



Penerapan Algoritma K-Medoids Pada Clustering Penerima Bantuan Pangan Non Tunai (BPNT)

Tiara Ramayanti, Elin Haerani*, Jasril, Lola Oktavia

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia, Kota, Indonesia

Email: ¹tiararamayantii@gmail.com, ^{2,*}elin.haerani@uin-suska.ac.id, ³jasril@uin-suska.ac.id, ⁴lola.oktavia@uin-suska.ac.id

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Abstrak—Bantuan Pangan Non Tunai (BPNT) merupakan bantuan yang disalurkan pemerintah kepada masyarakat kurang mampu untuk meringankan beban keuangan yang semakin membebani kehidupan mereka. Dalam beberapa kasus, ditemukan banyaknya masyarakat yang menerima BPNT tidak tetap sasaran sehingga diperlukan analisis pola karakteristik penerima BPNT agar bantuan menjadi tepat sasaran. Banyaknya kriteria yang harus dipertimbangkan untuk mengetahui masyarakat yang berhak menerima BPNT maka diperlukan algoritma yang tepat dalam menentukan cluster yang tepat dalam menganalisis pola karakteristik. Penelitian ini menerapkan algoritma K-Medoids dalam mengelompokkan data BPNT yang diperoleh dari penelitian Firza Syahputra pada tahun 2020-2021 dengan jumlah 732 data yang terdiri dari 42 atribut, namun hanya digunakan 25 atribut sebagaimana yang terdapat pada Pedoman Umum Bantuan Pangan Non Tunai 2019, sehingga pemerintah dapat mempertimbangkan faktor-faktor yang menjadi karakteristik penerima bantuan. Indikator pengujian Silhouette Coefficient diterapkan untuk memaksimalkan hasil clustering. Hasil clustering yang diperoleh adalah terbentuk 3 pengelompokan dengan koefisien Silhouette 0.4439221599010089. Hasil analisis menunjukkan bahwa clustering yang dilakukan menggunakan algoritma K-medoids adalah dapat diasumsikan bahwa cluster mengelompok berdasarkan pengelompokan Cluster 0 berhak menerima, Cluster 1 dipertimbangkan dan Cluster 2 tidak berhak menerima BPNT.

Kata Kunci: Bantuan Pangan Non Tunai (BPNT); Clustering; K-Medoids; Silhouette Coefficient

Abstract—Bantuan Pangan Non Tunai (BPNT) is assistance distributed by the government to underprivileged communities to ease the financial burden that is increasingly burdening their lives. In a number of cases, it was found that the number of people who received BPNT was not properly targeted, so it was necessary to analyze the pattern of the characteristics of BPNT recipients so that the assistance was right on target. There are many criteria that must be considered to determine the people who are entitled to receive BPNT, so an appropriate algorithm is needed to determine the right cluster when analyzing characteristic patterns. This study applies the K-Medoids algorithm to classify BPNT data obtained from Firza Syahputra's research in 2020–2021, with a total of 732 attributes, so that the government can consider the factors that characterize beneficiaries. Perform tests using the Silhouette coefficient, which is useful for maximizing clustering results. The clustering result is three clusters, and the silhouette coefficient is 0.4439221599010089. The results of the analysis show that clustering performed using the K-Medoids algorithm can assume that clusters are grouped according to grouping: cluster 0 is eligible to receive BPNT, cluster 1 is considered, and cluster 2 is not eligible to receive BPNT.

Keywords: Bantuan Pangan Non Tunai (BPNT); Clustering; K-Medoids; Silhouette Coefficient

1. PENDAHULUAN

Sebagai negara berkembang, kemiskinan menjadi masalah yang sering terjadi di Indonesia [1]. Kemiskinan merujuk pada kondisi ketika individu, keluarga, atau komunitas cukup akses terhadap sumber daya ekonomi, seperti pendapatan, pekerjaan yang layak, pangan, perumahan, pendidikan, layanan kesehatan, dan fasilitas dasar lainnya. Kemiskinan dapat menyebabkan kondisi hidup yang sulit, ketidakstabilan ekonomi, serta terbatasnya kesempatan untuk mencapai potensi penuh dan meningkatkan kesejahteraan. Badan Pusat Statistik (BPS) mencatat penduduk miskin yang dirilis pada September 2022 sebesar 9,57% yaitu mencapai 26,36 juta jiwa [2]. Tingkat pengangguran, pertumbuhan ekonomi, pertumbuhan penduduk, indeks pembangunan manusia, ketimpangan distribusi pendapatan semuanya mempengaruhi kemiskinan pada saat yang bersamaan [3]. Kemiskinan di suatu Negara memang harus dianggap sebagai masalah serius, karena kemiskinan membuat banyak orang Indonesia sulit memenuhi kebutuhan hidup saat ini [4].

Bantuan Pangan Non Tunai (BPNT) adalah bantuan guna meringankan beban keuangan yang semakin membebani kehidupan mereka yang diberikan oleh pemerintah. Untuk menentukan layak atau tidaknya warga negara untuk menerima BPNT pemerintah menggunakan kriteria sebagai berikut. 1) kondisi hidup 2) pendapatan bulanan dan 3) tanggungan. Pemerintah desa dapat memeriksa urutan kelayakan dan membuat keputusan tentang hal itu dengan menganalisis kriteria tersebut yang diatur dalam Peraturan Presiden tentang Penyaluran Bantuan Pendapatan Nomor 63 Tahun 2017 [5].

