

Penelitian Awal Pengelompokkan Data Penduduk Mempergunakan Algoritma K-means Clustering untuk Penentuan Penerimaan Bantuan Dana Desa

Yolendes Hermadi Dandi¹, Anastasia Rita Widiarti²
Fakultas Sains dan Teknologi / Program Studi Informatika
Universitas Sanata Dharma
Yogyakarta, Indonesia
e-mail: ¹marselius306@gmail.com, ²rita_widiarti@usd.ac.id

Abstrak

Persoalan penentuan yang berhak menerima bantuan dana desa oleh suatu pemerintah desa seringkali berimbas pada tidak berhasilnya tujuan pemberian dana, karena bisa berakibat keluarga yang seharusnya mendapatkan bantuan dana desa menjadi tidak memperoleh haknya. Salah satu yang dapat ditempuh adalah dengan memanfaatkan algoritma k-means clustering untuk membantu pengambil keputusan dalam memilih penduduk yang berhak menerima bantuan dana tersebut, dengan harapan hasilnya bisa lebih akurat. Dari hasil percobaan dengan menggunakan 188 data penduduk di desa X yang telah melewati pengolahan data awal termasuk membuang atribut yang tidak relevan, nilai silhouette terbesar pada hasil pengelompokkan data menjadi 2 bagian adalah 0,728278964. Nilai tersebut menunjukkan bahwa hasil pengelompokkan data yang terjadi sudah kuat dan setiap data sudah ditempatkan pada klaster yang sangat sesuai. Atribut-atribut yang diduga dapat digunakan sebagai basis penentuan pengambilan keputusan adalah jumlah keluarga, luas lahan, luas lantai, jenis lantai, bahan dinding, keberadaan jendela, bahan atap, jenis penerangan, tempat pembuangan sampah, sarana MCK, sumber air, dan sarana buang air besar. Dari nilai ground truth yang ditemukan sebesar 70,21%, dapat diketahui pula bahwa hasil pengelompokkan relevan dengan kenyataan di lapangan.

Kata kunci: clustering, k-means, silhouette index, bantuan dana.

Abstract

The problem of determining who is entitled to receive village fund assistance by a village government often has an impact on the unsuccessful purpose of providing funds, because it can result in families who should have received village fund assistance becoming unable to obtain their rights. One of the things that can be taken is to utilize the k-means clustering algorithm to assist decision makers in choosing residents who are eligible to receive the funding assistance, in the hope that the results can be more accurate. From the results of the experiment using 188 population data in village X who had passed the initial data processing including removing irrelevant attributes, the largest silhouette value in the results of grouping data into 2 parts was 0.728278964. The value indicates that the results of the data grouping that occurred were already strong and each data was already placed in a very suitable cluster. The attributes that are thought to be used as a basis for determining decision making are the number of families, land area, floor area, type of floor, wall material, presence of windows, roofing material, type of lighting, garbage disposal, toilet facilities, water sources, and disposal facilities. From the ground truth value found at 70.21%, it can also be seen that the results of the grouping are relevant to reality in the field.

Keywords: clustering, k-means, silhouette index, aid funds.

1. Pendahuluan

Wakil Menteri Desa, Pembangunan Tertinggal dan Transmigrasi Budi Arie Setiadi mengungkapkan adanya 3 persoalan utama yang menjadi penyebab penyaluran bantuan langsung tunai atau disingkat BLT dana desa tidak tepat sasaran. Dari kesimpulan yang ditarik berdasarkan catatan lebih dari 2000-an aduan terkait kesalahan dalam penyaluran BLT Dana Desa, hal tersebut disebabkan karena kesalahan administratif, adanya niat yang salah dari kepala desa, atau kepala daerah, dan adanya pembegalan dari alokasi bantuan sosial dan bantuan langsung [1].

