

# PENERAPAN *CLUSTERING* PADA RATA-RATA LAMA SEKOLAH DI PROVINSI INDONESIA MENGUNAKAN ALGORITMA *K-MEANS* DAN *FUZZY C-MEANS*

**Jordy Apriyanto**

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,  
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia  
*E-mail: jordy.535210073@stu.untar.ac.id,*

## ABSTRAK

Sekolah adalah bagian dari sistem pengajaran yang ditujukan untuk memfasilitasi perkembangan mutu kapasitas manusia. Untuk mengelompokkan data ke dalam grup atau kluster, menggunakan dua algoritma clustering, yaitu *K-Means* dan FCM. Dengan pengumpulan dataset dari Badan Pusat Statistik yang terdiri dari 582 provinsi/kabupaten/kota di Indonesia. Capaian dari kajian ini menyatakan bahwa algoritma FCM menghasilkan rata-rata *silhouette score* sebesar 0.5369645 dengan melakukan eksperimen sebanyak 10 kali. *K-Means* lebih sederhana dan cepat, cocok untuk dataset yang besar sedangkan FCM memberikan fleksibilitas dengan keanggotaan *fuzzy*, sehingga cocok untuk dataset dengan batas kelompok yang tidak jelas. Kedua algoritma memberikan hasil yang stabil dalam eksperimen akan tetapi algoritma *clustering* FCM merupakan algoritma yang sedikit lebih unggul dari pada algoritma *K-Means* pada hal *silhouette score*, namun perbedaannya tidak signifikan.

**Kata kunci**— Sekolah, *Clustering*, *K-Means*, FCM, *Silhouette Score*

## ABSTRACT

The school is part of the teaching system aimed at facilitating the development of human capacity quality. To cluster data into groups or clusters, two clustering algorithms are used: *K-Means* and FCM. The dataset is collected from the Central Statistics Agency, consisting of 582 provinces/regencies/cities in Indonesia. The findings of this study indicate that the FCM algorithm produces an average *silhouette score* of 0.5369645 after conducting 10 experiments. *K-Means* is simpler and faster, making it suitable for large datasets, while FCM provides flexibility with fuzzy membership, making it more suitable for datasets with unclear group boundaries. Both algorithms yield stable results in the experiments; however, the FCM clustering algorithm is slightly superior to the *K-Means* algorithm in terms of *silhouette score*, although the difference is not significant.

**Keywords**— School, *Clustering*, *K-Means*, FCM, *Silhouette Score*

## 1. PENDAHULUAN

Pendidikan merujuk pada hak yang harus disetujui oleh setiap orang. Pendidikan bisa didapatkan dengan jalur resmi maupun alternatif. Salah satu metode formal yang umumnya dipilih individu untuk memperoleh pendidikan adalah dengan mengikuti sistem pendidikan di sekolah [1]. Perkembangan teknologi memberi dampak besar terhadap dunia pendidikan. Fungsi pendidikan tidak sebatas untuk mendapatkan ilmu pengetahuan, tetapi juga untuk membentuk sikap dan perilaku peserta didik. Saat ini, sistem pendidikan di Indonesia lebih mengutamakan aspek pengetahuan tanpa

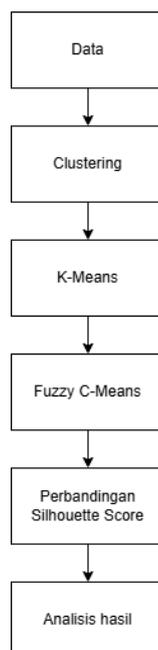
memberikan perhatian yang cukup pada pengembangan nilai-nilai karakter, sehingga diperlukan pembiasaan dan teladan dalam membentuk karakter siswa [2].

