster yang dihasilkan, Pemerintah Kota Pangkalpinang sebagai entitas penyelenggara pendidikan dan potensi peserta didik dapat melakukan penentuan dan klasifikasi sekolah-sekolah tingkat dasar. Hal ini mencakup aspek peningkatan kuantitas sekolah dasar di daerah-daerah yang mengalami kekurangan sekolah dasar, penambahan tenaga pendidik untuk sekolah yang membutuhkan peningkatan guru, peningkatan jumlah siswa bagi sekolah yang memerlukan alokasi siswa yang lebih sedikit berdasarkan wilayah tertentu, dan peningkatan fasilitas ruang kelas di sekolah yang tidak memiliki ruang untuk pertemuan tatap muka dalam proses belajar.

Kata Kunci: Data Science; Machine Learning; Klasifikasi; Sekolah Dasar; Big Data

Abstract—The majority of parents take into account their child's educational standing to some extent. The school's status, number of schools, number of teachers, number of students, and number of classrooms are crucial considerations for parents when selecting a school. The problem is that data regarding the classification of elementary schools in the city of Pangkalpinang is not yet available so that parents and related agencies do not yet know the status & classification of schools in their area. The utilisation of machine learning has been possible for analysing data from Pangkalpinang City School, owing to the advancements in data science technology. This study generates a categorization of school data using clusters of school status. The research used an unsupervised machine learning (ML) model called K-means clustering for classification purposes. The dataset containing 14 sub-district records in Pangkalpinang, utilised for the k-means clustering technique, was acquired from the official website of the Ministry of Education and Culture (<https://dapo.kemdikbud.go.id/>). The authenticity of the data was verified by the Pangkalpinang City Education and Culture Office. This research use data modelling to establish school standards and utilises an algorithm to assess the precision of school categorization according to its parameters. According to the cumulative Sillhouette scores obtained from the school status, Cluster 1 for 21.43% of the total, Cluster 2 for 28.25%, Cluster 3 for 14.28%, Cluster 4 for 21.42%, and Cluster 5 for 14.29%. The cluster with the lowest attribute values, specifically cluster 2, exhibits the highest number of clusters as shown from the cumulative plot findings. The Pangkalpinang City Government can determine and categorise elementary-level schools by aggregating the number of resulting clusters, as the entity responsible for education and potential pupils. This encompasses measures such as expanding the number of primary schools in areas facing a scarcity of such institutions, augmenting the teaching workforce in schools that necessitate additional educators, accommodating more students in schools that have a need for smaller student-to-teacher ratios in specific regions, and enhancing classroom infrastructure in schools lacking adequate space for in-person instruction.

Keywords: Data Science; Machine Learning; Classification; Elementary School; Big Data

1. PENDAHULUAN

Beranjak dari visi misi dinas pendidikan dan kebudayaan kota Pangkalpinang yang ingin mewujudkan Pangkalpinang sebagai kota sejahtera, nyaman, unggul dan makmur serta meningkatkan kualitas sumber daya manusia yang berkeadilan, maka peran sekolah dasar menjadi tombak dalam mewujudkan visi misi yang dicita-citakan tersebut. Sekolah dasar adalah jenjang dasar pendidikan formal yang sangat penting untuk keberlangsungan dan pelaksanaan kejenjang yang lebih tinggi. Oleh karena itu, peraturan yang ditetapkan oleh pemerintah, terutama di Pangkalpinang, untuk menerima siswa baru harus sesuai dengan persyaratan usia calon siswa. Permendikbud 1/2021 pasal 4 menetapkan bahwa calon siswa sekolah dasar harus berusia 7 tahun atau paling rendah 6 tahun pada tanggal 1 Juli tahun sebelumnya[1].

Sebagian besar orang tua bergantung pada klasifikasi status sekolah anak mereka saat memilih sekolah mereka. Terutama jenis sekolah dan fasilitas adalah faktor utama yang memengaruhi keputusan orang tua tentang sekolah mana yang akan mereka pilih untuk anaknya. Permasalahannya adalah data terkait klasifikasi sekolah dasar di kota Pangkalpinang belum tersedia sehingga orang tua murid & dinas terkait pun belum tahu akan status & klasifikasi sekolah-sekolah yang ada di lingkungan mereka. Dengan mempartisi data menjadi beberapa kelompok, penelitian ini bertujuan

untuk menunjukkan bahwa informasi tentang status sekolah dasar di Pangkalpinang dapat diolah melalui klasifikasi berdasarkan kelompok dan domain data untuk setiap sekolah dasar sehingga mengurangi kesulitan orang tua dalam menentukan sekolah terbaik berdasarkan parameter data yang tersedia & membantu instansi terkait untuk membantu meningkatkan mutu pendidikan di daerah tersebut[2].

Studi ilmiah tentang algoritma dan model statistik yang digunakan sistem komputer untuk melakukan tugas tertentu tanpa diprogram secara eksplisit dikenal sebagai machine learning (ML) [3]. Ada sejumlah algoritma berbeda yang dapat digunakan seseorang dalam machine learning [4]. Algoritma machine learning biasanya terbagi menjadi beberapa tipe diantaranya yang sering digunakan adalah unsupervised learning dan supervised learning [5]. Supervised learning dimana data latih memerlukan label biasanya benar atau salah [6] sedangkan unsupervised learning data tidak perlu lagi diberi label [7].

Cara kerja unsupervised learning berupa memecah data maupun objek-objek yang memiliki value sama dalam ruang lingkup tertentu ke dalam sejumlah kelompok atau cluster [8]. Algoritma unsupervised learning yang digunakan pada penelitian ini adalah K-means clustering. Algoritma k-means biasanya merupakan teknik pengelompokan yang paling populer dan umum [9]. Algoritma k-means ini bekerja pada setiap iterasi dari hasil pengelompokan hirarki cluster [10].