Untuk melindungi rakyat dari kemiskinan pemerintah melakukan perlindungan sosial dengan cara memberikan BPNT. Program BPNT bertujuan untuk meringankan beban seseorang untuk memenuhi kebutuhan pangan dan memberikan gizi yang seimbang. Dengan adanya penyaluran BPNT yang diberikan kepada masyarakat. Kebutuhan sembako tersebut akan diarahkan ke daerah atau desa yang membutuhkan. Dalam BPNT, bantuan pangan diberikan dalam bentuk kartu elektronik atau kartu sembako yang diberikan kepada penerima manfaat. Kartu ini berisi saldo yang dapat digunakan untuk membeli bahan pangan di warung atau toko pangan yang telah bekerja sama dengan program BPNT. Penerima manfaat dapat memilih dan membeli bahan pangan yang mereka butuhkan sesuai dengan saldo yang tersedia di kartu. BPNT merupakan langkah awal pemerintah



dalam menyelesaikan masalah kemiskinan dan ketimpangan akses terhadap pangan di Indonesia. Dinas Sosial Kota Pekanbaru menentukan siapa yang berhak menerima BPNT. Namun data yang banyak menyebabkan Dinas Sosial Kota Pekanbaru kesulitan menjalankan program ini sehingga dikhawatirkan BPNT diberikan kepada masyarakat yang tidak berhak menerimanya [6].

Data mining adalah metode di mana data yang diperoleh diproses untuk mengetahui model data sehingga informasi yang tersembunyi dapat diekstraksi dari data tersebut [7]. Pada data mining terdapat salah satu teknik yang digunakan dalam pengolahannya yaitu KDD (Knowledge Discovery in Databases). KDD merupakan kegiatan multidisiplin yang berkaitan dengan suatu metodologi untuk memperoleh ilmu pengetahuan [8]. Tahapan awalnya dimulai dengan studi literatur yang dilanjutkan dengan pengumpulan data. Pengumpulan data mungkin memerlukan penggunaan perangkat keras khusus, tenaga kerja manual, atau alat perangkat lunak untuk mengumpulkan dokumen.

Selanjutnya dilakukan pemilihan atribut yaitu ukuran data dikurangi melalui pemilihan subset data dan pemilihan subset atribut. Kemudian, Pembersihan data, menghilangkan noise dan ketidakonsistenan atau ketidakrelevanan data [5] dan Transformasi data, menyesuaikan dengan format pemrosesan data mining, data diubah atau digabungkan menjadi format yang dibutuhkan [9].

K-medoids adalah metode pengelompokan data mining non-hierarkis berbasis partisi yang menggunakan medoid sebagai pusat cluster [10]. Indikator pengujian yang digunakan adalah Silhouette Coefficient (SC) yang menilai klaster berdasarkan ukuran jarak rata-rata antara satu titik data dengan titik data lainnya dalam klaster yang sama (kohesi) dan jarak rata-rata antar klaster yang berbeda (pemisahan) [11]. Metode ini ialah kombinasi dari proses kohesi dan pemisahan. Kohesi mengukur kedekatan objek dalam suatu cluster, sedangkan segregasi mengukur kedekatan antar cluster pembentuk [12].

Syed Ali Abbas dkk (2020) menyatakan dalam penelitiannya mengenai Analisis Cluster pada Data Kelahiran yang Dikumpulkan di Kota Muzaffarabad, Kashmir bahwa algoritma K-Medoids unggul dari K-Means dengan tingkat akurasi K-means, K-medoids dan Rank K-medoids rata-rata masing-masing adalah 67,58%, 69,58% dan 72,64% [13]. Mahdi Hashemzadeh dkk (2019) berhasil mengidentifikasi wilayah dalam kerangka input, yang kemungkinan besar merupakan nyala api dengan memanfaatkan fitur warna api menggunakan algoritma ICA K-Medoids [14].

Dalam penelitian pengelompokan kurva laktasi (LC) sapi Holstein di beberapa cluster berdasarkan karakteristik pemerahannya dan untuk mengetahui perbedaan fisiologis antar cluster, Mingyung Lee dkk (2020) berhasil mengelompokkan data laktasi individu dengan karakteristik pemerahan yang serupa [15]. Teknologi IoT, Big data, dan sensor, data multivariat sangat besar dan tidak dapat melakukan analisis dengan pendekatan tradisional sehingga Sivadi dkk (2019) mengembangkan algoritma K-Medoids untuk analisis data yang berukuran besar dan memperoleh hasil yang menggemblakan untuk waktu komputasi, NMI, kemurnian dan akurasi pengelompokan [16].

Selain itu, Soumita Modak dkk (2020) menggunakan pengelompokan k-medoids untuk mengungkap kelompok berbeda dari 1318 bintang variabel di Galaksi berdasarkan kurva cahayanya, di mana setiap kurva cahaya mewakili grafik kecerahan bintang terhadap waktu memperoleh dua kelompok biner gerhana optimal yang sesuai dengan sistem masif yang terang dan sistem yang lebih redup dan kurang masif [17]. Febiyaanti Alfiah dkk (2022) dalam kajian analisis k-medoids cluster tahun 2020 berdasarkan indikator kemiskinan di Jawa Timur, membagi daerah/kota dengan karakteristik yang sama menjadi dua kelompok sesuai dengan indikator kemiskinan di Jawa Timur, dan masing-masing kelompok memiliki karakteristik karakteristik sendiri [18].

Jefri Arianto (2019) dalam penelitiannya dengan menggunakan data mining, penduduk miskin di desa Sambirejo Timur diklasifikasikan melalui algoritma k-medoids, dan ditemukan bahwa pekerjaan, status keluarga, jumlah tanggungan keluarga, dan pendapatan bulanan merupakan faktor yang mempengaruhi yang menentukan siapa yang dapat menerima subsidi pemerintah desa. Keempat faktor tersebut digunakan untuk mengelompokkan masyarakat kurang beruntung di Desa Sambierejo Timur [19].

Perkembangan algoritma K-Medoids dalam beberapa penelitian sebelumnya memberikan hasil yang optimal sehingga membuat peneliti mengusulkan algoritma K-Medoids dalam pengelompokan penerima BPNT. Penelitian ini menggunakan data dari penelitian Firza Syahputra pada tahun 2020-2021 yang berjumlah 732 data terdiri dari 42 atribut [6] namun digunakan 25 atribut dalam menentukan pola karakteristik penerima BPNT sebagaimana yang terdapat pada Pedoman Umum Bantuan Pangan Non Tunai 2019 [20].