Dari hasil observasi melalui wawancara dengan salah satu kepala Dusun di desa X yang berada di suatu daerah pedalaman, terkait dengan penyaluran BLT terungkap bahwa pemerintah desa merasa kesulitan dalam mengelompokkan dan memilih keluarga yang menerima bantuan. Akibatnya, dirasakan bahwa masih terdapat keluarga yang membutuhkan bantuan namun belum mendapat perhatian. Sementara secara demografis, desa dengan total penduduk berjumlah 188 keluarga pada semester 1 Tahun 2020, dan pekerjaan utama sebagian besar penduduknya petani yang masih bergantung kepada alam dalam memenuhi kebutuhan sehari-hari, tentu membutuhkan sekali perhatian pemerintah dalam hal kesejahteraan. Maka dari itu, untuk membantu dan mengoptimalkan keputusan pemerintah desa dalam memilih keluarga yang akan menerima bantuan dana desa, perlu dikembangkan suatu sistem pengelompokkan tingkat kesejahteraan penduduk desa tersebut.

Terdapat beberapa penelitian yang berkaitan dengan kasus-kasus serupa terkait implementasi pengelompokkan data untuk mengambil keputusan pemberian bantuan. Kusnadi dan Putri yang melakukan penelitian mengenai prioritas penerima bantuan bedah rumah, menyimpulkan bahwa metode k-means dapat dipergunakan sebagai salah satu metode pengklasteran yang membantu keputusan dalam menentukan kelompok penduduk prioritas yang mendapatkan beda rumah [2]. Peneliti lain Yurnas, Sarwinda, dan Muttakin yang meneliti tentang pengelompokan penerima bantuan kesejahteraan masyarakat, menyatakan bahwa penerapan dan pendekatan data mining dengan algoritma k-means *clustering* mampu untuk mengelompokkan data penduduk dan memiliki tingkat akurasi yang cukup tinggi [3]. Pada penelitian yang dilakukan oleh Muhidin dan Baragigiratri tentang pemetaan penduduk calon penerima bantuan renovasi rumah, mengatakan bahwa penelitian tersebut menghasilkan 3 kluster dengan akurasi yang didapatkan menggunakan Daves Bouldin Index (DBI) sebesar 0.214 [4]. Penelitian yang dilakukan oleh Noviana, Jasmir, dan Novianto yaitu tentang penerapan menentukan kelompok prioritas penerima bantuan beras rastra mendapatkan hasil akurasi penentuan kelompok prioritas penerima bantuan beras sejahtera dari 3 kluster, yaitu kluster 1 terdapat 46 data penduduk dengan persentase 21%, kluster kedua terdapat 143 data penduduk dengan persentase 66%, dan kluster ketiga terdapat 28 data penduduk dan hasil persentase 13% [5]. Penelitian yang dilakukan oleh Azis, Defit, dan Yunus untuk klasterisasi dana bantuan pada Program Keluarga Harapan (PKH), menemukan bahwa perhitungan yang dilakukan dalam pengelompokan menggunakan 10 data desa menunjukkan untuk (C1) sebanyak 5 data desa, untuk (C2) sebanyak 4 data desa, dan untuk (C3) sebanyak 1 data desa [6]. Penelitian yang dilakukan oleh Desi, dkk. menyatakan bahwa implementasi algoritma k-means untuk penerimaan siswa baru berdasarkan nilai rapot dan hasil tes cukup efektif dan sangat [7]. Nabila, dkk. juga menemukan bahwa k-means *clustering* untuk pengelompokkan kasus covid-19 di Provinsi Lampung memberikan hasil pengujian dalam satuan DBI mendekati 0 [8]. Paper ini menyajikan pendalaman implementasi k-means untuk mengelompokkan tingkat kesejahteraan di desa X sehingga diharapkan mampu membantu pemerintah desa setempat untuk menentukan siapa yang perlu mendapat BLT dengan cara yang efektif dan berkeadilan.

2. Metode Penelitian

Bagian ini menjelaskan gambaran mengenai data dan proses untuk mengolah data sehingga dapat diperoleh informasi baru yang bermakna bagi pengambil keputusan dalam hal ini pemerintah desa untuk memutuskan penduduk yang bisa mendapatkan bantuan BLT.