Di Indonesia, penelitian tentang klasifikasi data dan algoritma K-Means telah banyak dilakukan. Salah satu contohnya adalah penelitian yang dilakukan oleh Amin Septianingsih, yang menemukan empat klaster: klaster 1 terdiri dari 10 orang, klaster 2 terdiri dari 19 orang, klaster 3 terdiri dari 3 orang, dan klaster 4 terdiri dari 2 orang[11]. Dibandingkan dengan klaster lain, klaster 1 memiliki persentase terendah untuk variabel seperti luas hunian kurang dari 7,2 m², atap ijuk yang paling luas, dan kepemilikan rumah sewa atau kontrak. Klaster 2 memiliki persentase tertinggi untuk variabel seperti akses hunian yang layak dan air minum yang layak. Klaster 3 memiliki persentase tertinggi untuk variabel seperti akses hunian yang layak, air minum yang layak, penerangan listrik yang layak, dan sanitasi yang layak. Lantai Tanah dan Dinding Terlulus: Variabel lantai tanah dan dinding terlulus bambu memiliki persentase terendah dibandingkan dengan variabel lain dalam klaster 4, termasuk sanitasi layak, kepemilikan rumah sewa atau kontrak, dan penerangan listrik, Lantai Tanah dan Dinding Terlulus: Variabel lantai terlulus tanah dan dinding bambu pada klaster 4 memiliki persentase terendah dibandingkan dengan variabel lain: penerangan listrik, sanitasi layak, atap ijuk terlulus, dan kepemilikan rumah sewa atau kontrak. Sebaliknya, variabel lantai terlulus tanah dan dinding bambu memiliki persentase tertinggi [12].

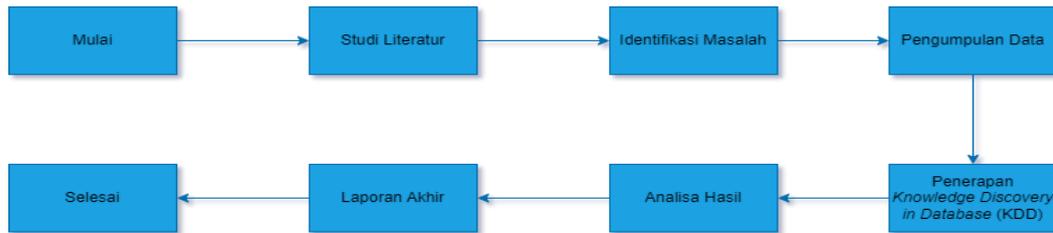
Penelitian tambahan tentang Klasterisasi Sekolah Menggunakan Algoritma K-Means Berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik menunjukkan bahwa setiap Kelompok memiliki jumlah anggota yang berbeda. Cluster 0 memiliki 178 sekolah, Cluster 1 memiliki 3 sekolah, dan Cluster 2 memiliki 43 sekolah. Hasil pemeringkatan Cluster diambil dari Cluster 1, Cluster 0, dan Cluster 2 secara berurutan. Selanjutnya, algoritma K-Means diuji kinerjanya dengan membagi menjadi tiga kelompok. Nilai Davies Bouldin Index adalah -0,695. Hasil penelitian menunjukkan bahwa 43 sekolah termasuk dalam cluster rendah. Oleh karena itu, penelitian ini mengusulkan bahwa dinas pendidikan pemerintah harus mendampingi dan membeli kebutuhan sekolah sekolah tersebut [13].

Metode klasifikasi pada penelitian ini menggunakan model pembelajaran mesin (ML) unsupervised seperti K-Means Clustering. Sebagai referensi, peneliti menggunakan data dari <https://dapo.kemdikbud.go.id/> untuk mengumpulkan data tentang jumlah sekolah dasar di masing-masing dari 7 kecamatan yang ada di kota Pangkalpinang. Dengan menggunakan algoritma tersebut, penelitian ini bertujuan untuk memodelkan data, menentukan standar sekolah, dan menganalisis ketepatan klasifikasi sekolah berdasarkan parameternya. Pentingnya penelitian ini karena hasilnya dapat dijadikan bahan evaluasi oleh dinas pendidikan dan kebudayaan kota pangkalpinang kedepan, bahwa parameter apa saja yang perlu diperbaiki atau dikembangkan sehingga dapat meningkatkan kualitas sumber daya manusia dan ekonomi di wilayah sekolah tersebut serta bisa digunakan sebagai dasar pertimbangan orang tua dalam pengambilan keputusan pemilihan sekolah dasar bagi anaknya berdasarkan parameter data yang didapatkan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