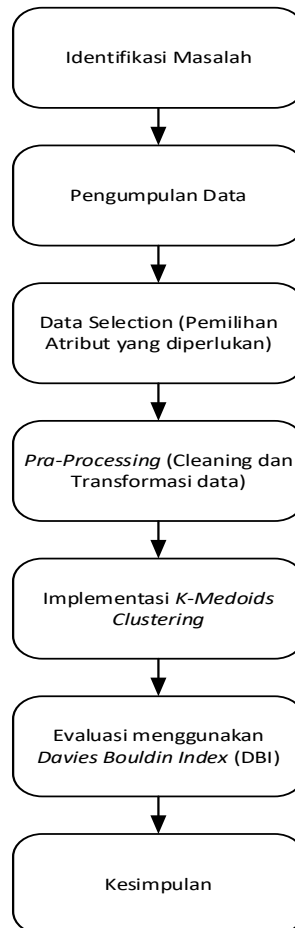
Analisis mendalam terhadap pola karakteristik penerima yang tidak tepat sasaran seperti identifikasi faktor-faktor yang memengaruhi ketidaksesuaian tersebut, kesalahan dalam pengumpulan data, perubahan status ekonomi, atau perubahan dalam komposisi rumah tangga. Dengan memahami pola ini, dapat dilakukan penyesuaian yang lebih baik dalam penentuan penerima. Banyaknya kriteria yang harus dipertimbangkan untuk mengetahui masyarakat yang berhak menerima BPNT membuat Dinas Sosial Kota Pekanbaru kesulitan dalam menjalankan program BPNT maka diperlukan metode data mining dalam menganalisis pola karakteristik penerima BPNT. Oleh karena itu, penulis akan menerapkan algoritma K-Medoids dalam clustering data BPNT sehingga pemerintah dapat mempertimbangkan faktor-faktor yang menjadi karakteristik penerima bantuan.

2. METODOLOGI PENELITIAN



2.1 Tahapan Penelitian

Teknik KDD (Knowledge Discovery in Databases) adalah proses yang melibatkan langkah-langkah sistematis untuk mengekstraksi pengetahuan yang berharga dan berarti dari data yang ada.. Adapun tahapan KDD yang diadopsi untuk penelitian dapat dilihat pada Gambar 1.



Gambar 1.Tahapan Penelitian

2.1.1 Pengumpulan Data

Metodologi pengumpulan data memainkan peran penting dalam memastikan validitas, reliabilitas, dan relevansi data yang dikumpulkan. Data penerima BPNT diperoleh dari penelitian Firza Syahputra pada tahun 2020-2021 dengan jumlah 732 data dan memiliki 42 atribut.

2.1.2 Data Selection

Pada tahapan ini data awal memiliki 42 atribut dengan jumlah data 732 data berubah menjadi 25 atribut dengan jumlah data 732 data. Atribut yang dihilangkan berjumlah 17 atribut dikarenakan Dinas Sosial hanya menggunakan 25 atribut dalam menentukan penerima bantuan pangan BPNT. Adapun atribut yang digunakan adalah status penguasaan tempat tinggal, luas lantai, jenis sumber air minum, jenis dinding terluas, bahan bakar masak, jenis lantai terluas, sumber penerangan utama, daya yang terpasang, , fasilitas BAB, pemanas air, lemari es/ kulkas, televisi, AC, telepon gengam, emas atau perhiasan, komputer/ laptop, mobil, sepeda, sepeda motor, lahan, rumah di tempat lain, perahu motor, perahu, motor tempel.

2.1.3 Pre-Processing

Tahapan preprocessing dibagi menjadi dua bagian yaitu:

- Data Cleaning, merupakan proses mempersiapkan dan membersihkan data mentah (raw data) sebelum dianalisis atau dimodelkan. Tujuan dari data cleaning adalah untuk memastikan data yang digunakan berkualitas tinggi, bebas dari kesalahan, inkonsistensi, dan missing values. Pada tahap ini dilakukan pembersihan data yang bernilai null dan kosong sehingga jumlah data berubah menjadi 683 data.
- Data Transformation, merupakan proses mengubah atau mengolah data dalam bentuk tertentu agar lebih sesuai untuk analisis atau pemodelan. Tahap ini dilakukan dengan mengubah data yang berformat huruf menjadi



angka. Terdapat 24 parameter yang berformat nominal, pada tahap ini diubah semuanya dalam format numerik. Proses transformasi dijabarkan pada Tabel 1.

Tabel 1. Transformasi Data

| Atribut | Sub Atribut | Nilai perubahan |
|----------------------------------|----------------------------|-----------------|
| Status penguasaan tempat tinggal | dinas | 5 |
| | milik sendiri | 4 |
| | bebas sewa | 3 |
| | kontrak/sewa | 2 |
| | lainnya | 1 |
| Jenis sumber air minum | kemasan bermerk | 10 |
| | kemasan bermerk isi ulang | 9 |
| | isi ulang | 8 |
| | leding enceran | 7 |
| | leding meteran | 6 |
| | sumur bor pompa | 5 |
| | sumur terlindungi | 4 |
| | sumur tidak terlindungi | 3 |
| | mata air terlindungi | 2 |
| | mata air tak terlindungi | 1 |
| Jenis dinding terluas | tembok | 5 |
| | plasteran | 4 |
| | kayu | 3 |
| | batang kayu | 2 |
| Bahan bakar masak | lainnya | 1 |
| | listrik | 5 |
| | gas 3 kg | 4 |
| | minyak tanah | 3 |
| Jenis lantai terluas | gas kota / bio gas | 2 |
| | briket | 1 |
| | marmar/granit | 9 |
| | marmar/granit lainnya | 8 |
| | keramik | 7 |
| | sementara/batu merah | 6 |
| | kayu/papan kualitas tinggi | 5 |
| | kayu/papan kualitas rendah | 4 |
| | parket/vinil/permadani | 3 |
| | lainnya | 2 |
| Sumber penerangan utama | tanah | 1 |
| | listrik PLN | 3 |
| | listrik non PLN | 2 |
| | bukan listrik | 1 |
| Daya terpasang | 2200 watt | 4 |
| | 1300 watt | 3 |
| | 900 watt | 2 |
| Fasilitas BAB | tanpa meteran | 1 |
| | sendiri | 4 |
| | bersama | 3 |
| Pemanas air | umum | 2 |
| | tidak ada | 1 |
| | ya | 2 |
| Lemari es/kulkas | tidak | 1 |
| | ya | 2 |
| Televisi | tidak | 1 |
| | ya | 2 |
| AC | tidak | 1 |
| | ya | 2 |
| Telepon rumah/telepon genggam | tidak | 1 |
| | ya | 2 |
| Emas/perhiasan & tabungan | tidak | 1 |
| | ya | 2 |



