

Implementasi Metode K-Medoids untuk Clustering Penerima Bantuan Berdasarkan Normalisasi Data Masyarakat Miskin dengan Metode Desimal Scaling

Rispani

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email: rispani0605@gmail.com

Abstrak—Kantor Dinas Sosial adalah penyalur bantuan untuk sosial masyarakat miskin yang sesuai dengan peraturan ditetapkan pemerintah Menteri Sosial Republik Indonesia Nomor 20 Tahun 2019. yaitu merupakan bantuan kepada masyarakat miskin yang sifatnya tidak secara terus menerus dan selektif dalam bentuk barang atau uang kepada penerima masyarakat miskin yang bertujuan untuk peningkatan kesejahteraan masyarakat miskin yang tidak mampu dan rentan terhadap risiko sosial. Data penerima bantuan masyarakat miskin yang perlu diolah dan dilakukan penormalisasian data yang bertujuan agar mendapatkan informasi yang diinginkan dari data penerima bantuan masyarakat miskin guna untuk memudahkan dalam pengelompokan penerima bantuan masyarakat miskin di Kantor Dinas Sosial Aceh Tenggara. Penerima bantuan masyarakat miskin yang perlu diolah dan dilakukan penormalisasian data yang bertujuan agar mendapatkan informasi yang diinginkan dari data penerima bantuan masyarakat miskin guna untuk memudahkan dalam pengelompokan dengan Metode Desimal Scaling dalam menormalisasikan data penerima bantuan masyarakat miskin dijadikan data dalam memperoleh informasi yang diinginkan kemudian dicluster dengan menggunakan Metode K-Medoids untuk clustering penerima bantuan berdasarkan normalisasi data penerima bantuan masyarakat miskin untuk membantu mempermudah dalam pengelompokan data penerima bantuan masyarakat miskin yang terbaik serta menggunakan sebuah sistem yang mampu memberikan solusi untuk mengcluster data penerima bantuan miskin dengan menggunakan metode K-Medoids dan aplikasi Ripedminer.

Kata Kunci: Normalisasi; Clustering; Data Bantuan; RapidMiner

Abstract—The Social Welfare Office is the distributor of aid for the economically disadvantaged population in accordance with the regulations set by the Minister of Social Affairs of the Republic of Indonesia Number 20 of 2019. This aid is provided to the economically disadvantaged population selectively, not continuously, in the form of goods or cash, aiming to improve the welfare of the economically disadvantaged and socially vulnerable. The data of aid recipients from the economically disadvantaged population needs to be processed and normalized to obtain the desired information, facilitating the grouping of aid recipients at the Southeast Aceh Social Welfare Office. The aid recipients' data is processed and normalized to ease the grouping process using Decimal Scaling method, enabling the extraction of desired information. Subsequently, the data is clustered using the K-Medoids method to group aid recipients based on the normalized data, thus simplifying the identification of the most suitable aid recipients. This research employs a system capable of providing a solution for clustering aid recipient data using the K-Medoids method and the RapidMiner application.

Keywords: Normalization; Clustering; Aid Data; RapidMiner

1. PENDAHULUAN

Penerima Bantuan Sosial adalah bantuan dari pemerintah yang diberikan kepada masyarakat miskin yang tidak mampu dan penyandang masalah kesejahteraan sosial melalui Kantor Dinas Sosial yang sesuai dengan peraturan ditetapkan Menteri Sosial Republik Indonesia Nomor 20 Tahun 2019 yaitu merupakan bantuan kepada masyarakat miskin yang sifatnya tidak secara terus menerus dan hanya dalam bentuk pangan (sembako) kepada penerima masyarakat miskin yang bertujuan untuk peningkatan kesejahteraan masyarakat miskin yang tidak mampu dan rentan terhadap risiko sosial[1][2]. Dari jumlah data yang diterima di tahun 2020 oleh Kantor Dinas Sosial dari data masyarakat miskin sekitar 8118 identitas lebih dalam pendistribusiannya dari pihak Dinas Sosial kepada penerima bantuan masyarakat miskin.

Dari penelitian di dapat untuk data masyarakat miskin yang perlu diolah dan dilakukan penormalisasian data yang bertujuan agar mendapatkan informasi yang diinginkan dari data masyarakat miskin guna untuk memudahkan dalam pengelompokan penerima bantuan masyarakat miskin di Kantor Dinas Sosial Aceh Tenggara, Untuk itu perlunya penormalisasian data masyarakat miskin salah satunya dengan menggunakan Metode Desimal Scaling dalam menormalisasikan data tersebut akan dijadikan data dalam memperoleh informasi yang diinginkan, Kemudian di cluster dengan menggunakan Metode K-Medoids untuk clustering penerima bantuan berdasarkan normalisasi data masyarakat miskin untuk membantu mempermudah dalam pengelompokan data penerima bantuan masyarakat miskin yang terbaik.

Penelitian sejenis pernah dilakukan oleh Nurhayati, dkk (2018) dengan judul Analysis of K-Mend and K-Medoids Performance Using Big Data Technology. Penelitian ini membandingkan Algoritma K-Mens dan K-Medoids dengan menguji data dengan aplikasi berbasis Java, Hadoop, dan Hive. Parameter yang digunakan dalam pengujian ini adalah akurasi, eksekusi. Dari pengujian tersebut diperoleh kesimpulan bahwa algoritma K-Medoids lebih baik bila dibandingkan algoritma K-Means[3].

Menurut Hae-Sang Park dan ChiHyuck Jun, pembahasan tentang sebuah algoritma baru untuk K-medoids dalam pengelompokan yang berjalan seperti K-means dan tes beberapa metode untuk memilih medoids awal. Algoritma yang diusulkan menghitung jarak metrik dan menggunakan untuk menemukan medoids baru di setiap iterasi langkah. Untuk mengevaluasi algoritma yang diusulkan, peneliti menggunakan beberapa data set nyata dan buatan serta dibandingkan dengan hasil algoritma lain dalam hal pembuatan indeks. Hasil penelitian menunjukkan bahwa algoritma yang diusulkan

membutuhkan waktu berkurang secara signifikan dalam perhitungan dengan kinerja yang sebanding terhadap partisi di sekitar medoids.