2.1. Data Penelitian

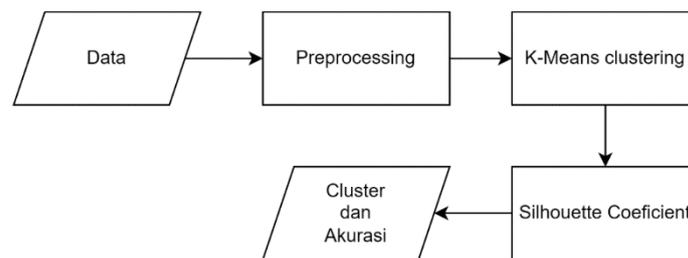
Data yang digunakan dalam penelitian diperoleh dalam bentuk *file* xls yang berisikan informasi hasil survei secara langsung kepada 188 keluarga di desa X. Terdapat 31 keterangan atau atribut pada file tersebut, yang berisi informasi data alamat provinsi, kabupaten/kota, kecamatan, desa, RT, RW, nama kepala keluarga, alamat, nomor HP, nomor telepon kabel/rumah, jumlah keluarga, status kepemilikan tempat tinggal, status lahan tinggal, luas lantai, luas lahan, bahan dinding, keberadaan jendela, bahan atap, jenis penerangan, energi yang digunakan untuk memasak, fasilitas pembuangan sampah, sarana MCK, sumber air, sarana buang air besar, tempat pembuangan limbah cair, lokasi rumah berada di bawah SUTET/SUTT/SUTTAS, rumah di bantaran sungai, rumah di lereng bukit / gunung, kondisi rumah, dan rata-rata pengeluaran per bulan. Pada data tersebut juga terdapat keterangan 39 data mendapatkan bantuan dan sisanya tidak mendapatkan bantuan. Berdasarkan informasi awal dari data tersebut, dan mendasarkan pada fakta-fakta terkait tingkat kesejahteraan, perlu dilakukan berbagai upaya penyiapan data mengikuti proses KDD [9].

Pertama-tama adalah dilakukan proses secara manual untuk menghapus atribut-atribut yang dinilai tidak berkontribusi pada tujuan penelitian, yaitu atribut yang mempunyai informasi yang sama untuk semua keluarga sehingga tidak bisa menjadi pembeda, atribut terkait informasi alamat tempat tinggal, atribut yang bersifat rahasia atau yang berkaitan dengan informasi pribadi seseorang. Maka untuk proses selanjutnya

informasi dari atribut-atribut yang digunakan adalah data dari jumlah keluarga, kepemilikan tempat tinggal, kepemilikan lahan, luas lantai, luas lahan, jenis lantai, bahan dinding, keberadaan jendela, bahan atap, jenis penerangan, energi untuk masak, keberadaan tempat sampah, sarana MCK, sumber air, sarana buang air besar, dan rata-rata pengeluaran per bulan.

2.2. Prosedur Penelitian

Menilik dari hasil-hasil penelitian terkait pemanfaatan algoritma *k-means clustering* untuk membantu proses penentuan berdasarkan pada kelompok tertentu [2]- [8], maka untuk mengelompokkan tingkat kesejahteraan penduduk di desa X guna menentukan mendapat BLT atau tidak, dipilih algoritma tersebut, dengan mengambil ukuran keberhasilan proses dari nilai *silhouette index* atau SI. Gambar 1 menjelaskan bagan prosedur yang ditempuh dalam penelitian dan diimplementasikan menggunakan bahasa pemrograman python.



Gambar 5. Bagan aliran proses mengubah data input menjadi *output*.

Pada tahap *Preprocessing*, data yang telah diproses secara manual sebelumnya akan melalui proses pembersihan data secara otomatis untuk menghilangkan atribut yang memiliki nilai kosong dengan fungsi `dropna()` pada *library* `pandas`, dan akan menghilangkan salah satu dari data yang memiliki nilai atribut sama dengan fungsi `drop_duplicate()` pada *library* `pandas`.