Sekolah adalah bagian dari sistem pendidikan yang bermaksud untuk memperbaiki mutu sumber daya manusia. Melalui pendidikan yang diterima oleh generasi muda disekolah, pola pikir dan kreativitas mereka dapat berkembang, yang pada akhirnya berkontribusi pada kemajuan negara, peningkatan kesejahteraan, dan pertumbuhan ekonomi [3]. Pengajaran dapat dilakukan pada tingkat sekolah dasar sebagai tahap awal pengajaran secara resmi di Indonesia. Sekolah dasar menjadi lembaga yang menjalankan pendidikan selama kurang lebih enam tahun untuk siswa-siswa nya yang berusia antara 6 hingga 12 tahun [4].

Dalam kasus ini, dibahas topik mengenai rata-rata lama sekolah diberbagai provinsi di Indonesia dengan menggunakan metode Clustering [5]. Clustering adalah metode pada data mining yang digunakan dengan tujuan untuk mengklasifikasikan statistik ke dalam golongan-golongan yang mempunyai persamaan dari suatu fitur khusus [6]. Metode clustering yang diterapkan adalah K-Means dan FCM. K-Means merupakan metode yang dimanfaatkan untuk mengelompokkan data berdasarkan jaraknya dari pusat kelompok, yang termasuk dalam salah satu algoritma kelompokisasi data. Metode ini bisa diterapkan, dijalankan, dan diproses dengan cepat serta mudah disesuaikan [7]. FCM adalah metode pengelompokkan data yang menentukan tingkat keanggotaan setiap titik data dalam sebuah kelompok [8]. FCM adalah algoritma yang bersifat iteratif, yang melibatkan proses pengulangan dalam pengelompokkan informasi. Tujuan dari FCM adalah untuk menemukan pusat kelompok yang digunakan dalam proses penentuan data yang termasuk dalam kelompok tertentu [9].

## 2. METODE PENELITIAN

Kajian ini meliputi kategori kajian eksperimen yang diawali pada tahap pengumpulan statistik. Data dikumpulkan dengan cara mengambil informasi dari sebuah situs web milik pemerintah. Tahap selanjutnya merupakan melakukan langkah clustering, yang dilakukan memanfaatkan algoritma K-Means dan FCM.



**Gambar 1.** Flowchart Metode Penelitian

## 2.1 Data

Informasi yang dipakai pada eksperimen berasal dari website Badan Pusat Statistik. Dataset ini terdiri dari 582 provinsi/kabupaten/kota di Indonesia. Dataset ini berisikan rata-rata lama sekolah pada tahun 2022, 2023, dan 2024 pada sestiap provinsi/kabupaten/kota di Indonesia.

## 2.2 Clustering

Pada pengelolaan data rata-rata lama sekolah di provinsi,kabupaten, atau kota di Indonesia, metode algoritma yang diterapkan untuk mengelompokkan data rata-rata lama pendidikan adalah dengan menggunakan teknik K-Means dan FCM.

### 2.2.1 K-Means

K-Means merujuk pada metode pengelompokan hal yang cukup sederhana yang memisahkan kumpulan statistik berubah menjadi beberapa grup dengan banyaknya kelompok yang telah diputuskan [10]. Algoritma ini bertujuan untuk mengategorikan statistik pada sejumlah kategori, dimana statistic pada sebuah kategori mempunyai karakteristik yang serupa diantara mereka dan berbeda dengan kategori lainnya [11]. Berikut adalah langkah-langkah pada metode K-Means: [12]

1. Menginput data yang akan dikelompokkan
2. Pilih kelompok yang diperlukan
3. Pilih secara acak sejumlah data yang setara dengan jumlah kelompok sebagai titik pusat
4. Tentukan batas dalam statistik dan titik pusat memanfaatkan rumus berikut pada kesamaan (1):

$$(a, b) = \sqrt{(xa - yb)^2 + (xa - yb)^2 + (xn - yn)^2} \quad (1)$$

Keterangan:

D (a, b) = selisih posisi antara statistik ke-i dan fokus kelompok ke-j

Xku = nilai statistik ke-u dengan karakteristik ke-k

Xko = nilai fokus kelompok ke-o pada karakteristik ke-k

5. Perbarui posisi focus kelompok berdasarkan pembaruan keanggotaan kelompok yang baru
6. Jika posisi fokus kelompok sudah stabil dan konsisten, oleh karena itu tahap klastering berakhir, jika masih ada perubahan dengan demikian lakukan Kembali tahap ke-4 hingga fokus kelompok tetap