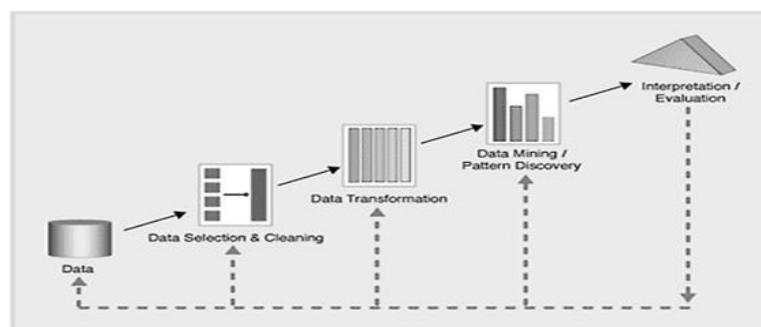
Penelitian ini dimulai dengan membaca literatur, mengidentifikasi masalah, mengumpulkan data, menggunakan metode knowledge discovery in database, melakukan analisis hasil, dan menyimpulkan dengan laporan akhir dan selesai. Pemanfaatan algoritma pada komputer sebagai penghasil produk-produk berbasis teknologi sudah sangat banyak [14]. Algoritma K-means merupakan salah satu algoritma partisi, karena K-Means didasarkan pada estimasi awal dengan mendefinisikan centroid grup [15]. Algoritma K-means menggunakan proses iteratif untuk mendapatkan database cluster. Dibutuhkan nomor cluster awal yang diinginkan sebagai masukan dan menghasilkan pusat massa akhir sebagai keluaran [16]. Metode clustering K-means akan memilih pola k sebagai titik awal secara acak atau centroid acak [17]. Banyaknya iterasi untuk mencapai pusat cluster akan dipengaruhi oleh calon pusat cluster secara acak. Maka ada cara untuk mengembangkan algoritma dengan menentukan centroid cluster dilihat dari kepadatan data awal yang tinggi agar mendapatkan performa yang lebih tinggi [18]. Penelitian ini dilakukan di Universitas Muhammadiyah Bangka Belitung dan akan mengumpulkan data dari dinas pendidikan dan kebudayaan kota Pangkalpinang, seperti yang ditunjukkan dalam jadwal tabel penelitian. Tahapan penelitian yang dikerjakan dapat dilihat pada gambar 1.



Gambar 1. Tahapan penelitian

Berikut merupakan deskripsi dari tahapan penelitian yang akan dilakukan :

1. Untuk memulai penelitian ini, peneliti membaca literatur tentang topik penelitian. Mereka melakukan ini dengan membaca buku, jurnal, makalah, dan laporan penelitian yang terkait dengan topik penelitian.
2. Setelah itu, identifikasi masalah dilakukan dengan tujuan untuk mengumpulkan sejumlah masalah yang dapat diselesaikan oleh peneliti. Dengan kata lain, identifikasi masalah yang baik akan menggambarkan masalah yang menjadi subjek penelitian.
3. Selanjutnya, untuk mendapatkan data dan informasi yang diperlukan untuk sekolah dasar di kota Pangkalpinang, dinas pendidikan dan kebudayaan kota Pangkalpinang disurvei, dokumentasi, dan wawancara.
4. Setelah data dikumpulkan, metode Knowledge Discovery in Database (KDD) diterapkan, seperti yang ditunjukkan pada Gambar 2 di bawah. Untuk mengetahui lebih banyak tentang hasil pengolahan data, lima langkah penting yang diambil dalam penelitian ini adalah :
 - a) Data understanding yaitu proses memahami data berdasarkan kebutuhan data (Data Requirement). Proses ini meliputi pengumpulan data (initial data collection) dan pendeskripsian data (data description).
 - b) Data preparation yaitu proses persiapan yang terdiri dari pemilihan data (pilihan data) dan pembersihan (pembersihan data). Dalam proses ini, data sekolah dasar dipilih sesuai dengan kebutuhan dan dibersihkan dari data yang tidak relevan atau jenis data yang salah.
 - c) Data transformations merupakan proses integrasi data yang telah dipilih oleh penelitian sehingga dapat digunakan untuk proses data mining. Tujuan dari proses ini adalah untuk mengubah data menjadi data yang dapat diolah dengan menggunakan algoritma yang ditetapkan oleh penelitian.
 - d) Data mining Proses mengumpulkan pola dan karakteristik data, Pada tahap ini, metode clustering digunakan; tujuannya adalah memodelkan data, menentukan standar sekolah, dan menganalisis ketepatan klasifikasi sekolah berdasarkan parameternya dengan menggunakan algoritma k-means clustering. Pada langkah clustering ini, data dikelompokkan berdasarkan status sekolah, jumlah sekolah, jumlah guru, jumlah siswa, dan jumlah ruang kelas.
 - e) Tahap analisis dan evaluasi dilakukan untuk mengidentifikasi pola yang dihasilkan dari mining process. Pola informasi yang dihasilkan dari mining process harus ditampilkan dalam bentuk yang mudah dipahami dan dipahami oleh pihak yang berkepentingan. Dengan menunjukkan nilai performa algoritma K-Means, evaluasi yang dilakukan akan melihat pola-pola dalam proses klasterisasi.



Gambar 2. Knowledge Discovery in Database[19][20]

5. Setelah pengolahan data dilakukan, baik dengan menggunakan metodologi Knowledge Discovery in Database (KDD), analisis hasil dilakukan. Ini dilakukan dengan melihat hasil penelitian sebelumnya untuk memahami hasil pengolahan data.
6. Terakhir menyusun laporan akhir dengan membuat sebuah artikel ilmiah berdasarkan pengamatan, pengolahan data dan analisis hasil yang sudah dilaksanakan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pemilihan Data

Data penelitian dalam format csv dikumpulkan dari <https://dapo.kemdikbud.go.id/> dan dipantau secara real-time. Tabel 1 menunjukkan hasil pemantauan data pada 2 Oktober 2023 pukul 13.00 wib. Kumpulan data ini disusun berdasarkan