Pada tahap *K-Means Clustering* akan dilakukan proses pengelompokan data menggunakan algoritma *k-means clustering*. Algoritma *k-means* dipilih karena sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, dan populer digunakan dalam praktiknya [10]- [11]. Algoritma ini mendefinisikan *centroid* atau pusat kluster dari rata-rata poin kluster tersebut [9].

Setelah proses pengelompokan selesai, untuk mengukur keberhasilan prosesnya dilakukan perhitungan nilai SI nya pada tahap *Silhouette Coefficient*. Dari nilai SI dapat dilihat kualitas sebuah kluster dengan cara menghitung nilai kohesi dan separasinya [9]. Sehingga setelah keseluruhan proses selesai, akan diperoleh hasil kluster dan nilai akurasinya. Dalam implementasinya akan digunakan fungsi `silhouette_score()` yang didapatkan dari *library* `sklearn`.

2.3. Desain Percobaan

Untuk mendapatkan hasil pengelompokan yang optimal, dilakukan berbagai rencana skenario pengujian. Akan digunakan pengujian dengan mengubah jumlah kluster dari 2, 3, ke 4, dikombinasikan dengan variasi jumlah atribut yang digunakan yaitu sebanyak 16, 15, 14, 13, dan 12 atribut, dan menggunakan perhitungan jarak euclidean. Setelah diperoleh kombinasi atribut yang terbaik pada pengujian dengan 2 kluster, maka akan dilanjutkan penggunaan atribut-atribut tersebut untuk pengujian pada 3 dan 4 kluster. Pada setiap skenario yang digunakan, percobaan dilakukan sebanyak 100 kali. Skenario terakhir akan dicoba pula memanfaatkan rumus jarak manhattan dari kombinasi terbaik pada skenario pengujian-pengujian sebelumnya [12]. Tabel 1 menunjukkan rangkuman skenario pengujian yang akan dilakukan.

Tabel 4. Skenario pengujian.

Bagian ke	Jumlah kluster	Jumlah atribut	Rumus Jarak
1	2	16, 15, 14, 13, 12	euclidean
2	3	16, 15, 14, 13, 12	euclidean
3	4	16, 15, 14, 13, 12	euclidean
4	Sesuai hasil terbaik dari percobaan di Bagian ke 1, 2, 3	Sesuai hasil terbaik dari percobaan di Bagian ke 1, 2, 3	manhattan

Dari seluruh percobaan yang dilakukan dengan jumlah kluster 2, akan dipilih skenario atribut-atribut dan rumus jarak terbaik untuk implementasi akhir pembagian data. Jumlah kluster 2 dipilih karena

secara realita data penduduk hanya dibagi menjadi 2 bagian yaitu yang mendapat bantuan atau tidak mendapat bantuan dana desa oleh pemerintah desa setempat. Hasil dari percobaan akan dibandingkan dengan pembagian yang dilakukan oleh pemerintah desa sebagai pembanding, sehingga dapat diperoleh nilai *ground truthnya*.

3. Hasil dan Pembahasan

Bagian ini memaparkan rangkaian hasil-hasil yang diperoleh pada semua bagian dari skenario pengujian seperti tertera pada Tabel 1 serta pembahasannya, dan telaah hasil penerapan percobaan pada data di lapangan.

3.1. Hasil Percobaan dan Pembahasannya

Sesuai dengan skenario pengujian yang telah ditetapkan, percobaan dimulai dengan menerapkan jumlah kluster sebanyak 2 dengan menggunakan 16 atribut. Program akan dijalankan sebanyak 100 kali, sehingga akan diperoleh informasi 100 nilai SI, dan 100 kemungkinan hasil pengelompokkan. Tabel 2 menunjukkan rangkuman dari 100 kali percobaan tersebut.

Tabel 2. Rangkuman hasil percobaan dengan jumlah kluster 2 dan 16 atribut.