Langkah berikutnya adalah mengidentifikasi statistik menurut fokus kelompok yang paling dekat, lalu mengalkulasi posisi inti kelompok yang terbaru menggunakan konsep pada persamaan (2): [13]

$$C_i = \frac{\sum d_i}{n_k} \quad (2)$$

Keterangan:

Ci = merujuk pada pusat kelompok ke-i

d = data pada titik ke-i

n = menunjukkan jumlah kelompok k

### 2.2.2 Fuzzy C-Means

FCM merujuk pada metode klastering data yang menggunakan bobot kelas keikutsertaan untuk menetapkan kelompok yang sesuai bagi setiap data [14]. Data dapat digabungkan berdasarkan tingkat keikutsertaan yang nilainya berkisar antara satu dengan yang lain, dan beberapa tipe statistik hanya menunjukkan keanggotaan Sebagian [15]. Semakin tinggi nilai keikutsertaan suatu statistik, semakin besar kelas keikutsertaannya dan kebalikannya lebih kecil bobot keikutsertaan statistik, menjadi lebih lemah bobot keikutsertaannya [16]. Berikut ini adalah algoritma untuk klasterisasi FCM: [17]

1. Masukkan statistik yang terdapat dalam kelompok X, dalam bentuk matriks berukuran n x p (n adalah banyaknya sampel statistik dan p adalah banyaknya karakteristik pada seluruh sampel).

- Xko merujuk pada statistic uji coba ke-k ( $k = 3, 4, \dots, n$ ) dan karakteristik ke-o ( $o = 4, 5, 6, 7, \dots, m$ )
2. Menentukan input yang dibutuhkan dalam perhitungan FCM, yaitu:
    - a. Banyaknya kelompok ( $c$ )
    - b. Eksponen ( $m$ )
    - c. Batas pengulangan terbesar
    - d. Nilai kesalahan terkecil
    - e. Kinerja sasaran
    - f. Pengulangan dasar
  3. Menghasilkan bobot secara tidak teratur. Dapat dilihat pada persamaan (3):

$$U_0 = \begin{pmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \mu_{1c}(x_c) \\ \mu_{11}(x_1) & \mu_{12}(x_2) & \mu_{nc}(x_c) \end{pmatrix} \quad (3)$$

Ketentuan yang perlu dipenuhi oleh matriks partisi dalam pengelompokan Fuzzy bisa diamati di kesamaan (4):

$$\mu_{ik} = [0, 1]; (1 \leq i \leq c; 1 \leq k \leq n)$$

$$\sum_{i=1}^n \mu_{ik} = 1; 1 \leq i \leq c$$

$$0 < \sum_{i=1}^c \mu_{ik} < c; 1 \leq k \leq n$$
(4)

Kemudian, hitung total untuk setiap atribut, Dapat dilihat pada persamaan (5) :

$$Q_j = \sum_{i=1}^c \mu_{ik} \quad (5)$$

Dengan  $o = 4, 5, \dots, m$  (banyaknya baris pada matriks yang berisi nilai acak dapat disesuaikan dengan jumlah ketentuan yang telah ditentukan). Selanjutnya dapat dilihat pada persamaan (6):

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad (6)$$

4. Memperkirakan kelompok ke-k, dengan  $k = 1, 2, \dots, c$ ; dan  $j = 1, 2, \dots, m$ . Dapat dilihat pada persamaan (7) dan (8):

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^m * X_{kj})}{d \sum_{k=1}^n (\mu_{ik})^m} \quad (7)$$

$$V = \begin{matrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{c1} & \cdots & v_{cm} \end{matrix} \quad (8)$$