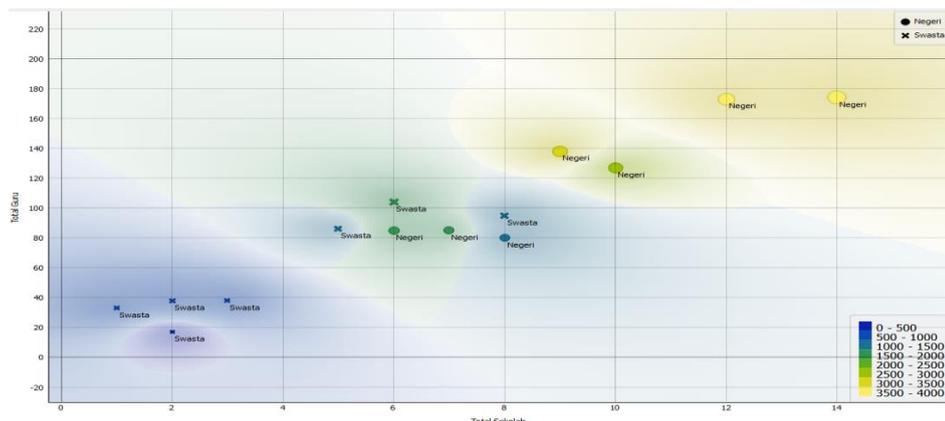
status sekolah di Pangkalpinang, termasuk total sekolah, guru, peserta didik, dan total ruang kelas SD. Data tersebut dimodelkan menggunakan algoritma K-Means, dengan 6 atribut yaitu kecamatan, status, total sekolah, total guru, total peserta didik, total ruang kelas dan 14 record data kecamatan untuk klasifikasi dan prediksi.

Tabel 1. Prediksi dan Klasifikasi Dataset Hasil Pilihan Data

Kecamatan	Status	Total Sekolah	Total Guru	Total Peserta Didik	Total Ruang Kelas
Gerunggang	Negeri	14	174	4087	118
Gerunggang	Swasta	5	86	1202	61
Gabek	Negeri	9	138	3030	96
Gabek	Swasta	6	104	1791	63
Bukit Intan	Negeri	10	127	2847	84
Bukit Intan	Swasta	3	38	631	29
Girimaya	Negeri	8	80	1433	61
Girimaya	Swasta	8	95	1387	75
Pangkal Balam	Negeri	7	85	1668	64
Pangkal Balam	Swasta	2	17	178	12
Rangkui	Negeri	12	173	3553	125
Rangkui	Swasta	2	38	664	24
Taman Sari	Negeri	6	85	1760	61
Taman Sari	Swasta	1	33	536	20

Data pada tabel 1 di atas memberikan informasi yang relevan tentang situasi pendidikan di berbagai kecamatan, dengan pemisahan antara sekolah negeri dan swasta. Analisis lebih lanjut atau visualisasi data mungkin diperlukan untuk mendapatkan wawasan yang lebih mendalam tentang pola atau tren yang mungkin terjadi di wilayah tersebut.

3.2 Hasi Dari Proses Preprocessing Data

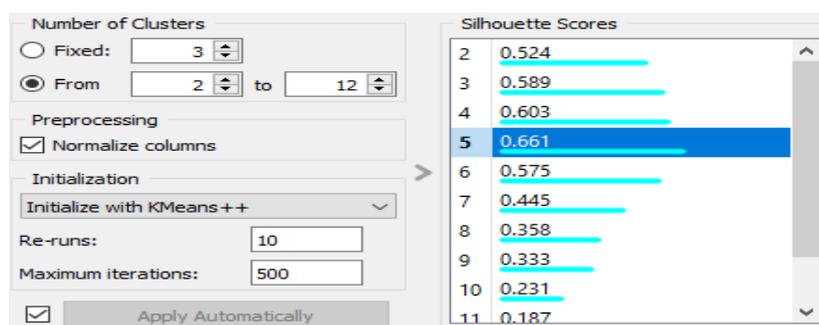


Gambar 3. Plotting Data

Dataset yang tercantum dalam tabel 1 diproses untuk preprocessing setelah pemilihan data selesai. Semua metrik, termasuk total sekolah, total guru, total peserta didik, dan total ruang kelas, dihubungkan menggunakan teknik impute berdasarkan rata-rata atau frekuensi tertinggi. Seperti yang ditunjukkan pada gambar 3, rata-rata adalah jarak rata-rata antara elemen di masing-masing kelompok dan elemen lainnya di kelompok yang berbeda.

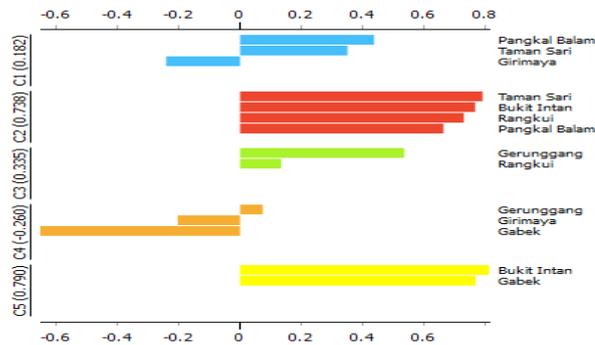
3.3 Hasil Data Mining

Metode K-Means Clustering mengatur data berdasarkan hasil preprocessing data dan status sekolah, seperti yang ditunjukkan pada gambar di bawah.