Nilai SI	Jumlah data di kluster 1 dan 2	Jumlah percobaan
0,625256013	[38], [150]	86
0,889598471	[185], [3]	11
0,910504057	[2], [186]	3

Bila ditilik dari nilai SI pada Tabel 2, maka hanya 14 kali percobaan dari 100 kali percobaan yang bisa memperoleh struktur kelompok yang kuat, sedangkan struktur kelompok lainnya dikategorikan baik. Maka dapat disimpulkan bahwa pengelompokkan dengan jumlah kluster 2 dan 16 atribut memiliki kecenderungan mempunyai struktur kelompok kategori baik. Dengan cara pengujian yang sama, dilakukan kembali 100 kali percobaan untuk kombinasi atribut dan hasil kategori struktur pengelompokkan seperti terlihat pada Tabel 3.

Tabel 3. Rangkuman hasil percobaan dengan jumlah kluster 2 dengan 15, 14, 13, dan 12 atribut.

Jumlah atribut	Atribut yang tidak digunakan	Kategori struktur	Prosentase jumlah percobaan dalam kategori struktur
15	Pengeluaran	Kuat	99
15	Sampah	Baik	90
15	Penerangan	Baik	89
14	Pengeluaran dan Penerangan	Kuat	100
14	Pengeluaran dan MCK	Kuat	100
14	Penerangan dan MCK	Baik	88
13	pengeluaran per bulan, MCK, dan penerangan	Kuat	100
13	pengeluaran per bulan, MCK, sampah	Kuat	100
13	penerangan, MCK, sampah	Baik	83
12	tempat tinggal, lahan, masak, dan jendela	Baik	87
12	tempat tinggal, lahan, masak, dan pengeluaran per bulan	Kuat	100
12	tempat tinggal, lahan, masak, dan MCK	Kuat	100

Dari kategori-kategori struktur kelompok seperti terlihat pada Tabel 3, dapat disimpulkan bahwa pada keseluruhan disain percobaan menggunakan variabel-variabel yang ditentukan telah dominan menunjukkan bahwa struktur kluster kuat. Artinya bahwa jika data dibagi menjadi 2 kelompok menggunakan atribut yang dipilih dalam penelitian ini, semua hasil pengelompokkannya memungkinkan digunakan sebagai acuan dalam penentuan keputusan pembagian tingkat kesejahteraan penduduk.

Dari rangkaian percobaan pada bagian pertama tersebut, juga diperoleh informasi nilai SI tertinggi per ketentuan jumlah atribut yang dipilih, terutama untuk jumlah atribut 12 sampai dengan 15. Atas dasar nilai SI tertinggi per variabel jumlah atribut yang ditentukan, maka percobaan pada bagian kedua yaitu membagi data menjadi 3 dan percobaan bagian ketiga yaitu membagi data menjadi 4 dilakukan. Hasil dari rangkaian percobaan bagian 3 dan 4 terlihat pada Tabel 4.

Tabel 4. Hasil percobaan dengan jumlah kluster 3 atau 4 untuk 100 kali per percobaan.

Jumlah kluster	Jumlah atribut	Kategori struktur	Prosentase jumlah percobaan
3	16	Baik	100

3	15	Baik	90
3	14	Baik	94
3	13	Baik	90
3	12	Baik	96
4	16	Baik	93
4	15	Baik	99
4	14	Baik	100
4	13	Baik	100
4	12	Baik	100

Dari rangkaian percobaan pada bagian kedua dan ketiga, dapat disimpulkan bahwa pembagian data menjadi 3 dan 4 kelompok tidak lebih baik hasilnya. Dari semua percobaan yang dilakukan, struktur kelompok yang dihasilkan berada pada kategori baik. Artinya bahwa pembagian kelompok yang paling baik adalah tetap membaginya menjadi 2 saja.

Dari keseluruhan rangkaian percobaan pada bagian pertama, kedua, dan ketiga seperti terlihat pada skenario pengujian di Tabel 1, maka pada percobaan di skenario keempat digunakan variabel jumlah kluster sebanyak 2 dan jumlah atribut sebanyak 12 karena data nilai SI tertinggi diperoleh pada kombinasi ini, namun dengan rumus jarak manhattan. Dari 100 kali percobaan klustering, diperoleh 1 saja nilai SI untuk semua percobaan yaitu sebesar 0,728278964. Kluster sudah dikategorikan kuat meski masih tipis kekuatannya. Hasil penilaian kluster ini relevan jika dibandingkan dengan hasil-hasil pada penelitian sebelumnya [2]- [8]. Dari percobaan juga diperoleh informasi jumlah anggota kelompok data di kluster satu sebanyak 35, dan di kluster dua sebanyak 153.