5. Nilailah peran sasaran pada pengulangan ke-t, Pt, dengan memanfaatkan rumus setelah ini. Dapat dilihat pada persamaan (9):

$$P_t = \sum_{k=1}^n \sum_{i=1}^c \left( \sum_{j=1}^m (X_{kj} - V_{ij})^2 \right) (\mu_{ik})^2 \quad (9)$$

6. Menganalisis perubahan yang terjadi pada matriks partisi. Dapat dilihat pada persamaan (10):

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (10)$$

### 2.3 Pengujian Silhouette Score

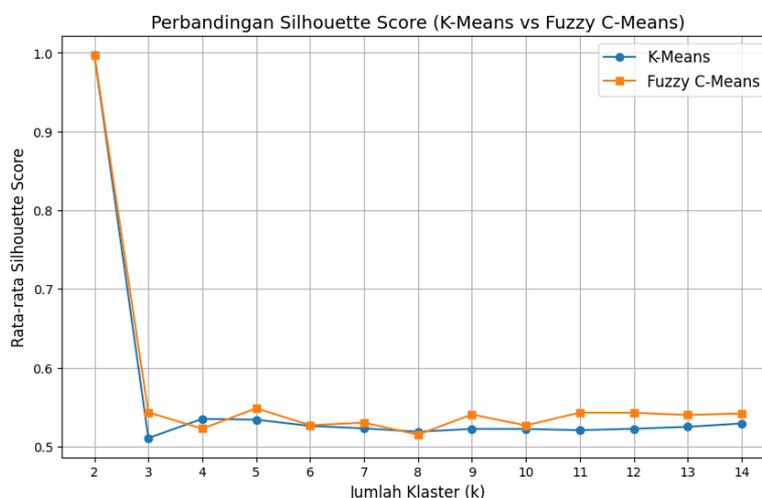
*Silhouette score* merupakan kombinasi dari dua pendekatan, yaitu pendekatan kohesi dan pemisahan. Kohesi bertujuan untuk menilai sejauh mana objek-objek dalam satu kelompok saling berdekatan, sementara pemisahan bertujuan untuk menilai sejauh mana jarak antara kelompok yang satu dengan kelompok lainnya [19]. *Silhouette score* didapatkan dengan mengedepankan rentang antar setiap statistik pada sebuah golongan menggunakan rentang statistik ke golongan selain itu [20]. *Silhouette score* akan menghasilkan nilai yang menggambarkan tingkat kecocokan atau kesesuaian objek yang dianalisis [21]. Berikut ada langkah-langkah untuk menghitung silhouette score, dapat dilihat pada persamaan (11): [22]

1. Tentukan nilai a(i) dengan cara menghitung nilai tengah kedekatan pada pusat tersebut melalui semua pusat lain dalam kelompok yang serupa
2. Tentukan nilai b(i) dengan cara menyusun nilai tengah kedekatan antara pusat tersebut dengan semua pusat dalam kelompok paling dekat
3. Gunakan rumus berikut untuk menghitung silhouette score atau nilai s(i) untuk titik ke-i. Dapat diamati dalam kesamaan (11)

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (11)$$

## 3. HASIL DAN PEMBAHASAN

Kajian ini mengulas mengenai perbandingan clustering memanfaatkan kedua algoritma yaitu K-Means dan algoritma FCM. Dataset ini terdiri dari 582 provinsi/kabupaten/kota di Indonesia. Gambar 2 merupakan perbandingan silhouette score dari kedua algoritma.



**Gambar 2.** Perbandingan Silhouette Score K-Means VS Fuzzy C-Means

Selanjutnya, melakukan eksperimen terhadap kedua algoritma yang digunakan dengan melakukan eksperimen sebanyak 10 kali. Hasil eksperimen dari kedua pendekatan yang dipakai bisa diamati pada tabel 1.