Decimal Scaling Normalization merupakan metode normalisasi dengan menggerakkan nilai desimal dari data ke arah yang diinginkan. Clustering merupakan suatu metode pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki 18 kedekatan (kemiripan)[4][5]. Cluster adalah kumpulan beberapa record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidak miripan dengan beberapa record di cluster yang lain. Clustering berbeda dengan klasifikasi, yaitu tidak ada variabel target dalam clustering[6]. Clustering tidak mencoba untuk melakukan klasifikasi, memperkirakan, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma clustering mencoba untuk melakukan pembagian terhadap semua data menjadi beberapa kelompok yang memiliki kedekatan/kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai lebih besar, sedangkan kemiripan dengan record dikelompok lain akan bernilai lebih kecil[7].

Data mining adalah suatu proses penambahan informasi penting dari suatu data. Informasi penting ini didapat dari suatu proses yang amat rumit seperti menggunakan artificial intelligence, teknik statistik, ilmu matematika, machinelearning dan lain sebagainya. Teknik-teknik rumit tersebut nantinya akan mengidentifikasi dan mengekstraksi informasi yang bermanfaat dari suatu database besar[8][9][10].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metodologi penelitian merupakan suatu strategis yang disusun sedemikian rupa, untuk mencari tahu suatu kebenaran atau kenyataan terhadap yang dikaji berdasarkan masalah yang terjadi. pada metode penelitian ini disusun sangat terstruktur guna untuk mendapatkan suatu kebenaran atau fakta dari beberapa sumber masalah dengan kajian yang ada untuk menjangkau hasil akhir dan memberikan solusi yang ideal terhadap waktu penelitian yang dilakukan berdasarkan sumber data yang ada dan kondisi suatu tempat sehingga dengan strategis yang disusun tersebut dapat diolah dengan baik. Ada beberapa tahapan-tahapan yang dilakukan penulis dalam penelitian ini untuk mengumpulkan suatu data. Tahapan tersebut adalah sebagai berikut:

1. Identifikasi masalah

Identifikasi masalah yaitu uraian masalah yang melatar belakangi pembuatan skripsi ini tentang permasalahan dalam normalisasi data masyarakat miskin dan mengcluster data penerima bantuan masyarakat miskin di Kantor Dinas Sosial Aceh Tenggara.

2. Studi lapangan

Pada tahapan ini penulis melakukan pengamatan secara langsung pada Kantor Dinas Sosial Aceh Tenggara untuk mengetahui informasi mengenai masalah yang ada.

3. Studi literature

Tahapan ini dilakukan untuk menghimpun data-data atau sumber-sumber yang berhubungan dengan data masyarakat miskin, dapat diperoleh dari berbagai sumber jurnal, buku, dokumentasi, internet dan pustaka.

4. Pengumpulan data

Tahapan ini bermaksud untuk mencari data pada Kantor Dinas Sosial Aceh Tenggara, yang akan dijadikan untuk menjawab permasalahan. Pada pengumpulan data ini sangat diperlukan untuk memperoleh data yang berkualitas. strategi pengumpulan data dapat dilakukan dengan cara:

a. Wawancara (Interview)

Kegiatan ini dengan bertanya langsung kepada narasumber dan pihak-pihak yang terkait proses data masyarakat miskin untuk mendapatkan data sesuai kenyataan.

b. Pengamatan (Observasi)

Tahapan ini melakukan pengamatan langsung terhadap situasi atau peristiwa yang ada pada kantor dinas sosial aceh tenggara.

5. Analisa

Pada tahap proses menganalisa data didapatkan dari objek penelitian yang dilakukan di lokasi, dan serta penyelesaian masalah menggunakan Metode Desimal Scaling untuk menormalisasi data masyarakat miskin dan menggunakan Metode K-Medoids untuk mengcluster penerima bantuan masyarakat miskin.

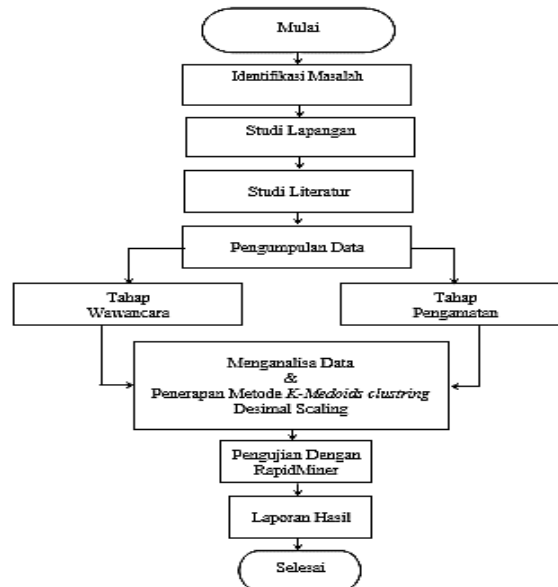
6. Pengujian Hasil

Tahapan ini adalah tahap akhir dimana dari sekumpulan data yang diperoleh akan diuji melalui perangkat lunak yaitu aplikasi RapidMiner yang digunakan sebagai aplikasi pengujian.

7. Penulisan Laporan

Penulisan laporan yang dilakukan untuk mendokumentasikan seluruh kegiatan penelitian dalam bentuk skripsi dan nantinya akan dibuat dalam bentuk artikel ilmiah yang akan dipublikasikan.

Dari tahapan tersebut untuk memudahkan penulis dapat mengumpulkan data melalui penelitian sehingga dapat memanfaatkan data tersebut sebagai bahan pengujian untuk memenuhi data yang kurang sehingga data yang dapat terpenuhi diperoleh data benar-benar sesuai dengan yang dibutuhkan dalam pengumpulan tersebut untuk memperlancar melakukan pengujian.



Gambar 1. Flowchart Tahapan Penelitian

2.2 Penerima Bantuan Masyarakat Miskin

Penerima bantuan masyarakat miskin adalah seseorang atau keluarga miskin yang tidak mampu dan penyandang masalah kesejahteraan sosial. Pengelolaan data yang menerima bantuan masyarakat miskin yang disalurkan melalui Kantor Dinas Sosial dan diberikan kepada masyarakat miskin dengan data yang sesuai syarat yang ditetapkan pemerintah Menteri Sosial Republik Indonesia Nomor 20 Tahun 2019. Hal itu guna untuk meningkatkan kesejahteraan masyarakat miskin melalui oleh Kantor Dinas Sosial[11][12].