| Atribut | Sub Atribut | Nilai perubahan |
|----------------------|-------------|-----------------|
| Komputer/laptop | tidak | 1 |
| | ya | 2 |
| Mobil | tidak | 1 |
| | ya | 2 |
| Sepeda | tidak | 1 |
| | ya | 2 |
| Sepeda motor | tidak | 1 |
| | ya | 2 |
| Lahan | tidak | 1 |
| | ya | 2 |
| Rumah di tempat lain | tidak | 1 |
| | ya | 2 |
| Perahu motor | tidak | 1 |
| | ya | 2 |
| Perahu | tidak | 1 |
| | ya | 2 |
| Motor tempel | tidak | 1 |
| | ya | 2 |

Penjelasan mengenai Tabel 1 adalah sebagai berikut.

- Status penguasaan tempat tinggal, setiap sub atribut diubah dari nilai 5 sampai 1 yang dimulai dari dinas berturut-turut sampai lainnya. Sebagai contoh Dinas mewakili nilai 5, Milik sendiri mewakili nilai 4, bebas sewa mewakili nilai 3, kontrak/sewa mewakili nilai 2, dan lainnya mewakili nilai 1.
- Jenis sumber air minum, setiap sub atribut diubah dari nilai 10 sampai 1 yang dimulai dari bermerk berturut-turut sampai air kemasan mata air tak terlindungi.
- Jenis dinding terluas, setiap sub atribut diubah dari nilai 5 sampai 1 yang dimulai dari tembok berturut-turut sampai lainnya.
- Bahan bakar masak, setiap sub atribut diubah dari nilai 5 sampai 1 yang dimulai dari listrik berturut-turut sampai air kemasan briket.
- Jenis lantai terluas, setiap sub atribut diubah dari nilai 9 sampai 1 yang dimulai marmer/granit dari berturut turut sampai lainnya.
- Sumber penerangan utama, setiap sub atribut diubah dari nilai 3 sampai 1 yang dimulai dari listrik PLN berturut-turut sampai bukan listrik.
- Daya terpasang, setiap sub atribut diubah dari nilai 4 sampai 1 yang dimulai dari 2200 Watt berturut-turut sampai tanpa meteran.
- Fasilitas BAB, setiap sub atribut diubah dari nilai 4 sampai 1 yang dimulai dari sendiri berturut-turut sampai tidak ada.
- Pemanas air, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Lemari es/kulkas, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Televisi, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- AC, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Telepon rumah/telepon genggam, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Emas/perhiasan & tabungan, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Komputer/leptop, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Mobil, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Sepeda, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Sepeda motor, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Lahan, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Rumah di tempat lain, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Perahu motor, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Perahu, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.
- Motor tempel, setiap sub atribut diubah menjadi nilai 2 dan 1 yang mana 2 mewakili ya dan 1 mewakili tidak.

2.4 K-Medoids Clustering



Algoritma K-medoids bekerja dengan menemukan k sebagai objek yang representif untuk meminimalkan jumlah ketidaksamaan objek data. Adapun tahapan Algoritma K-medoids adalah sebagai berikut.

- Inisialisasi: memilih titik k secara acak dari dataset sebagai medoids awal. Medoids adalah titik-titik data aktual yang mewakili setiap kluster.
- Assign dan Update: Untuk setiap titik data, hitung jaraknya terhadap setiap medoid. Assign titik tersebut ke kluster yang memiliki medoid terdekat. Perhitungan jarak dihitung dengan fungsi jarak salah satunya adalah fungsi Euclidean Distance dengan persamaan berikut t [13]:

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (1)$$

dimana:

x_i = variabel ke-i dari objek i

d = Euclidean Distance.

- Evaluasi: Hitung nilai total jarak dari setiap titik data ke medoid kluster yang sesuai. Ini mencerminkan kualitas klustering.
- Iterasi: Ulangi langkah 2 dan 3 hingga konvergensi atau hingga mencapai jumlah iterasi maksimum yang ditentukan.
- Hasil akhir: Klustering akan berhenti ketika konvergensi tercapai atau ketika mencapai jumlah iterasi maksimum. Hasil akhir adalah kluster yang terbentuk dan medoid untuk setiap kluster.

2.5 Silhouette Coefficient

Silhouette Coefficient (Koefisien Silhouette) adalah metrik evaluasi yang digunakan untuk mengukur sejauh mana setiap sampel dalam kluster memiliki kemiripan yang lebih besar dengan anggota kluster sendiri daripada dengan anggota kluster lainnya.. Di bawah ini adalah persamaan untuk menghitung Silhouette Coefficient [21].

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \quad (2)$$

Dengan

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, b(i) = \min d(i, P), \text{ dan } a(i) = \frac{1}{|Q|-1} \sum_{j \in Q, j \neq i} d(i, j) \quad (3)$$

dimana:

SC = Silhouette Coefficient

$b(i)$ = nilai minimal dari jarak rata-rata objek i ke semua objek pada kluster lain P

$a(i)$ = rata-rata jarak objek ke-i ke semua objek yang berada di dalam satu kluster Q

Nilai Silhouette Coefficient berkisar antara -1 hingga 1. Nilai yang lebih dekat ke 1 menunjukkan bahwa sampel tersebut secara keseluruhan lebih cocok dengan klusternya, sedangkan nilai yang lebih dekat ke -1 menunjukkan bahwa sampel tersebut mungkin ditempatkan di kluster yang salah. Nilai mendekati 0 menunjukkan bahwa sampel berada di dekat batas antara dua kluster yang berdekatan. Interpretasi nilai Silhouette Coefficient adalah sebagai berikut:

- Nilai mendekati 1: Indikasi baik bahwa sampel cocok dengan klusternya.
- Nilai mendekati 0: Indikasi bahwa sampel berada di dekat batas antara dua kluster yang berdekatan.
- Nilai mendekati -1: Indikasi bahwa sampel mungkin ditempatkan di kluster yang salah.