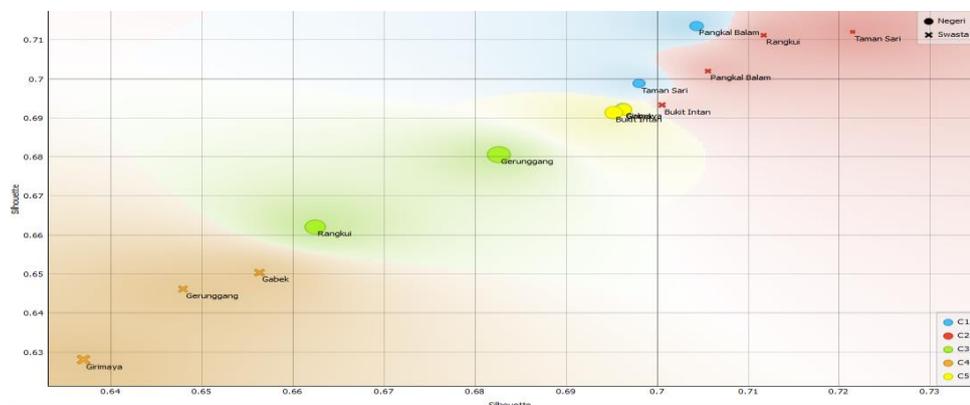
Gambar 4. Kmeans Konfigurasi

Konfigurasi metode clustering K-Means yang digunakan dalam penelitian ini dapat dilihat pada gambar 4 di atas. Metode ini memberikan label numerik pada kelompok yang terdiri dari dua hingga dua belas titik data yang dikumpulkan dari hasil prapemrosesan data sebelumnya. Selanjutnya, untuk menghitung kedekatan dengan titik pusat terdekat atau centroid, gunakan jumlah iterasi sebanyak mungkin, mulai dari 10 hingga 500. Ini harus dilakukan dengan inisialisasi K-Means++ dan mempertimbangkan rata-rata atau frekuensi tertinggi dari empat atribut: total sekolah, total guru, total peserta didik, dan total ruang kelas. Kelompok cluster data pengukuran simulasi digambarkan dengan plot silhouette menggunakan jarak euclidean di Gambar 5.



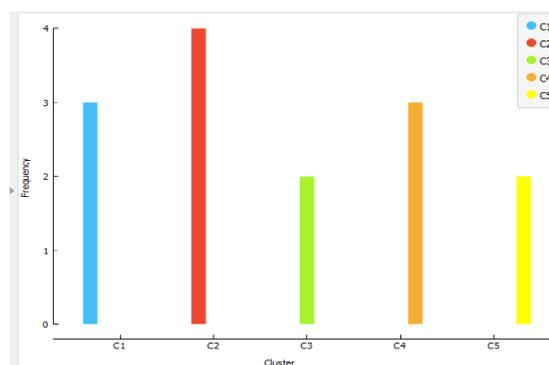
Gambar 5. Plot Silhouette

Studi ini menggunakan lima kluster, dan hasil dari setiap kluster akan menentukan sekolah mana yang termasuk dalam kategori yang memiliki jumlah terbesar, jumlah terkecil, atau jumlah terkecil dari empat atribut tersebut : total sekolah, total guru, total peserta didik dan total ruang kelas berdasarkan kecamatan, dan status sekolah di kota Pangkalpinang. Gambar 3 termasuk kluster 1 sebanyak 3 kecamatan berdasarkan status yaitu, pangkal balam, taman sari dan girimaya. Kluster 2 sebanyak 4 kecamatan berdasarkan status yaitu, taman sari, bukit intan, rangkui dan pangkal balam. Kluster 3 sebanyak 2 kecamatan berdasarkan status yaitu, gerunggung dan rangkui. Kluster 4 sebanyak 3 kecamatan berdasarkan status yaitu, gerunggung, girimaya dan gabek. Kluster 5 sebanyak 2 kecamatan berdasarkan status yaitu, bukit intan dan gabek seperti tergambar pada gambar 6 di bawah ini.



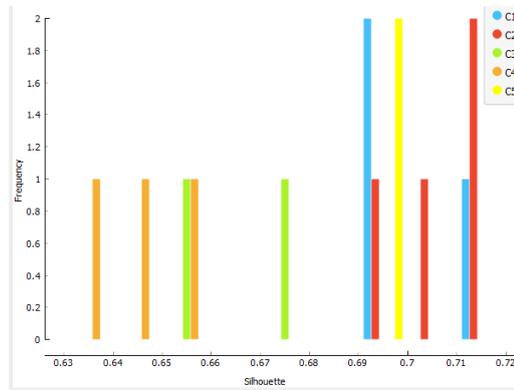
Gambar 6. Plot silhouette cluster dari kecamatan

Gambar 7 menunjukkan hasil kumulatif data yang didistribusikan berdasarkan jumlah data yang ada di cluster Sekolah Dasar di Kota Pangkalpinang berdasarkan status dan kecamatan. Gambar 8 menunjukkan bahwa Kluster 2, yang terdiri dari empat kecamatan berbasis pada status sekolahnya memiliki jumlah kluster terbesar.



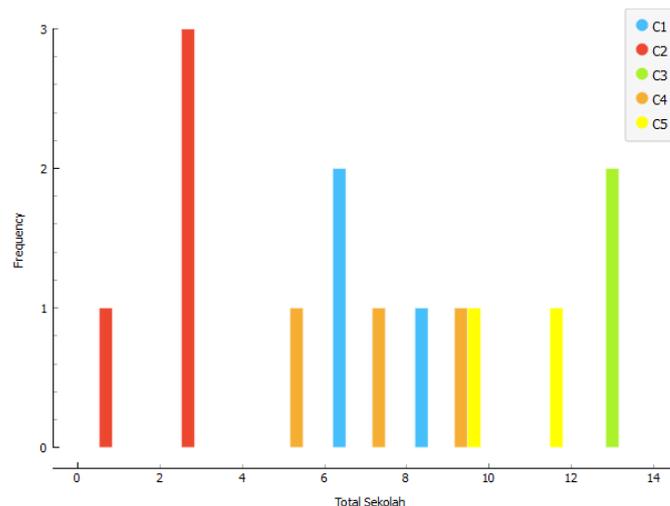
Gambar 7. Jumlah Kluster

Kemudian pada gambar 8 didasari dari nilai sillhouette yang kurang 0,71 dari 3 data kecamatan sekolah dasar yang didapatkan dari status sekolah dengan persentase 100% sehingga dapat diasumsikan bahwa klaster 1 memiliki persentase 21,43%, klaster 2 memiliki persentase 28,57%, klaster 3 memiliki persentase 14,28%, klaster 4 memiliki persentase 21,42%, dan klaster 5 memiliki persentase 14,29%.