Dari keseluruhan rangkaian percobaan, nilai SI terbaik didapatkan di jumlah kluster 2 dengan 12 atribut, yaitu atribut jumlah keluarga, luas lahan, luas lantai, jenis lantai, bahan dinding, keberadaan jendela, bahan atap, jenis penerangan, sarana pembuangan sampah, MCK, sumber air, dan sarana buang air besar. Baik menggunakan rumus jarak euclidean maupun manhattan, ternyata hasil nilai SI yang didapatkan sama besarnya, sehingga dapat disimpulkan bahwa penerapan kedua rumus jarak berbeda tidak membuat perbedaan pada nilai SI yang dihasilkan. Hal ini sedikit berbeda dari penelitian Sinwar dan Kaushik yang menyatakan bahwa penggunaan rumus jarak euclidean lebih baik dibandingkan dengan rumus jarak manhattan [12]. Dari percobaan-percobaan yang telah dilakukan, dan hasil yang didapatkan dari percobaan, maka dapat juga disimpulkan bahwa yang mempengaruhi nilai SI pada penelitian ini adalah atribut yang dipilih, penentuan titik pusat awal, dan jumlah kluster dan hal ini sejalan dengan sifat dari algoritma k-means sendiri bahwa perbedaan pada nilai-nilai awal pusat data yang kluster, inisialisasi nilai awal pusat kluster akan mempengaruhi pengelompokan yang terjadi [9].

3.2. Pembahasan Penerapan Hasil Percobaan

Mempergunakan hasil temuan nilai SI tertinggi pada berbagai skenario pengujian yang diimplementasikan pada percobaan, maka diperoleh hasil pengelompokan data penduduk menjadi 2 bagian, seperti terlihat sebagian hasilnya pada Gambar 2. Terdapat kolom kluster yang berisi informasi nama kelompok, dan kolom Nama yang telah disamarkan.

cluster	Nama
0	J [redacted]
1	a [redacted]
1	N [redacted]
1	S [redacted]
0	A [redacted]
1	R [redacted]
1	T [redacted]
0	S [redacted]

Gambar 2. Tampilan sebagian data hasil pengelompokkan dengan atribut Nama yang disamarkan.

Dari hasil pengelompokkan yang telah didapatkan tersebut, dapat diambil kesimpulan bahwa data yang berada di kelompok 0 adalah keluarga yang mendapatkan bantuan dana desa, dan kelompok 1 adalah keluarga yang tidak mendapatkan bantuan dana desa. Dari hasil perhitungan secara manual data dalam kelompok yang terjadi, didapatkan jumlah data benar di kelompok 0 sebanyak 9 data, dan di kelompok 1 sebanyak 123 data dari total data sebanyak 188. Dari hasil ini, maka didapatkan nilai *ground truth* sebesar 70,12%. Besarnya nilai *ground truth* yang diperoleh menunjukkan bahwa penerapan k-means clustering pada pengelompokkan data penduduk desa X untuk kepentingan pemberian bantuan dana desa dapat memberikan manfaat bagi pemerintah sebagai pengambil keputusan.

4. Kesimpulan

Mengacu pada hasil percobaan untuk mendalami persoalan kesulitan pengelompokan data tingkat kesejahteraan penduduk di desa X, didapatkan kesimpulan bahwa penggunaan algoritma k-means *clustering* untuk pengelompokan data dari data yang dipunyai mampu memberikan panduan bagi pemerintah desa tersebut. Hal ini dapat dilihat dari nilai SI yang dominan kuat untuk semua skenario percobaan, dan dari nilai *ground truth*nya. Pengelompokan data menjadi 2 bagian dan penggunaan atribut-atribut jumlah keluarga, luas lahan, luas lantai, jenis lantai, bahan dinding, keberadaan jendela, bahan atap, jenis penerangan, sarana pembuangan sampah, fasilitas MCK, sumber air, dan sarana buang air besar sudah mampu memberikan nilai SI yang kuat, sehingga hasil pengelompokan oleh sistem dapat digunakan sebagai acuan bagi pemerintah desa dalam mengambil keputusan mengenai siapa penduduk yang mendapatkan bantuan dana desa.