2.3 Data Mining

Data mining adalah suatu proses penambahan informasi penting dari suatu data. Informasi penting ini didapat dari suatu proses yang amat rumit seperti menggunakan artificial intelligence, teknik statistik, ilmu matematika, machinelearning dan lain sebagainya[13]. Teknik-teknik rumit tersebut nantinya akan mengidentifikasi dan mengekstraksi informasi yang bermanfaat dari suatu database besar. Contoh mudahnya misalnya membaca data nomor kartu keluarga, Setelah selesai mendapatkan informasi bahwa mayoritas orang dengan nomor kartu keluarga, maka hal tersebut dapat dikatakan sebagai proses[10][14].

Jadi, berdasarkan penjelasan di atas dapat disimpulkan bahwa data mining merupakan suatu proses penambahan data dalam jumlah data yang sangat besar dengan menggunakan metode statistika, matematika hingga memanfaatkan teknologi artificial intelligenceter kini.

2.4 Normalisasi Data

Normalisasi merupakan sebuah teknik dalam logical desain sebuah basis data yang mengelompokkan atribut dari suatu relasi sehingga membentuk struktur relasi yang baik tanpa redundansi. Normalisasi adalah proses pembentukan struktur basis data sehingga sebagian besar ambiguity bisa dihilangkan[15]. Normalisasi merupakan parameter digunakan untuk menghindari duplikasi terhadap tabel dalam basis data dan juga merupakan proses merekomposisikan sebuah tabel yang masih memiliki beberapa anomali atau ketidak wajaran sehingga menghasilkan tabel yang lebih sederhana dan struktur yang bagus, yaitu sebuah tabel yang tidak memiliki data redundancy dan memungkinkan user untuk melakukan insert, delete, dan update pada baris (record) tanpa menyebabkan inkonsistensi data[16].

2.5 Metode Desimal Scaling

Decimal Scaling Normalization merupakan metode normalisasi dengan menggerakkan nilai desimal dari data ke arah yang diinginkan, clustering merupakan suatu metode pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki 18 kedekatan (kemiripan)[17]. Cluster adalah kumpulan beberapa record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidak miripan dengan beberapa record di cluster yang lain. clustering berbeda dengan klasifikasi, yaitu tidak ada variabel target dalam clustering. Clustering tidak mencoba untuk melakukan klasifikasi, memperkirakan, dan memprediksi nilai dari variabel target. Akan tetapi, algoritma clustering mencoba untuk melakukan pembagian terhadap semua data menjadi beberapa kelompok yang memiliki kedekatan dan kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai lebih besar, sedangkan kemiripan dengan record dikelompok lain akan bernilai lebih kecil maka rumus yang digunakan[5].

$$newdata = \frac{data}{10^i newdata} = data \text{ hasil normalisasi} \quad (1)$$

Keterangan :

newdata = data hasil normalisasi

i = adalah nilai scaling yang kita inginkan

2.6 Metode K-Medoids

K-Medoids merupakan salah satu metode partitional clustering yang meminimalkan jarak antara titik berlabel berada dalam cluster dan titik yang ditunjuk sebagai pusat klaster, tidak sama dengan algoritma K-Means, K-Medoids memilih data points sebagai pusat (medoids)[18]. K-Medoids adalah teknik partisi klasik Clustering yang mengelompokkan data set dari objek ke dalam kelompok yang dikenal apriori[19]. Dibandingkan dengan K-Means, K-Medoids lebih kuat untuk menangani kebisingan (noise) dan orang asing (outlier) karena meminimalkan beberapa dissimilarities berpasangan, bukan jumlah kuadrat jarak Euclidean. Sebuah medoids dapat diartikan sebagai objek cluster yang rata-rata perbedaan untuk semua objek dalam cluster minimal yaitu titik paling berlokasi di cluster maka rumus yang digunakan Rumus[20]:

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

Keterangan:

d(x,y) = tingkat perbedaan

n = nilai vektor

x_i = vektor citra input

y_i = vektor citra pembanding/output

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Masalah yang didapat dari data masyarakat miskin di Kantor Dinas Sosial adalah kurangnya dalam pengolahan data untuk penerimaan bantuan masyarakat miskin karna data yang tidak normal sehingga perlu di normalisasikan untuk memudahkan pengelompokan data dalam penerima bantuan masyarakat miskin, tentu dengan cara tersebut dapat mengetahui kelompok data penerima bantuan masyarakat miskin yang terbaik dan diperlukan suatu sistem yang dapat mempermudah pihak Kantor Dinas Sosial dalam mengelompokkan penerima bantuan masyarakat miskin di Kantor Dinas Sosial (Tanah Merah, Badar, Kutacane, Aceh Tenggara) yang artinya memerlukan dalam pengelompokan data penerima bantuan masyarakat miskin. Adapun data yang digunakan dalam penelitian ini merupakan data pilihan yang diberikan oleh pihak Kantor Dinas Sosial.

Data yang digunakan adalah data dua tahun terakhir (2020) tahap analisa dan pembahasan ini, algoritma yang dipakai dalam menormalisasikan data masyarakat miskin menggunakan Metode Desimal Scaling dan mengelompokkan data penerima bantuan masyarakat miskin menggunakan algoritma K-Medoids dan sebagai alat bantu penerapan data mining dan RapidMiner terutama membantu menyelesaikan masalah klasifikasi. Dalam menganalisa pengelompokan penerima bantuan masyarakat miskin diperlukan data yang real dari pihak Kantor Dinas Sosial Aceh Tenggara. Data masyarakat miskin yang benar-benar dan sesuai dengan kondisi data yang didapat dari hasil pengumpulan penelitian.

3.2 Normalisasi Data Dengan Desimal Scaling

Berikut ini merupakan tabel 1 dari data masyarakat miskin 2020, data ini meliputi data masyarakat miskin di Kantor Dinas Sosial Aceh Tenggara.