Untuk mengevaluasi performa keseluruhan klustering, Silhouette Coefficient dapat dihitung dengan mengambil rata-rata Silhouette Coefficient dari semua sampel dalam dataset. Nilai yang lebih tinggi menunjukkan hasil klustering yang lebih baik, dengan kluster yang lebih terpisah dan lebih konsisten.

3. HASIL DAN PEMBAHASAN

3.1 K-Medoids Clustering

Penerapan algoritma k-medoids clustering menggunakan python 3.9 dengan memanggil data yang sudah melewati tahap pre-processing. Diawali dengan mengimpor library yang diperlukan, seperti NumPy dan scikit-learn. NumPy akan digunakan untuk manipulasi data dalam bentuk array, sementara scikit-learn akan digunakan untuk menghitung jarak antara dua titik. Kemudian menentukan jumlah kluster yang diinginkan (k) dan inisialisasikan medoids awal secara acak dari data yang tersedia. Selama proses clustering, hasilnya dapat berbeda-beda tergantung pada inisialisasi medoids awal, jumlah kluster yang dipilih, dan metrik jarak yang digunakan. Oleh karena itu, penting untuk melakukan beberapa iterasi atau pengulangan algoritma K-Medoids dengan inisialisasi yang berbeda untuk mendapatkan hasil clustering yang lebih stabil. Digunakan metric euclidean untuk menentukan fungsi jarak yang akan digunakan. Pada proses ini diinisialisasikan k adalah sama dengan 3 sebagaimana ditampilkan pseudocode pada Gambar 2.



```
import numpy as np
from sklearn_extra.cluster import KMedoids

# Create a KMedoids object with k=3 clusters
kmedoids = KMedoids(n_clusters=3, metric='euclidean', random_state=0)

# Run the k-medoids algorithm
kmedoids.fit(x_pred)

# Get the cluster labels and medoids
labels = kmedoids.labels_
medoids = kmedoids.cluster_centers_

# Print the results
print("Cluster labels: ", labels)
#print("Medoids: ", medoids)
sampel1 ['CLUSTER'] = kmedoids.labels_
print(sampel1 ['CLUSTER'].value_counts())
#sampel1
medoids
```

Gambar 2. K-Medoids Clustering

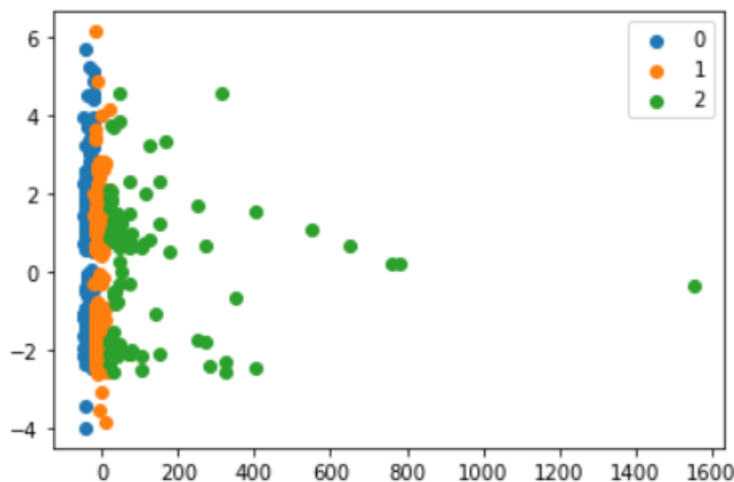
Hasil clustering K-Medoids menggunakan Python adalah pembagian data ke dalam kelompok-kelompok atau kluster berdasarkan medoids yang dipilih. Setiap sampel akan diatribusikan ke kluster yang memiliki medoid terdekat dengan sampel tersebut. Proses Clustering menggunakan K-Medoids menghasilkan 3 cluster sesuai dengan jumlah inisialisasi nilai k. Cluster 0 berjumlah 309, Cluster 1 berjumlah 275 dan Cluster 2 berjumlah 98 item. Hasil clustering ditampilkan pada Gambar 3 berikut ini.

```
Cluster labels: [0 0 0 0 0 1 0 0 0 0 0 0 1 2 1 1 1 1 1 0 0 0 0 1 2 1 1 1 0 0 1 2 1 1 1 0 0 1 2 1 0 0 1
1 0 0 0 0 0 0 1 2 0 0 0 0 1 2 2 1 0 0 1 1 1 1 0 0 0 0 0 1 1 1 0 0 0 0 1 1 1 0 1 2 1
0 0 0 1 0 0 0 0 0 0 1 1 0 0 1 1 1 2 2 1 2 1 0 2 2 1 2 1 1 1 1 1 2 2 0 0 0 1
1 1 1 0 0 0 0 0 0 1 2 1 1 1 2 0 0 1 2 2 2 1 1 0 1 1 1 1 1 2 1 2 2 0 0 1
1 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0 2 1 1 1 0 1 1 1 2 2 2 2 1
0 0 1 2 0 1 0 0 1 2 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 1 1 0 1 1 2 1 1 0 0
0 0 0 0 0 1 0 1 1 1 2 1 1 1 2 1 0 2 1 1 0 2 1 0 0 0 0 1 1 1 1 0 1 1 1
0 0 2 2 2 1 2 2 0 1 0 2 0 0 0 0 0 2 1 0 2 2 2 0 1 0 0 0 0 0 0 2 1 1
0 1 0 0 1 0 1 1 2 2 1 1 0 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 1 1 2 2 1
0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 1 0 1 2 2 1 2 2 1 1 0 0 1 2 1 0 1 1
0 0 0 0 0 0 1 1 1 0 1 1 0 0 1 1 1 1 0 0 0 0 1 1 1 1 2 1 1 1 0 1 1 1 0 0
0 1 2 1 0 0 0 1 2 2 1 0 0 2 1 1 2 2 1 1 1 2 1 1 0 1 1 1 2 2 1 1 1 2 2 1
1 1 1 2 2 2 1 1 1 1 0 0 0 0 2 1 0 2 1 1 0 1 2 2 1 1 1 1 1 1 0 1 0 1 1 0 1
1 1 0 1 1 1 1 1 0 1 2 1 0 0 0 1 0 1 0 1 1 1 0 0 0 1 1 0 0 0 2 0 0 0
0 0 2 2 1 1 1 1 0 0 0 1 1 1 0 1 1 1 0 2 1 2 2 1 1 1 1 0 1 1 1 0 0 1 1
1 2 0 0 0 1 1 1 2 0 0 1 2 2 1 0 0 1 0 0 0 0 0 0 1 2 1 0 0 1 1 1 0 0 0
0 2 0 0 0 1 2 1 0 0 1 2 2 1 0 0 0 1 2 1 0 0 0 1 1 1 0 0 0 1 1 1 2 2
1 1 1 2 1 1 0 1 0 0 0 1 1 0 1]
```

```
0    309
1    275
2     98
Name: CLUSTER, dtype: int64
```