**Tabel 1.** Hasil Eksperimen Kedua Metode

| Eksperimen ke | Silhouette Score K-Means | Silhouette Score Fuzzy C-Means |
|---------------|--------------------------|--------------------------------|
| 1             | 0.510744                 | 0.543269                       |
| 2             | 0.535090                 | 0.522655                       |
| 3             | 0.534114                 | 0.548534                       |
| 4             | 0.526110                 | 0.527046                       |
| 5             | 0.523015                 | 0.530226                       |
| 6             | 0.518797                 | 0.515079                       |
| 7             | 0.522359                 | 0.540664                       |
| 8             | 0.522289                 | 0.526761                       |
| 9             | 0.520706                 | 0.543024                       |
| 10            | 0.522541                 | 0.542881                       |
| Rata-rata     | 0.5235765                | 0.5340139                      |

Berdasarkan hasil eksperimen pada tabel 1, kedua metode menghasilkan nilai silhouette score yang hampir sama secara rata-rata, dengan sedikit keunggulan pada algoritma FCM. Variasi nilai pada setiap eksperimen terlihat kecil, yang menunjukkan bahwa hasil kelompokisasi relatif konsisten untuk kedua algoritma. Algoritma K-Means memiliki variasi nilai silhouette score yang cukup kecil, dengan nilai tertinggi 0.535090 (pada eksperimen ke-2) dan nilai terendah 0.510744 (pada eksperimen ke-1) sedangkan FCM menunjukkan hasil yang konsisten, dengan nilai tertinggi 0.548534 (pada eksperimen ke-3) dan nilai terendah 0.515079 (pada eksperimen ke-6). Kedua algoritma memberikan hasil yang stabil dalam eksperimen.

#### 4. KESIMPULAN

Kesimpulan yang diperoleh dari hasil clustering nilai tengah lama sekolah di provinsi, kabupaten, atau kota di Indonesia dengan menggunakan dua algoritma yang berbeda-beda, menunjukkan performa yang kompetitif dan stabil dengan nilai silhouette score rata-rata yang identik. Namun, algoritma FCM sedikit lebih unggul dibandingkan menggunakan metode K-Means. Algoritma pertama, mengaplikasikan algoritma K-Means dan menghasilkan nilai tengah indeks score sama dengan 0.5235765 dan algoritma kedua, dengan menggunakan algoritma FCM dengan menghasilkan nilai tengah indeks score sama dengan 0.5340139. Sehingga algoritma clustering FCM merupakan metode yang mempunyai kinerja sedikit lebih optimal dari pada metode K-Means pada hal silhouette score. Namun, perbedaannya tidak signifikan, K-Means lebih sederhana dan cepat, cocok untuk dataset yang besar sedangkan FCM memberikan fleksibilitas dengan keanggotaan fuzzy, sehingga cocok untuk dataset dengan batas kelompok yang tidak jelas.

#### DAFTAR PUSTAKA

- [1] N. I. Fitri Hayati, "Karakteristik Perkembangan Siswa Sekolah Dasar : Sebuah Kajian Literatur," *Jurnal Pendidikan Tambusai*, vol. 5, 2021.
- [2] Z. H. R. Mitha Amelia, "Impelementasi Pendidikan Karakter Melalui Budaya Sekolah di Sekolah Dasar," *JURNAL BASICEDU*, vol. 5, 2021.
- [3] D. N. Risdo Rolita Simanjorang, "FUNGSI SEKOLAH," *Jurnal Pendidikan Sosial dan Humaniora*, vol. 2, 2023.
- [4] N. I. Riri Zulvira, "Karakteristik Siswa Kelas Rendah Sekolah Dasar," *Jurnal Pendidikan Tambusai*, vol. 5(1), pp. 1846-1851, 2021.