Tabel 1. Data Masyarakat Miskin 2020

No	Nama	Tahun	Umur	Tanggungan	Pendidikan	Pekerjaan
1	Satunan	2020	62	4	SD	Petani/Kebun
2	Rajadin	2020	50	2	SMP	Petani/Kebun
3	Sahbidun	2020	51	2	SD	Petani/Kebun
4	Saparudin	2020	40	3	SMP	Petani/Kebun
5	Hadimah	2020	48	1	SD	Mengurus Rumah Tangga
6	Cungek	2020	38	1	SD	Mengurus Rumah Tangga
7	Hajidan Bahri	2020	40	4	SMA	Petani/Kebun
8	Rustam Efendi	2020	27	3	SMA	Petani/Kebun
9	Atinah	2020	28	3	SMA	Mengurus Rumah Tangga
10	Rusiah	2020	42	1	SD	Mengurus Rumah Tangga
11	Asrah	2020	39	2	SMP	Mengurus Rumah Tangga
12	Erna Wati	2020	31	2	SMP	Mengurus Rumah Tangga
13	Izhar Saidi	2020	33	2	SMA	Petani/Kebun
14	Rojikin	2020	24	2	SMA	Petani/Kebun

No	Nama	Tahun	Umur	Tanggung	Pendidikan	Pekerjaan
15	Anggara	2020	23	2	SMA	Petani/Kebun
16	Rejeki	2020	30	3	SMP	Petani/Kebun
17	Melidawati	2020	29	1	SMP	Mengurus Rumah Tangga
18	Ruzaini	2020	29	2	SMA	Mengurus Rumah Tangga
19	karmida	2020	20	1	SMA	Mengurus Rumah Tangga
20	Mawardi	2020	46	4	SMP	Petani/Kebun

Penyesuaian format data perlu dilakukan untuk mempermudah pengolahan sehingga butuh transformasi data bagi data yang dapat diolah kedalam bentuk angka. Dalam hal ini data keterangan harus ditransfer dalam bentuk angka. Hasil transformasi data sesuai dengan data masyarakat miskin.

Tabel 2. Transformasi Nilai Untuk Keterangan

No	Pendidikan	Pekerjaan
1	1 = SMA	1 = Petani/Kebun
2	2 = SMP	2 = Mengurus Rumah Tangga
3	3 = SD	

Tabel 3. Transformasi Data Masyarakat Miskin 2020

No	Nama	Tahun	Umur	Tanggung	Pendidikan	Pekerjaan
1	Satunan	2020	62	4	1	1
2	Rajadin	2020	50	2	2	1
3	Sahbidun	2020	51	2	2	1
4	Saparudin	2020	40	3	2	1
5	Hadimah	2020	48	1	1	2
6	Cungek	2020	38	1	1	2
7	hajidan bahri	2020	40	4	3	1
8	rustam efendi	2020	27	3	3	1
9	Atinah	2020	28	3	3	2
10	Rusiah	2020	42	1	1	2
11	Asrah	2020	39	2	2	2
12	erna wati	2020	31	2	2	2
13	izhar saidi	2020	33	2	3	1
14	Rojikin	2020	24	2	3	1
15	Anggara	2020	23	2	3	1
16	Rejeki	2020	30	3	2	1
17	Melidawati	2020	29	1	2	2
18	Ruzaini	2020	29	2	3	2
19	Karmida	2020	20	1	3	2
20	Mawardi	2020	46	4	2	1

Adapun proses perhitungan metode desimal scaling untuk normalisasi data penerima batuan masyarakat miskin di kantor dinas sosial sebagai berikut:

$$newdata = \frac{data}{10^i} \quad newdata = data \text{ hasil normalisasi}$$

Keterangan :

newdata = data hasil normalisasi

i = adalah nilai scaling yang kita inginkan

Normalisasi Data Umur

$$Newdata_1 = \frac{62}{10^2} = 0,62$$

$$Newdata_2 = \frac{50}{10^2} = 0,50$$

$$Newdata_3 = \frac{51}{10^2} = 0,51$$

$$Newdata_4 = \frac{40}{10^2} = 0,40$$

$$Newdata_5 = \frac{48}{10^2} = 0,48$$

$$Newdata_6 = \frac{38}{10^2} = 0,38$$

$$Newdata_7 = \frac{40}{10^2} = 0,40$$

$$Newdata_8 = \frac{27}{10^2} = 0,27$$

$$Newdata_9 = \frac{28}{10^2} = 0,28$$

$$Newdata_{10} = \frac{42}{10^2} = 0,42$$

$$Newdata_{11} = \frac{39}{10^2} = 0,39$$

$$Newdata_{12} = \frac{31}{10^2} = 0,31$$

$$Newdata_{13} = \frac{33}{10^2} = 0,33$$

$$Newdata_{14} = \frac{24}{10^2} = 0,24$$

$$Newdata_{15} = \frac{23}{10^2} = 0,23$$

$$Newdata_{16} = \frac{30}{10^2} = 0,30$$

$$Newdata_{17} = \frac{29}{10^2} = 0,29$$

$$Newdata_{18} = \frac{29}{10^2} = 0,29$$

$$Newdata_{19} = \frac{20}{10^2} = 0,20$$

$$Newdata_{20} = \frac{46}{10^2} = 0,46$$

Lakukan perulangan seperti diatas untuk menghasilkan normalisasi data tanggungan, pendidikan dan pekerjaan. Sehingga dihasilkan hasil normalisasi yang dapat dilihat pada tabel 4.

Tabel 4. Hasil Normalisasi Data Masyarakat Miskin

No	Nama	Tahun	Umur	Tanggungan	Pendidikan	Pekerjaan
1	Satunan	2020	0,62	0,04	0,01	0,01
2	Rajadin	2020	0,50	0,02	0,02	0,01
3	Sahbidun	2020	0,51	0,02	0,02	0,01
4	Saparudin	2020	0,40	0,03	0,02	0,01
5	Hadimah	2020	0,48	0,01	0,01	0,02
6	Cungek	2020	0,38	0,01	0,01	0,02
7	hajidan bahri	2020	0,40	0,04	0,03	0,01
8	rustam efendi	2020	0,27	0,03	0,03	0,01
9	Atinah	2020	0,28	0,03	0,03	0,02
10	Rusiah	2020	0,42	0,01	0,01	0,02
11	Asrah	2020	0,39	0,02	0,02	0,02
12	erna wati	2020	0,31	0,02	0,02	0,02
13	izhar saidi	2020	0,33	0,02	0,03	0,01
14	Rojikin	2020	0,24	0,02	0,03	0,01
15	Anggara	2020	0,23	0,02	0,03	0,01
16	Rejeki	2020	0,30	0,03	0,02	0,01
17	melidawati	2020	0,29	0,01	0,02	0,02
18	Ruzaini	2020	0,29	0,02	0,03	0,02
19	Karmida	2020	0,20	0,01	0, 03	0,02
20	Mawardi	2020	0,46	0,04	0,02	0,01

3.3 Penerapan Algoritma K-Medoids

Algoritma K-Medoids adalah salah satu algoritma yang digunakan untuk klasifikasi atau Clustering data. Adapun langkah-langkah pengerjaan Algoritma K-Medoids sebagai berikut.