Gambar 3. Hasil Clustering K-Medoids

Dari hasil pengelompokan ditampilkan sebaran data K-medoids clustering Gambar 6.



Gambar 4. Sebaran Cluster

Dari gambar dapat diketahui bahwa cluster 0 yang diwakili dengan warna biru mengelompok lurus vertikal di sebelah kanan warna oranye yang mewakili cluster 1, dan cluster 2 mengelompok disebelah kanan cluster 1



yang sebaran datanya cukup menyebar dibandingkan cluster lainnya. Untuk mengetahui karakteristik setiap cluster dilakukan dengan menganalisis titik medoids yang terbentuk pada setiap cluster yang akan dibahas dibagian analisis hasil clustering.

3.2 Silhouette Coefficient

Proses evaluasi dilakukan dengan menggunakan Silhouette Coefficient pada k-medoids clustering diperoleh cluster dengan nilai k=3 adalah cluster terbaik karena memiliki nilai Silhouette Coefficient paling tinggi. Tahapan evaluasi dilakukan dengan menghitung nilai koefisien Silhouette yang dicobakan pada algoritma K-Medoids dalam 10 pengujian yaitu nilai k 2 sampai 10 untuk melihat cluster terbaik.

```
from sklearn.cluster import KMeans
from sklearn.datasets import load_iris
from sklearn.metrics import silhouette_score, calinski_harabasz_score
import matplotlib.pyplot as plt

# Create list of number of clusters
clusters = range(2, 11)

# Initialize empty lists for silhouette scores
silhouette_scores = []

# Loop through different number of clusters and calculate silhouette and calinski-harabasz scores
for k in clusters:
    kmedoids = KMedoids(n_clusters=k, metric='euclidean', random_state=0)
    labels = kmedoids.fit_predict(x_pred)
    silhouette_scores.append(silhouette_score(x_pred, labels))

# Plot silhouette scores
plt.plot(clusters, silhouette_scores, marker='o')
plt.xlabel('Number of clusters')
plt.ylabel('Silhouette Score')

# Show plot
plt.show()
```

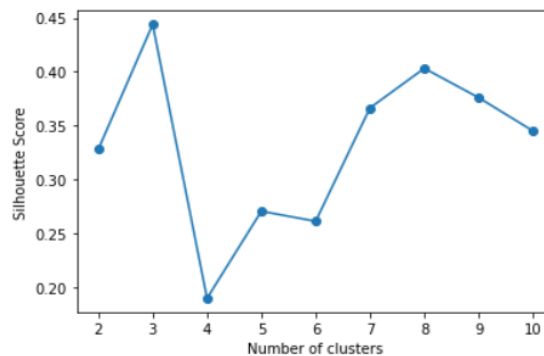
Gambar 5. Silhouette Coefficient

Evaluasi hasil klastering menggunakan Silhouette Coefficient memberikan gambaran tentang sejauh mana klastering yang dihasilkan oleh algoritma K-Medoids berhasil memisahkan data menjadi kelompok-kelompok yang serupa. Hasil evaluasi menggunakan Silhouette Coefficient dapat ditampilkan pada Tabel 2.

Tabel 2. Silhouette Coefficient

| Nilai k | Silhouette Coefficient |
|---------|---------------------------|
| 2 | 0.32822482887377086 |
| 3 | 0.4439221599010089 |
| 4 | 0.19002952742220164 |
| 5 | 0.2707216543729025 |
| 6 | 0.26142579235470664 |
| 7 | 0.3665014508775353 |
| 8 | 0.4031910442833763 |
| 9 | 0.37602441332237435 |
| 10 | 0.34514871606719394 |

Jika ditampilkan dalam bentuk grafik nilai Silhouette Coefficient adalah seperti Gambar 6.



Gambar 6. Grafik Silhouette Coefficient

Grafik puncak Silhouette dapat digunakan untuk memvisualisasikan nilai Silhouette Coefficient pada berbagai jumlah cluster yang berbeda. Grafik ini membantu kita memilih jumlah cluster yang optimal berdasarkan



nilai tertinggi dalam grafik puncak. Berdasarkan grafik, nilai $k=3$ menempati puncak grafik yang menggambarkan bahwa jumlah cluster terbaik untuk pengelompokan penerima BPNT adalah 3 cluster.