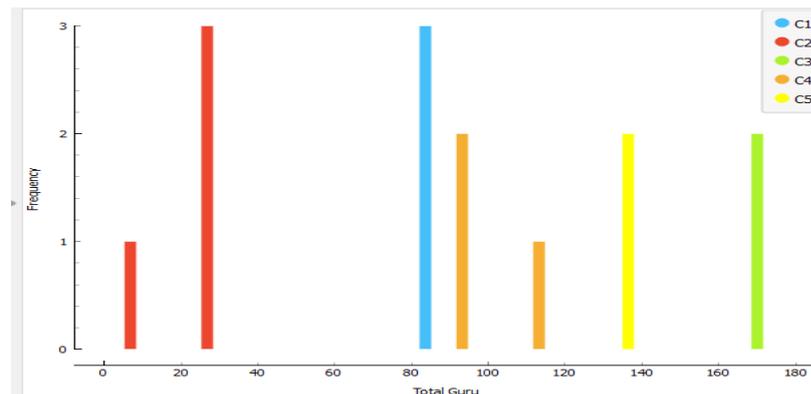
Gambar 8. Silhouette Score

Pada kumulatif total sekolah pada gambar 9 dengan persentase 100% maka dapat diasumsikan klaster 1 memiliki persentase 21,43%, klaster 2 memiliki persentase 28,25%, klaster 3 memiliki persentase 14,29%, klaster 4 memiliki persentase 21,42%, dan klaster 5 memiliki persentase 14,28% dari atribut total sekolah pada gambar 7 dengan persentase 100%.



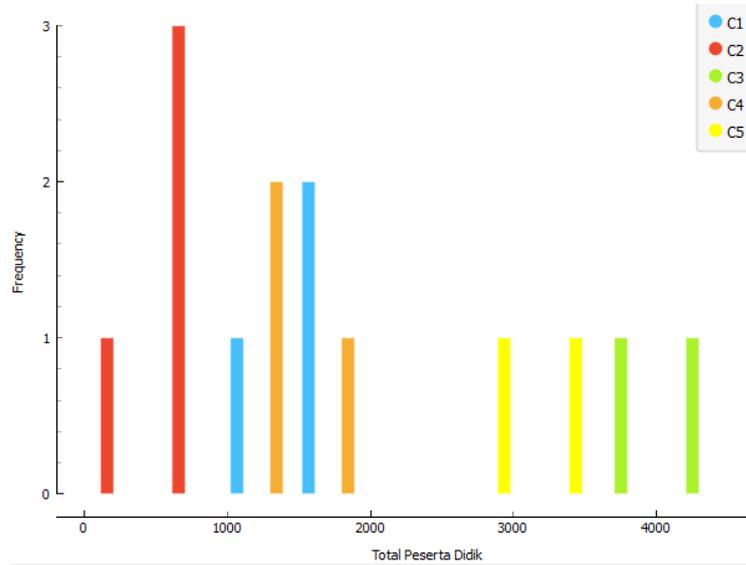
Gambar 9. Total Sekolah

Pada kumulatif berdasarkan total guru pada gambar 10 di bawah kurang dari 180 dari 14 record data sekolah dasar berdasarkan kecamatan dengan persentase 100% maka dapat diasumsikan klaster 1 memiliki persentase 21,43%, klaster 2 memiliki persentase 28,57%, klaster 3 memiliki persentase 14,29%, klaster 4 memiliki persentase 21,43%, dan klaster 5 memiliki persentase 14,29% masing-masing didasari atribut total guru yang tidak sampai 180 dari 14 record data sekolah dasar berdasarkan kecamatan dengan persentase 100%.



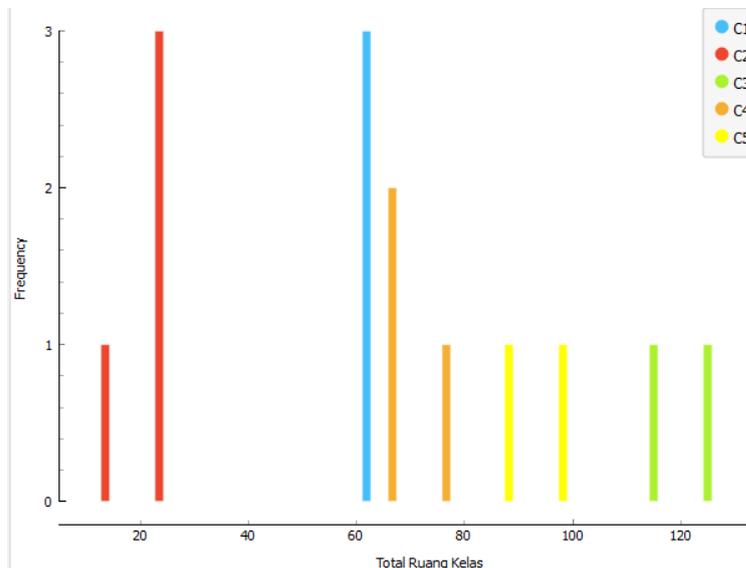
Gambar 10. Total Guru

Kumulatif berdasarkan total peserta didik pada gambar 11 kurang dari 4000 dari 14 record data sekolah dasar berdasarkan kecamatan di kota Pangkalpinang dapat dilihat masing-masing kluster mendapatkan nilai kluster 1 memiliki tingkat persentase 21,43%, kluster 2 memiliki tingkat persentase 28,57%, kluster 3 memiliki tingkat persentase 14,28%, kluster 4 memiliki persentase 21,43%, dan kluster 5 memiliki persentase 14,29%.