1. Menentukan pusat cluster k sebagai inisial centroid atau nilai tengah (medoids) sebanyak k-cluster.
2. Cari semua poin yang paling dekat dengan medoids, dengan cara menghitung jarak vector antar dokumen dengan menggunakan Euclidian Distance.

3. Secara acak, pilih poin yang bukan medoids.
4. Hitung total jarak antar medoids.
5. Jika TD baru < TD awal, tukar posisi medoids baru, jadilah medoids yang baru. Jika nilai cost baru lebih besar dibandingkan dengan cost lama maka iterasi dihentikan.
6. Ulangi langkah 2-5 sampai medoids tidak berubah.

Tabel 5. Data Masyarakat Miskin

D	X1	X2	X3	X4
D1	0,62	0,04	0,01	0,01
D2	0,50	0,02	0,02	0,01
D3	0,51	0,02	0,02	0,01
D4	0,40	0,03	0,02	0,01
D5	0,48	0,01	0,01	0,02
D6	0,38	0,01	0,01	0,02
D7	0,40	0,04	0,03	0,01
D8	0,27	0,03	0,03	0,01
D9	0,28	0,03	0,03	0,02
D10	0,42	0,01	0,01	0,02
D11	0,39	0,02	0,02	0,02
D12	0,31	0,02	0,02	0,02
D13	0,33	0,02	0,03	0,01
D14	0,24	0,02	0,03	0,01
D15	0,23	0,02	0,03	0,01
D16	0,30	0,03	0,02	0,01
D17	0,29	0,01	0,02	0,02
D18	0,29	0,02	0,03	0,02
D19	0,20	0,01	0,03	0,02
D20	0,46	0,04	0,02	0,01

Iterasi 1 :

Menentukan pusat cluster k secara acak

Medoid 1 (D_5) = (0,48 ; 0,01 ; 0,01 ; 0,02) y_1

Medoid 2 (D_{10}) = (0,42 ; 0,01 ; 0,01 ; 0,02) y_2

Medoid 3 (D_{15}) = (0,23 ; 0,02 ; 0,03 ; 0,01) y_3

Medoid 4 (D_{20}) = (0,46 ; 0,04 ; 0,02 ; 0,01) y_4

Mencari nilai C1, C2, C3 dan C4 dengan menggunakan rumus Euclidian Distance. Sehingga didapatkan hasil yang dapat dilihat pada tabel 6.

Tabel 6. Euclidian Distance Iterasi 1

No	A	B	C	D	dc1	dc2	dc3	dc4	c1	c2	c3	c4
1	0,62	0,04	0,01	0,01	-0,16	-0,22	-0,39	-0,15				Yes
2	0,50	0,02	0,02	0,01	-0,01	-0,08	-0,25	-0,01				Yes
3	0,51	0,02	0,02	0,01	-0,02	-0,09	-0,26	-0,02				Yes
4	0,40	0,03	0,02	0,01	0,08	0,01	-0,16	0,08		Yes		
5	0,48	0,01	0,01	0,02	0	-0,06	-0,23	0,01	Yes			
6	0,38	0,01	0,01	0,02	0,01	-0,09	-0,15	-0,04	Yes			
7	0,40	0,04	0,03	0,01	0,09	-0,01	-0,07	-0,04		Yes		
8	0,27	0,03	0,03	0,01	0,16	-0,03	-0,09	-0,13	Yes			
9	0,28	0,03	0,03	0,02	0,07	0,06	0	-0,11			Yes	
10	0,42	0,01	0,01	0,02	0,16	0,06	0	-0,19			Yes	
11	0,39	0,02	0,02	0,02	0,11	-0,11	-0,06	-0,09			Yes	
12	0,31	0,02	0,02	0,02	0,19	-0,03	0,02	-0,01				Yes
13	0,33	0,02	0,03	0,01	0,17	-0,05	0	0,06			Yes	
14	0,24	0,02	0,03	0,01	0,26	0,04	0,09	-0,03				Yes
15	0,23	0,02	0,03	0,01	0,27	0,05	0,1	0,07			Yes	

No	A	B	C	D	dc1	dc2	dc3	dc4	c1	c2	c3	c4
16	0,30	0,03	0,02	0,01	0,19	0,14	0	-0,02			Yes	
17	0,29	0,01	0,02	0,02	0,21	0,12	0,02	-0,02			Yes	
18	0,29	0,02	0,03	0,02	0,19	0,12	0	0				Yes
19	0,20	0,01	0,03	0,02	0,29	0,22	0,1	0,08			Yes	
20	0,46	0,04	0,02	0,01	0,02	-0,05	-0,17	-0,19	Yes			

$$Total\ cost = -0,15 + + - 0,01 + -0,02 + 0,08 + 0,01 + 0,01 + 0,09 + 0,16 + 0,07 + 0,16 + 0,11 + 0,19 + 0,17 + 0,26 + 0,27 + 0,19 + 0,21 + 0,19 + 0,29 + 0,02 = 2,3$$

Iterasi 2:

Berdasarkan jumlah anggota medoids 1,2,3,4 maka, pusat medoids baru adalah sebagai berikut:

$$Medoid\ 1\ (D_1) = (0,62 ; 0,04 ; 0,01 ; 0,01) y1$$

$$Medoid\ 2\ (D_2) = (0,50 ; 0,02 ; 0,02 ; 0,01) y2$$

$$Medoid\ 3\ (D_3) = (0,51 ; 0,02 ; 0,02 ; 0,01) y3$$

$$Medoid\ 4\ (D_4) = (0,40 ; 0,03 ; 0,02 ; 0,01) y4$$

Lakukan berulang seperti langkah iterasi 1. Sehingga didapatkan hasil seperti yang dapat dilihat pada tabel 7.