3.3 Analisis Hasil Clustering

Analisis nilai medoids dalam konteks klastering sangat penting untuk memahami karakteristik klaster yang dihasilkan oleh algoritma k-medoids. Berikut pada gambar 7 adalah nilai medoid yang terbentuk.

```
pd.DataFrame(medoids)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 2.0 | 15.0 | 6.0 | 3.0 | 5.0 | 3.0 | 2.0 | 4.0 | 4.0 | 2.0 | ... | 1.0 | 1.0 | 1.0 | 2.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| 1 | 2.0 | 42.0 | 7.0 | 4.0 | 6.0 | 3.0 | 2.0 | 4.0 | 4.0 | 2.0 | ... | 1.0 | 1.0 | 1.0 | 2.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| 2 | 4.0 | 96.0 | 6.0 | 4.0 | 6.0 | 3.0 | 3.0 | 4.0 | 4.0 | 2.0 | ... | 1.0 | 2.0 | 2.0 | 2.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |

3 rows x 24 columns

Gambar 7. Nilai Medoids

Dari tabel medoids diatas ditransformasikan nilai kembali ke format awal sehingga diperoleh analisis nilai medoids. Adapun analisis yang diperoleh pada tabel 2 yaitu:

1. Cluster 0 memiliki ciri karakteristik status penguasaan tempat tinggal adalah kontrak/sewa, luas lantai 15, jenis sumber air minum adalah sumur bor pompa, jenis dinding terluas adalah kayu, bahan bakar masak adalah gas 3 kg, jenis lantai terluas adalah sementara/batu merah, sumber penerangan utama adalah listrik PLN, daya yang terpasang adalah 900 Watt, fasilitas BAB sendiri, tidak ada pemanas air, ada lemari es/kulkas, ada televisi, tidak ada AC, ada telepon rumah/telepon genggam, tidak ada emas/perhiasan & tabungan, tidak ada komputer/laptop, tidak ada mobil, tidak ada sepeda, ada sepeda motor, tidak ada lahan, tidak ada rumah di tempat lain, tidak ada perahu motor, tidak ada perahu, tidak ada motor tempel sehingga diasumsikan termasuk kategori berhak menerima BPNT.
2. Cluster 1 memiliki ciri karakteristik status penguasaan tempat tinggal adalah kontrak/sewa, luas lantai 42, jenis sumber air minum adalah sumur leding meteran, jenis dinding terluas adalah plasteran, bahan bakar masak adalah gas 3 kg, jenis lantai terluas adalah keramik, sumber penerangan utama adalah listrik PLN, daya yang terpasang adalah 900 Watt, fasilitas BAB sendiri, tidak ada pemanas air, ada lemari es/kulkas, ada televisi, tidak ada AC, ada telepon rumah/telepon genggam, tidak ada emas/perhiasan & tabungan, tidak ada komputer/laptop, tidak ada mobil, tidak ada sepeda, ada sepeda motor, tidak ada lahan, tidak ada rumah di tempat lain, tidak ada perahu motor, tidak ada perahu, tidak ada motor tempel sehingga diasumsikan termasuk kategori dipertimbangkan menerima BPNT.
3. Cluster 2 memiliki ciri karakteristik status penguasaan tempat tinggal adalah milik sendiri, luas lantai 96, jenis sumber air minum adalah leding meteran, jenis dinding terluas adalah plasteran, bahan bakar masak adalah gas 3 kg, jenis lantai terluas adalah sementara/batu merah, sumber penerangan utama adalah listrik PLN, daya yang terpasang adalah 1300 Watt, fasilitas BAB sendiri, tidak ada pemanas air, ada lemari es/kulkas, ada televisi, tidak ada AC, ada telepon rumah/telepon genggam, tidak ada emas/perhiasan & tabungan, ada komputer/laptop, tidak ada mobil, ada sepeda, ada sepeda motor, tidak ada lahan, tidak ada rumah di tempat lain, tidak ada perahu motor, tidak ada perahu, tidak ada motor tempel sehingga diasumsikan termasuk kategori tidak berhak menerima BPNT.

4. KESIMPULAN

Berdasarkan analisa dan pembahasan yang telah penulis lakukan dapat disimpulkan dengan menerapkan algoritma K-Medoids Clustering bahwa cluster terbaik yang terbentuk adalah 3 cluster beranggotakan cluster 0 berjumlah 309 items, cluster 1 berjumlah 275 items dan cluster 2 berjumlah 98 dengan koefisien Silhouette **0.4439221599010089** yang merupakan nilai paling tinggi dari nilai pengujian lainnya. Hasil analisis clustering menunjukkan bahwa clustering membagi kelompok berdasarkan pola karakteristik setiap cluster sehingga diamsusikan bahwa Cluster 0 Berhak Menerima BPNT, Cluster 1 Dipertimbangkan dan Cluster 2 Tidak Berhak Menerima BPNT

REFERENCES

- [1] F. Alfiah, A. Almadayani, D. Al Farizi, and E. Widodo, "Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020," JURNAL ILMIAH SAINS, vol. 22, no. 1, p. 1, Dec. 2021, doi: 10.35799/jis.v22i1.35911.



- [2] Badan Pusat Statistika, “Profil Kemiskinan di Indonesia September 2022,” 2022. <https://www.bps.go.id/pressrelease/2023/01/16/2015/persentase-penduduk-miskin-september-2022-naik-menjadi-9-57-persen.html#:~:text=Jumlah%20penduduk%20miskin%20pada%20September,53%20persen%20pada%20September%202022.> (accessed Jun. 26, 2023).
- [3] V. G. Sianturi, M. Syafii, and A. A. Tanjung, “ANALISIS DETERMINASI KEMISKINAN DI INDONESIA STUDI KASUS (2016-2019),” *Samuka*, vol. 5, no. 2, 2021, doi: doi.org/10.33059/jse.v5i2.4270.