Gambar 11. Total Peserta Didik

Didasari atribut total ruang kelas pada gambar 12 di bawah yang tidak sampai 120 data berdasarkan 14 record data sekolah dasar menurut kecamatan di Pangkalpinang dengan tingkat capaian 100%, bisa disepakati bahwa kluster 1 memiliki presentase 21,43%, kluster 2 memiliki presentase 28,57%, dan kluster 3 memiliki presentase 14,28%. Kesimpulannya, kluster 2 memiliki presentase terbesar, yaitu 28,57%.



Gambar 12. Total Ruang Kelas

Hasil plotting skor silhouette pada kluster 1, Girimaya, Pangkal Balam, dan Taman Sari ditunjukkan pada Tabel 2 di bawah ini. Dengan demikian, dapat diprediksi bahwa Girimaya terkelompok dalam kluster 1 dengan skor siluet 0.692766, Pangkal Balam terkelompok dalam kluster 1 dengan skor siluet 0.711739 dan Taman Sari terkelompok dalam kluster 1 dengan status sekolah negeri dengan skor silhouette 0.711739.

Tabel 2. Silhouette score kluster 1

Kecamatan	Cluster	Silhouette	Status	Total Sekolah	Total Guru	Total Peserta Didik	Total Ruang kelas
Girimaya	C1	0.692766	Negeri	8	80	1433	61
Pangkal Balam	C1	0.711739	Negeri	7	85	1668	64
Taman Sari	C1	0.695238	Negeri	6	85	1760	61

Menurut Tabel 3 di bawah ini, skor silhouette untuk kluster 2 Bukit Intan, Pangkal Balam, Rangkui, dan Taman Sari diplot. Hasil plot ini menunjukkan bahwa Bukit Intan terkelompok dalam kluster 2 dengan skor silhouette 0,697004, Pangkal Balam terkelompok dalam kluster 2 dengan skor silhouette 0,705043, dan Rangkui terkelompok dalam kluster 2 beratribut status sekolah swasta dengan skor silhouette 0,705043.

Tabel 3. Sillhouette score kluster 2

Kecamatan	Cluster	Sillhouette	Status	Total Sekolah	Total Guru	Total Peserta Didik	Total Ruang kelas
Bukit Intan	C2	0.697004	Swasta	3	38	631	29
Pangkal Balam	C2	0.705043	Swasta	2	17	178	12
Rangkui	C2	0.714424	Swasta	2	38	664	24
Taman Sari	C2	0.714494	Swasta	1	33	536	20

Pada Tabel 4 di bawah, interpretation Sillhouette score kluster 3 berdasarkan kecamatan Gerunggang dan Rangkui dapat diprediksi bahwa Gerunggang terkelompok kluster 3 dengan skor silhouette 0.678648 dan Rangkui terkelompok kluster 3 dengan skor silhouette 0.657787 berdasarkan status sekolah negeri.

Tabel 4. Sillhouette score kluster 3

Kecamatan	Cluster	Sillhouette	Status	Total Sekolah	Total Guru	Total Peserta Didik	Total Ruang kelas
Gerunggang	C3	0.678648	Negeri	14	174	4087	118
Rangkui	C3	0.657787	Negeri	12	173	3553	125

Pada Tabel 5 di bawah, interpretation Sillhouette score kluster 4 berdasarkan kecamatan Gerunggang, gabek dan girimaya dapat diprediksi bahwa Gerunggang terkelompok dengan kluster 4 dengan status sekolah swasta dengan skor silhouette 0.643121, Gabek terkelompok dengan kluster 4 dengan status sekolah swasta yang dengan skor silhouette 0.654157 dan Girimaya terkelompok dengan kluster 4 dengan status sekolah swasta dengan skor silhouette 0.634809.

Tabel 5. Sillhouette score kluster 4

Kecamatan	Cluster	Sillhouette	Status	Total Sekolah	Total Guru	Total Peserta Didik	Total Ruang kelas
Gerunggang	C4	0.643121	Swasta	5	86	1202	61
Gabek	C4	0.654157	Swasta	6	104	1791	63
Girimaya	C4	0.634809	Swasta	8	95	1387	75

Tabel 6 di bawah ini menunjukkan interpretasi skor silhouette pada kluster 5 Gabek dan Bukit Intan. Hasil plotting silhouette score menunjukkan bahwa kecamatan Gabek memiliki skor silhouette 0.691555 dan kecamatan Bukit Intan memiliki skor silhouette 0.694601 dengan status sekolah negeri.

Tabel 6. Sillhouette score kluster 5

Kecamatan	Cluster	Sillhouette	Status	Total Sekolah	Total Guru	Total Peserta Didik	Total Ruang kelas
Gabek	C5	0.691555	Negeri	9	138	3030	96
Bukit Intan	C5	0.694601	Negeri	10	127	2847	84

4. KESIMPULAN

Berlandaskan temuan yang diperoleh dari penelitian yang memfokuskan pada pemodelan data mengenai status sekolah dasar di 7 kecamatan di kota Pangkalpinang, diambil dari basis data Kemendikbud yang terdokumentasi pada portal <https://dapo.kemdikbud.go.id/>. Dalam upaya untuk mengintegrasikan desain pemodelan pembelajaran mesin, hipotesis dibuat yang menunjukkan bahwa penggunaan algoritma klasifikasi K-Means untuk pemodelan dapat digunakan secara efektif untuk melakukan klasifikasi dan prediksi karena dapat secara efisien menangani jumlah data yang besar dan cocok untuk data yang memiliki struktur kluster yang jelas. Ini membuatnya efisien untuk sejumlah aplikasi praktis. Ini berkaitan dengan menentukan berapa banyak kluster sekolah yang dimasukkan ke dalam kategori terbanyak dan paling minim berdasarkan empat karakteristik utama yang dimiliki oleh masing-masing sekolah, yakni total sekolah, total guru, total peserta didik, serta total ruang kelas, yang disesuaikan dengan lokasi kecamatan dan status sekolah masing-masing. Jumlah kluster terbesar terlihat dari hasil plot kumulatif dengan nilai atribut paling rendah adalah kluster 2, yakni 4 kecamatan berdasarkan status sekolah swasta dengan nilai atribut total sekolah paling rendah yaitu hanya 8 sekolah, atribut total guru berjumlah 126 tenaga pengajar guru, atribut total peserta didik dengan jumlah 2009 siswa dan hanya memiliki 85 ruang kelas, lalu dengan skor silhouette yang kurang dari 0,71 dengan 14 record data sekolah dasar berdasarkan kecamatan dengan tingkat capaian 100%, bisa disepakati bahwa kluster 1 memiliki nilai 21,43% , kluster 2