Tabel 7. Euclidian Distance Iterasi 2

No	A	B	C	D	dc ₁	dc ₂	dc ₃	dc ₄	c ₁	c ₂	c ₃	c ₄
1	0,62	0,04	0,01	0,01	0	-0,13	-0,12	-0,22	Yes			
2	0,50	0,02	0,02	0,01	0,13	0	0,01	-0,09		Yes		
3	0,51	0,02	0,02	0,01	0,12	-0,01	0	-0,1			Yes	
4	0,40	0,03	0,02	0,01	0,22	0,09	0,1	0,26			Yes	
5	0,48	0,01	0,01	0,02	0,16	0,03	0,04	-0,06		Yes		
6	0,38	0,01	0,01	0,02	0	0,06	-0,08	-0,06	Yes			
7	0,40	0,04	0,03	0,01	-0,06	0	-0,14	-0,12		Yes		
8	0,27	0,03	0,03	0,01	0,08	0,14	0	0,02			Yes	
9	0,28	0,03	0,03	0,02	0,06	0,12	-0,02	0				Yes
10	0,42	0,01	0,01	0,02	-0,04	0,02	-0,12	-0,1		Yes		
11	0,39	0,02	0,02	0,02	0,03	-0,05	-0,03	-0,12	Yes			
12	0,31	0,02	0,02	0,02	-0,03	-0,11	-0,09	-0,18	Yes			
13	0,33	0,02	0,03	0,01	0,11	0,03	0,05	-0,04		Yes		
14	0,24	0,02	0,03	0,01	0,09	0,01	0,03	-0,06		Yes		
15	0,23	0,02	0,03	0,01	-0,01	-0,09	-0,07	-0,16	Yes			
16	0,30	0,03	0,02	0,01	0	-0,02	0	-0,1			Yes	
17	0,29	0,01	0,02	0,02	0,02	0	0,02	-0,08		Yes		
18	0,29	0,02	0,03	0,02	0	-0,02	0	-0,1		Yes		
19	0,20	0,01	0,03	0,02	0,1	0,08	0,1	0				Yes
20	0,46	0,04	0,02	0,01	-0,17	-0,19	-0,17	0				Yes

$$Total\ Cost\ baru = -0,12 + 0 + 0 + 0,1 + 0,03 + 0 + -0,6 + 0,08 + 0 + -0,01 + -0,12 + -0,11 + -0,04 + -0,06 + -0,16 + 0 + 0,02 + 0 + 0,1 + -0,17 = -1,06$$

$$Total\ Cost\ Lama = 2,3$$

$$Cost = Cost\ Baru > Cost\ Lama$$

$$Hasil\ Iterasi\ Cost2 = -1,06 > 2,3$$

Berdasarkan hasil iterasi cost maka proses dihentikan di iterasi 2 karena nilai iterasi kedua lebih kecil dibandingkan iterasi pertama sehingga didapatkan cluster yang terbaik sebagai berikut berdasarkan tabel dibawah ini:

Tabel 8. Hasil Clustering

No	Nama	Tahun	Umur	Tanggung	Pendidikan	Pekerjaan	Clustering
1	Satunan	2020	0,62	0,04	0,01	0,01	C1
2	Rajadin	2020	0,50	0,02	0,02	0,01	C1
3	Sahbidun	2020	0,51	0,02	0,02	0,01	C1

No	Nama	Tahun	Umur	Tanggungan	Pendidikan	Pekerjaan	Clustering
4	Saparudin	2020	0,40	0,03	0,02	0,01	C3
5	Hadimah	2020	0,48	0,01	0,01	0,02	C2
6	Cungkek	2020	0,38	0,01	0,01	0,02	C2
7	Hajidan Bahri	2020	0,40	0,04	0,03	0,01	C1
8	Rustam Effendi	2020	0,27	0,03	0,03	0,01	C3
9	Atinah	2020	0,28	0,03	0,03	0,02	C4
10	Rusiah	2020	0,42	0,01	0,01	0,02	C2
11	Asrah	2020	0,39	0,02	0,02	0,02	C2
12	Erna Wati	2020	0,31	0,02	0,02	0,02	C2
13	Izhar Saidi	2020	0,33	0,02	0,03	0,01	C3
14	Rojikin	2020	0,24	0,02	0,03	0,01	C3
15	Anggara	2020	0,23	0,02	0,03	0,01	C3
16	Rejeki	2020	0,30	0,03	0,02	0,01	C3
17	Melidawati	2020	0,29	0,01	0,02	0,02	C2
18	Ruzaini	2020	0,29	0,02	0,03	0,02	C1
19	Karmida	2020	0,20	0,01	0,03	0,02	C4
20	Mawardi	2020	0,46	0,04	0,02	0,01	C4

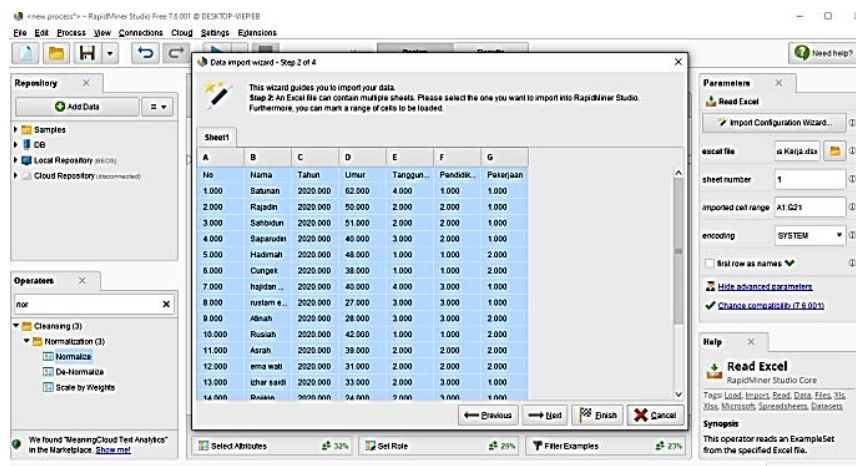
Berdasarkan dari tabel 8 dapat disimpulkan hasil clustering yang berbeda untuk data masyarakat miskin sebanyak 3 dari 20 sampel. Clustering 1 tidak mengalami perubahan data dan hasil clustering data klasifikasi penerima bantuan masyarakat miskin sehingga diperoleh data yang paling baik dari data masyarakat miskin sebelumnya.

3.4 Tampilan Sistem

Implementasi pengujian algoritma K-Medoids untuk pengelompokan penerima bantuan masyarakat miskin, yaitu berdasarkan spesifikasi cluster yang digunakan untuk melakukan pengujian menggunakan tools Rapidminer. Dimana kebutuhan sistem terbagi atas dua yaitu Perangkat Keras (Hardware) dan Perangkat Lunak (Software).