- [5] A. A. N. A. P. W. A. W. E. I. Siti Hajar, "Penerapan K-Means Clustering Pada Ekspor Minyak Kelapa Sawit Menurut Negara Tujuan," *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, pp. 314-318, 2020.
- [6] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)*, vol. 3 Nomor 1, pp. 40-56, 2024.
- [7] S. D. S. Andri Nofiar Am, "Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering," *Jurnal KomTekInfo*, vol. 5 No 3, pp. 1-9, 2019.
- [8] N. S. F. R. A. Nurfidah Dwitianti, "Penerapan Fuzzy C-Means Cluster dalam Pengelompokan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat," *Faktor Exacta*, vol. 3, pp. 201-209, 2019.
- [9] L. H. Brian Christian, "Penerapan Algoritma Fuzzy C-Means Pada Penentuan Lokasi Gudang Pendukung PT. XYZ," *JURNAL TEKNOLOGI INFORMASI*, vol. 16 No 1, pp. 31-48, 2019.
- [10] A. P. W. I. S. D. I. G. Khairunnissa Fanny Irnanda, "Penerapan K-Means pada Proporsi Individu dengan Keterampilan (Teknologi Informasi Komunikasi) TIK Menurut Wilayah," *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, pp. 452-456, 2019.
- [11] S. D. S. Haris Kurniawan, "Data Mining Menggunakan Metode K-Means Clustering untuk Menentukan Besar Uang Kuliah Tunggal," *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, vol. 1 No 2, pp. 80-89, 2020.
- [12] S. D. Y. Y. Ismail Virgo, "Kelompokisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering," *Jurnal Sistem Informasi dan Teknologi*, vol. 2 No 1, pp. 23-28, 2020.
- [13] D. C. R. N. M. H. Aldi Rizal, "Pengelompokan Karyawan Berdasarkan Kesalehan Menggunakan Perbandingan Fuzzy C-Means, K-Means, dan Probabilistic Distance Clustering," *JURNAL FOURIER*, vol. 11 No 2, pp. 69-77, 2022.
- [14] N. Ulinuha, "Provincial Clustering in Indonesia Based on Plantation," *ITSMART: Jurnal Teknologi dan Informasi*, vol. 9 No 1, 2020.
- [15] M. F. F. F. J. F. Y. I. U. R. G. Hamdi Syukron, "Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 76-83, 2022.
- [16] K. D. H. Anggara Cahya Putra, "Optimalisasi Penyaluran Bantuan Pemerintah Untuk UMKM Menggunakan Metode Fuzzy C-Means," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 474-482, 2021.
- [17] U. A. R. L. H. Ira Halimatuz Zahro, "Implementasi Algoritma Fuzzy C-Means untuk Pengelompokan Provinsi di Indonesia Berdasarkan Kualitas Perguruan Tinggi," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 1, pp. 80-86, 2024.
- [18] S. N. H. N. A. Rozzi Kesuma Dinata, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics Journal*, vol. 5, no. 1, pp. 10-17, 2020.
- [19] U. K. Krisman Pratama Simanjuntak, "Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 7-16, 2021.
- [20] Y. F. R. Christopher Kevin Herijanto, "PERBANDINGAN ALGORITME K-MEANS CLUSTERING DAN FUZZY C-MEANS UNTUK PENENTUAN STRATEGI PROMOSI UNIVERSITAS KATOLIK DARMA CENDIKA," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 3931-3937, 2024.
- [21] S. M. M. R. S. Febby Arisca Zulfani, "Analisis Metode Clustering K-Means pada Zonasi Daerah Terdampak Banjir di Kota Medan dengan Evaluasi Silhouette Coefficient," *Algoritma: Jurnal Matematika, Ilmu pengetahuan Alam, Kebumihan dan Angkasa*, vol. 2, no. 6, pp. 170-181, 2024.
- [22] F. I. S. A. L. A. Syahrul Mulyadi, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mini Batch K-Means Clustering," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1051-1062, 2024.
- [23] L. M. D. G. Hadi Syahputra, "Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means," *Jurnal KomtekInfo*, vol. 9, no. 1, pp. 29-33, 2022.
- [24] M. Mirhan Siregar, "Analisis Perilaku Sosial Masyarakat Terhadap Pandemi Menggunakan Metode Fuzzy C-Means," *Journal of Information Systems and Technology*, vol. 1, no. 2, pp. 75-82, 2024.
- [25] N. K. Zuhail, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," *PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS*, vol. 1, pp. 200-205, 2022.