memiliki nilai 28,57%, klaster 3 memiliki nilai 14,28%, klaster 4 memiliki nilai 21,42 % dan klaster 5 memiliki nilai 14,29%. Sedangkan jumlah cluster paling sedikit dengan nilai atribut paling tinggi adalah klaster 3, yakni 2 kecamatan berdasarkan status sekolah negeri dengan nilai atribut total sekolah paling tinggi yaitu terdapat 26 sekolah dasar, 347 tenaga pengajar guru, total peserta didik dengan jumlah 7640 siswa dan memiliki 243 ruang kelas. Dengan akumulasi jumlah klaster yang dihasilkan, Pemerintah Kota Pangkalpinang sebagai entitas penyelenggara pendidikan dan potensi peserta didik dapat melakukan penentuan dan klasifikasi sekolah-sekolah tingkat dasar. Hal ini mencakup aspek peningkatan kuantitas sekolah dasar di daerah-daerah yang mengalami kekurangan sekolah dasar, penambahan tenaga pendidik untuk sekolah yang membutuhkan peningkatan guru, peningkatan jumlah siswa bagi sekolah yang memerlukan alokasi siswa yang lebih sedikit berdasarkan wilayah tertentu, dan peningkatan fasilitas ruang kelas di sekolah yang tidak memiliki ruang untuk pertemuan tatap muka dalam proses belajar.

REFERENCES

- [1] T. Hartati, O. Nurdiawan, and E. Wiyandi, "Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," *J. Sains Teknol. Transp. Marit.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30.
- [2] D. T. Worung, S. R. U. A. Sompie, and A. Jacobus, "Implementasi K-Means dan K-NN pada Pengklasifikasian Citra Bunga," *J. Tek. Inform.*, vol. 15, no. 3, pp. 217–222, 2020, [Online]. Available: <https://ejournal.unsrat.ac.id/v3/index.php/informatika/article/view/31965>.
- [3] B. Mahesh, "Machine learning algorithms-a review," *Int. J. Sci. Res. (IJSR)*. [Internet], vol. 9, no. 1, pp. 381–386, 2020.
- [4] J. Bell, "What is machine learning?," *Mach. Learn. City Appl. Archit. Urban Des.*, pp. 207–216, 2022.
- [5] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *J. Artif. Intell. Res.*, vol. 70, pp. 245–317, 2021.
- [6] N. Li, M. Shepperd, and Y. Guo, "A systematic review of unsupervised learning techniques for software defect prediction," *Inf. Softw. Technol.*, vol. 122, p. 106287, 2020.
- [7] D. Yolanda, M. H. Hersyah, E. Marozi, and others, "Implementasi Metode Unsupervised Learning Pada Sistem Keamanan Dengan Optimalisasi Penyimpanan Kamera IP," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1099–1105, 2021.
- [8] F. Marisa, B. Ariefia, A. L. Maukar, and others, "Pendeteksian Daerah (Provinsi) Rawan Covid19 Dengan Metode Unsupervised Learning & Algoritma K-Medoids," *J. Teknol. Inf. dan Komun.*, vol. 12, no. 1, pp. 17–21, 2021.
- [9] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE access*, vol. 8, pp. 80716–80727, 2020.
- [10] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. Dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [11] A. Z. Saputra, N. Suarna, and G. D. Lestari, "Klasterisasi Nilai Ujian Sekolah Menggunakan Metode Algoritma K-Means," *J. Janitra Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 1–9, 2023, doi: 10.25008/janitra.v3i1.153.
- [12] A. Septianingsih, "Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 1, pp. 224–241, 2022.
- [13] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022.
- [14] Y. B. Pratama and N. P. Dalimunthe, "Implementasi Teknik Computer Vision Untuk Deteksi Viridiplantae Pada Lahan Pasca Tambang," vol. 3, no. 1, pp. 64–72, 2022, doi: 10.47065/bulletincsr.v3i1.193.
- [15] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhajja, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny.)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.
- [16] H. Hu, J. Liu, X. Zhang, and M. Fang, "An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis," *Pattern Recognit.*, vol. 139, Jul. 2023, doi: 10.1016/j.patcog.2023.109404.
- [17] R. M. Adnan, P. Khosravinia, B. Karimi, and O. Kisi, "Prediction of hydraulics performance in drain envelopes using Kmeans based multivariate adaptive regression spline," *Appl. Soft Comput.*, vol. 100, p. 107008, 2021.
- [18] E. Khaledian, S. Pandey, P. Kundu, and A. K. Srivastava, "Real-time synchrophasor data anomaly detection and classification using isolation forest, kmeans, and loop," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2378–2388, 2020.
- [19] S. Rusmayana, A. Faqih, and A. Bahtiar3, "PENERAPAN METODE ALGORITMA K-MEANS DALAM PEMETAAN PESERTA DIKLAT KETERAMPILAN PELAUT DI SMK N 1 MUNDU," *J. Sist. Inf. dan Manaj.*, vol. 10, no. 2, 2022.
- [20] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.