1. Operator RapidMiner

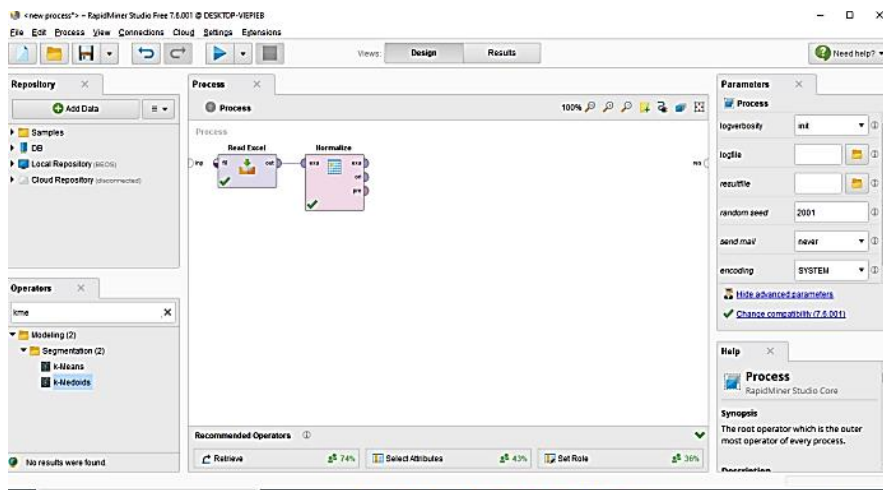
Jenis data yang diuji adalah data berupa *Microsoft Excel*, Yang file datanya disimpan dalam *FileExtension.xlsx* maka menu yang dipilih adalah menu operator dan pilih submenu *Data Access* → *Read Excel* yang dapat dilihat pada gambar 2.



Gambar 2. Tampilan Operator RapidMiner

2. Normalisasi Metode Desimal Scalling

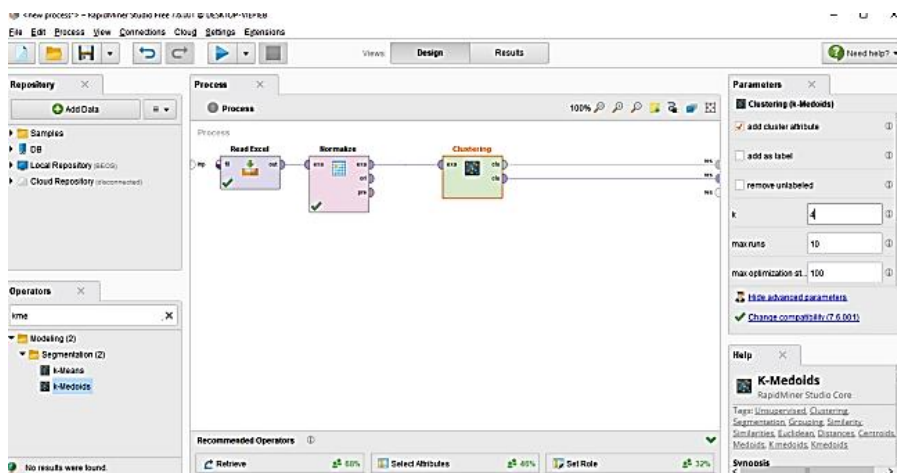
Normalization merupakan metode normalisasi dengan menggerakkan nilai desimal dari data ke arah yang diinginkan, clustering merupakan suatu metode pengelompokan record 3.



Gambar 3. Tampilan Normalisasi Metode Decimal Scalling

3. Klasifikasi

Metode algoritma *K-Medoids* di *double* klik atau di klik dan di *drag* ke halaman *main process* seperti pada tampilan gambar 4.



Gambar 4. Tampilan Klasifikasi Metode K-Medoids Pada RapidMiner

4. Pengelompokan Metode K-Medoids

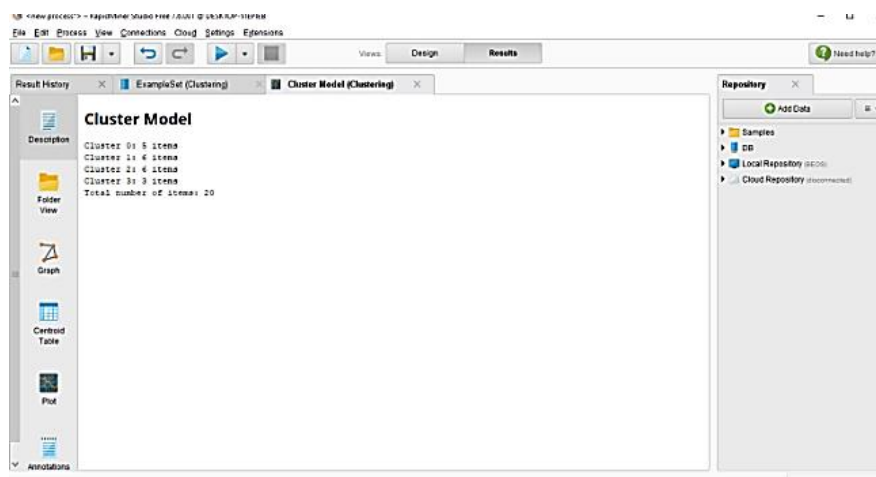
Hasil yang diperoleh dari pengujian bentuk *description rule* decisiontree algoritma *K-Medoids* seperti gambar 5.

Row No.	id	cluster	Umur	Tanggungan	Pendidikan	Pekerjaan
1	1	cluster_0	2.316	1.716	-1.503	-0.882
2	2	cluster_0	1.227	-0.245	-0.290	-0.882
3	3	cluster_0	1.318	-0.245	-0.290	-0.882
4	4	cluster_2	0.318	0.736	-0.290	-0.882
5	5	cluster_1	1.045	-1.225	-1.503	1.078
6	6	cluster_1	0.136	-1.225	-1.503	1.078
7	7	cluster_0	0.318	1.716	1.042	-0.882
8	8	cluster_2	-0.884	0.736	1.042	-0.882
9	9	cluster_3	-0.773	0.736	1.042	1.078
10	10	cluster_1	0.560	-1.225	-1.503	1.078
11	11	cluster_1	0.227	-0.245	-0.290	1.078
12	12	cluster_1	-0.500	-0.245	-0.290	1.078
13	13	cluster_2	-0.318	-0.245	1.042	-0.882
14	14	cluster_2	-1.135	-0.245	1.042	-0.882
15	15	cluster_2	-1.227	-0.245	1.042	-0.882
16	16	cluster_2	-0.591	0.736	-0.290	-0.882
17	17	cluster_1	-0.882	-1.225	-0.290	1.078