- [4] E. D. Ristika, W. P. Primandhana, and M. Wahed, “Analisis Pengaruh Jumlah Penduduk, Tingkat Pengangguran Terbuka Dan Indeks Pembangunan Manusia Terhadap Tingkat Kemiskinan Di Provinsi Jawa Timur,” *Eksis: Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 12, no. 2, p. 129, Nov. 2021, doi: [10.33087/eksis.v12i2.254](https://doi.org/10.33087/eksis.v12i2.254).
- [5] E. Widodo and A. Jaelani, “PENERAPAN DATA MINING UNTUK PREDIKSI PENERIMA BANTUAN PANGAN NON TUNAI (BPNT) DI DESA WANACALA MENGGUNAKAN METODE NAÏVE BAYES,” vol. 13, 2022, doi: doi.org/10.37366/sigma.v13i3.
- [6] F. SYAHPUTRA, “SISTEM PENDUKUNG KEPUTUSAN BANTUAN PANGAN NON TUNAI MENGGUNAKAN METODE FUZZY ANALYTICAL HIERARCHY PROCESS DAN TOPSIS,” 2022.
- [7] A. D. Andini and T. Arifin, “IMPLEMENTASI ALGORITMA K-MEDOIDS UNTUK KLASIFIKASI DATA PENYAKIT PASIEN DI RSUD KOTA BANDUNG,” *JURNAL RESPONSIF*, vol. 2, no. 2, pp. 128–138, 2020, doi: doi.org/10.51977/jti.v2i2.247.
- [8] L. Ardiantoro, S. Zahara, and N. Sunarmi, “PEMANFAATAN KNOWLEDGE DATA DISCOVERY(KDD) PADA POLA PERMAINAN ATLET BULUTANGKIS,” *JURNAL EXPLORE IT!*, vol. 11, no. 1, pp. 1–6, 2019, doi: doi.org/10.35891/explorit.v11i1.1467.
- [9] Pratama Zulfiansyah Muhammad, “PENENTUAN KRITERIA DAN PENERIMA BANTUAN PANGAN NON TUNAI DI DESA PARUNGKUDA DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES,” Universitas Bina Sarana Informatika, Sukabumi, 2020. [Online]. Available: <https://elibrary.bsi.ac.id/skripsi/o21920200030i06/penentuan-kriteria-dan-penerima-bantuan-pangan-non-tunai-di-desa-parungkuda-dengan-menggunakan-algoritma-na-ve-bayes>
- [10] C. Oktarina, K. Anwar Notodiputro, and Indahwati, “COMPARISON OF K-MEANS CLUSTERING METHOD AND K-MEDOIDS ON TWITTER DATA,” *Indonesian Journal of Statistics and Its Applications*, vol. 4, no. 1, pp. 189–202, 2020, doi: [10.29244/ijsa.v4i1.599](https://doi.org/10.29244/ijsa.v4i1.599).
- [11] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, “Electrical peak load clustering analysis using K-means algorithm and silhouette coefficient,” in *Proceeding - 2nd International Conference on Technology and Policy in Electric Power and Energy, ICT-PEP 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 258–262. doi: [10.1109/ICT-PEP50916.2020.9249773](https://doi.org/10.1109/ICT-PEP50916.2020.9249773).
- [12] U. IZZATY, I. R. HG, and D. DEVIANTO, “PENGKLASIFIKASIAN KABUPATEN/KOTA DI PROVINSI SUMATERA BARAT BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT DENGAN VALIDITAS KOEFISIEN SILHOUETTE,” *Jurnal Matematika UNAND*, vol. 9, no. 2, pp. 192–199, 2020, doi: doi.org/10.25077/jmu.9.2.192-198.2020.
- [13] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, “K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir,” *IEEE Access*, vol. 8, pp. 151847–151855, 2020, doi: [10.1109/ACCESS.2020.3014021](https://doi.org/10.1109/ACCESS.2020.3014021).
- [14] M. Hashemzadeh and A. Zademehti, “Fire detection for video surveillance applications using ICA K-medoids-based color model and efficient spatio-temporal visual features,” *Expert Syst Appl*, vol. 130, pp. 60–78, Sep. 2019, doi: [10.1016/j.eswa.2019.04.019](https://doi.org/10.1016/j.eswa.2019.04.019).
- [15] M. Lee, S. Lee, J. Park, and S. Seo, “Clustering and characterization of the lactation curves of dairy cows using K-medoids clustering algorithm,” *Animals*, vol. 10, no. 8, pp. 1–14, Aug. 2020, doi: [10.3390/ani10081348](https://doi.org/10.3390/ani10081348).
- [16] S. Balakrishna, M. Thirumaran, R. Padmanaban, and V. K. Solanki, “An efficient incremental clustering based improved K-Medoids for IoT multivariate data cluster analysis,” *Peer Peer Netw Appl*, vol. 13, no. 4, pp. 1152–1175, Jul. 2020, doi: [10.1007/s12083-019-00852-x](https://doi.org/10.1007/s12083-019-00852-x).
- [17] S. Modak, T. Chattopadhyay, and A. K. Chattopadhyay, “Unsupervised classification of eclipsing binary light curves through k-medoids clustering,” *J Appl Stat*, vol. 47, no. 2, pp. 376–392, Jan. 2020, doi: [10.1080/02664763.2019.1635574](https://doi.org/10.1080/02664763.2019.1635574).
- [18] F. Alfiah, A. Almadayani, D. Al Farizi, and E. Widodo, “Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020,” *JURNAL ILMIAH SAINS*, vol. 22, no. 1, p. 1, Dec. 2021, doi: [10.35799/jis.v22i1.35911](https://doi.org/10.35799/jis.v22i1.35911).
- [19] J. Arianto, “PENERAPAN DATA MINING UNTUK PENGELOMPOKAN PENDUDUK KURANG MAMPU DESA SAMBIREJO TIMUR DENGAN ALGORITMA K-MEDOIDS (STUDI KASUS KANTOR KEPALA DESA SAMBIREJO TIMUR),” *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: [10.30865/komik.v3i1.1660](https://doi.org/10.30865/komik.v3i1.1660).
- [20] Tim Pengendali Pelaksanaan Penyaluran Bantuan Sosial Secara Nontunai, *PEDOMAN UMUM BANTUAN PANGAN NONTUNAI 2019*. 2019.
- [21] M. A. Nahdliyah, T. Widiariyah, and A. Prahutama, “METODE k-MEDOIDS CLUSTERING DENGAN VALIDASI SILHOUETTE INDEX DAN C-INDEX (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018),” *JURNAL GAUSSIAN*, vol. 8, no. 2, pp. 161–170, 2019, doi: doi.org/10.14710/j.gauss.8.2.161-170.