Penelitian ini dapat dilanjutkan dengan mencoba menggagas penggunaan atribut-atribut lainnya yang telah disingkirkan dalam penelitian ini, meski pemilihan atribut sudah didasarkan pada cara umum memandang apakah seseorang berada di tingkat sejahtera atau tidak. Selain itu, dalam keseluruhan percobaan ini semua atribut masih dinilai mempunyai bobot yang sama, karena memang tidak ada informasi mengenai kedudukan informasi dalam atribut yang sesuai dikaitkan dengan kesejahteraannya.

Daftar Pustaka

- [1] G. Irawan, "Wamendes Ungkap 3 Persoalan Utama dalam Penyaluran BLT Dana," 4 Juni 2020. [Online]. Available: <https://www.tribunnews.com/nasional/2020/06/04/wamendes-ungkap-3-persoalan-utama-dalam-penyaluran-blt-dana-desa>.
- [2] Y. Kusnadi and M. S. Putri, "Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah," *Teknologi Informatika dan Komputer MH. Thamrin*, vol. 7, no. 1, pp. 17 - 24, 2021.
- [3] D. O. Yurnas, D. Sarwinda and F. Muttakin, "Pengelompokan Penerima Bantuan Kesejahteraan Masyarakat Dengan Pendekatan Data Mining Terintegrasi Sistem Pendukung Keputusan," *Prosiding Seminar Ilmiah Nasional Teknologi Komputer*, vol. 1, pp. 14 - 19, 2015.
- [4] A. Muhidin and I. Baragigiratri, "Pemetaan Penduduk Calon Penerima Bantuan Renovasi Rumah Desa Pesangkalan Menggunakan Algoritma K-Means," *Teknologi Pelita Bangsa - SIGMA*, vol. 8, no. 2, pp. 147 - 152, 2018.
- [5] T. Noviana, Jasmir and Y. Novianto, "Penerapan Data Mining Menentukan Kelompok Prioritas Penerima Bantuan Beras Rastra Dengan Clustering K-Means," *Jurnal Ilmiah Mahasiswa Teknik Informatika*, vol. 1, no. 3, pp. 159 - 174, 2019.
- [6] S. A. Azis, S. Defit and Y. Yunus, "Klasterisasi Bantuan Pada Program Keluarga Harapan (PKH) Menggunakan Metode K-Means," *Informatika Ekonomi Bisnis*, vol. 3, no. 2, pp. 53 - 59, 2021.
- [7] E. Desi and dkk, "implementasi Algoritma K-Means Untuk Penerimaan Siswa," *IT Jurnal*, vol. 10, no. 1, pp. 1 - 10, 2022.
- [8] Z. Nabila and dkk, "Analisis Data Mining Untuk Clustering Kasus Covid-19 di Provinsi Lampung Dengan Algoritma K-Means," *Media Infotama*, vol. 2, no. 2, pp. 100 - 108, 2021.
- [9] J. Han, M. Kamber and j. Pei, Data Mining: Concepts and Techniques, 3rd ed., *New York: Morgan Kaufman*, 2012, pp. 489 - 490.
- [10] X. Wu and V. Kumar, The Top Ten Algorithms in Data Mining, *Boca Raton: Chapman & Hall/CRC*, 2009, pp. 21 - 35.
- [11] A. Jain, Data Clustering: 50 Years Beyond K-Means, *East Lansing, Michigan: Michigan State University*, 2009, pp. 651 - 666.
- [12] D. Sinwar and R. Kaushik, "Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering," *International Journal For Research In Applied Science And Engineering.Technology (IJRASET)*, vol. 2, no. 5, pp. 270 - 274, 2014.