Gambar 5. Tampilan Pengelompokan Metode K-Medoids

5. Cluster Model

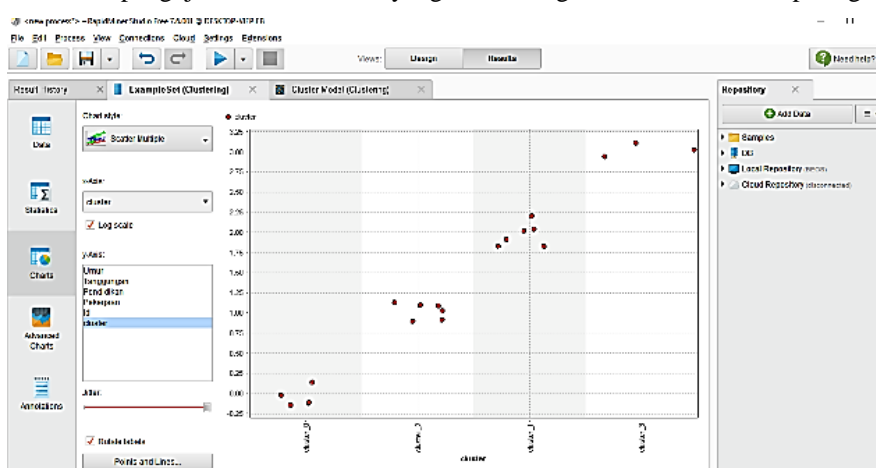
Hasil yang diperoleh dari pengujian bentuk *cluster* algoritma *K-Medoids* seperti gambar 6 di bawah ini:



Gambar 6. Tampilan Cluster Model

6. Hasil Keputusan

Hasil yang diperoleh dari pengujian bentuk *cluster* yang terbaik algoritma *K-Medoids* seperti gambar 7 di bawah ini:



Gambar 7. Tampilan Cluster Rapidminer

4. KESIMPULAN

Kesimpulan dari skripsi ini adalah bahwa data masyarakat miskin dinormalisasi dengan mengubah nilai sebelumnya menjadi nilai yang lebih kecil, sehingga dapat diproses lebih lanjut untuk mendapatkan informasi yang diinginkan. Selain itu, penerapan algoritma dalam proses pengelompokan data masyarakat miskin juga telah dilakukan, memungkinkan pembentukan kelompok yang sesuai. Terakhir, tahapan pengujian data menggunakan alat data mining telah dilakukan dalam serangkaian iterasi pada proses pengelompokan data masyarakat miskin, menghasilkan data pengelompokan yang terbaik.

REFERENCES

- [1] L. G. R. Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021.
- [2] A. Z. Wahidah, A. Widayani, S. I. Wardani, I. Rachmawati, and N. Latifah, "Prosedur Penyaluran Bantuan Sosial di Era Pandemi Covid-19 pada Dinas Sosial Kabupaten Blitar," *Competence J. Manag. Stud.*, vol. 16, no. 1, pp. 51–63, 2022.
- [3] N. S. Sinatrya and L. K. Wardhani, "Analysis of K-Means and K-Medoids's Performance Using Big Data Technology," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, IEEE, 2018, pp. 1–5.
- [4] L. M. Ginting, M. M. T. Sigiro, E. D. Manurung, and J. J. P. Sinurat, "Perbandingan Metode Algoritma Support Vector Regression dan Multiple Linear Regression Untuk Memprediksi Stok Obat," *J. Appl. Technol. Informatics Indones.*, vol. 1, no. 2, pp. 29–34, 2021.
- [5] G. A. B. Suryanegara and M. D. Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021.
- [6] V. Herlinda, D. Darwis, and D. Dartono, "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means," *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 2, pp. 94–99, 2021.
- [7] D. N. P. Sari and Y. L. Sukestiyarno, "Analisis cluster dengan metode K-Means pada persebaran kasus COVID-19 berdasarkan

- Provinsi di Indonesia,” in *PRISMA, Prosiding Seminar Nasional Matematika*, 2021, pp. 602–610.
- [8] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sist. J. Sist. Inf.*, vol. 7, no. 3, pp. 238–249, 2018.
- [9] J. Eska, “Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4. 5,” 2018.
- [10] R. Ordila, R. Wahyuni, Y. Irawan, and M. Y. Sari, “Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik Pt. Inecda),” *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020.
- [11] A. Mukhlisin and P. T. Prasetyaningrum, “ANALISIS DAN PERANCANGAN DECISION SUPPORT SYSTEM PENYALURAN BANTUAN PENYANDANG MASALAH KESEJAHTERAAN SOSIAL (PMKS) MENGGUNAKAN METODE WEIGHTED PRODUCT (WP),” *ANTIVIRUS J. Ilm. Tek. Inform.*, vol. 14, no. 1, pp. 1–14, 2020.
- [12] E. Y. Yunus, “Peran Tenaga Kesejahteraan Sosial Kecamatan (TKSK) dalam Mendampingi Masyarakat Penyandang Masalah Kesejahteraan Sosial (PMKS) di Kecamatan Kanigaran Kota Probolinggo,” *Publicio J. Ilm. Polit. Kebijak. dan Sos.*, vol. 3, no. 1, pp. 62–70, 2021.
- [13] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner,” *JBASE-Journal Bus. Audit Inf. Syst.*, vol. 4, no. 1, 2021.
- [14] Z. Nabila, A. R. Isnain, P. Permata, and Z. Abidin, “Analisis data mining untuk clustering kasus covid-19 di Provinsi Lampung dengan algoritma k-means,” *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021.
- [15] R. B. B. Sumantri and F. Mahardika, “Perancangan Basis Data Sistem Informasi Penjualan Barang (Studi Kasus: Minimarket ‘ABC’ Sidareja),” *J. Teknol. dan Bisnis*, vol. 2, no. 1, pp. 61–68, 2020.
- [16] H. H. Muflihini, H. Dhika, and S. Handayani, “Perancangan Sistem Informasi Inventory Pada Toko Rosadah,” *Bianglala Inform.*, vol. 8, no. 2, pp. 91–99, 2020.
- [17] M. R. Kusnaldi, T. Gulo, and S. Aripin, “Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal,” *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 330–338, 2022.
- [18] A. A. D. Sulistyawati and M. Sadikin, “Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan,” *Sist. J. Sist. Inf.*, vol. 10, no. 3, pp. 516–526, 2021.
- [19] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R. H. Zer, and D. Hartama, “Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia,” *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020.
- [20] A. Supriyadi, A. Triayudi, and I. D. Sholihati, “Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 2, pp. 229–240, 2021.