



Perbandingan Performa Cluster Model Algoritma K-Means Dalam Mengelompokkan Penerima Bantuan Program Keluarga Harapan

Warisa*, Nurahman

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Darwan Ali

Naskah Diterima :07 Februari 2023; Diterima Publikasi :17 Mei 2023

DOI: 10.21456/vol13iss1pp20-28

Abstract

Poverty has so far played a role as a problem faced by residents of the Mentawa Baru sub-district, Ketapang. The inability of this community is related to the need to meet education and health needs in social welfare. In assisting the grouping of beneficiary data is carried out using the K-Means algorithm. Apart from that, to increase performance, those who have gone through the first grouping process are then continued using feature selection in the decision tree tool. The algorithm used aims to classify PKH beneficiary data to help the government find out about the handling of the aid program in Mentawa Baru Ketapang sub-district. As for the results obtained from this study, namely, the accuracy of the initial clustering obtained a DBI value of -0.994 at K=8 while the second clustering value that had gone through feature selection with K=3 obtained a DBI value of -0.865. It is known from the performance testing of the comparison of the two clustering that the best performance value is found in the second cluster after going through feature selection.

Keywords : K-Means; Social Assistance; Decision Trees

Abstrak

Kemiskinan hingga saat ini berperan sebagai masalah yang dihadapi warga kecamatan mentawa baru ketapang. Ketidakmampuan masyarakat ini berkaitan dengan kebutuhan untuk memenuhi kebutuhan pendidikan dan kesehatan dalam kesejahteraan masyarakat sosial. Dalam membantu pengelompokan data penerima bantuan dilakukan menggunakan algoritma K-Means. Selain itu untuk menaikkan performa yang sudah melalui proses pengelompokan pertama, kemudian dilanjutkan menggunakan seleksi fitur pada tool decision tree. Algoritma yang digunakan bertujuan untuk mengelompokan data penerima bantuan PKH untuk membantu pemerintah mengetahui tentang penanganan program bantuan yang ada di kecamatan mentawa baru ketapang. Adapun hasil yang diperoleh dari penelitian ini yaitu, akurasi dari pengklasteran awal diperoleh nilai DBI sebesar -0.994 pada K=8 sedangkan nilai pengklasteran kedua yang telah melalui seleksi fitur dengan K=3 yang diperoleh nilai DBI sebesar -0.865. Diketahui dari pengujian performa dari perbandingan kedua pengklasteran tersebut bahwa nilai performa terbaik terdapat di pengklasteran kedua setelah melalui seleksi fitur.

Kata Kunci : K-Means; Bantuan Sosial; Decision Tree

1. Pendahuluan

Kecamatan Mentawa Baru Ketapang adalah salah satu kecamatan yang berada di Kab.Kotawaringin Timur, Kalimantan Tengah. Kecamatan ini berada di kota Sampit yang merupakan ibukota dari Kabupaten Kotawaringin Timur. Kecamatan ini merupakan salah satu dari 17 kecamatan yang ada di Kabupaten Kotawaringin Timur. Kecamatan ini memiliki 11 kelurahan/desa dengan luas wilayah sebesar 726,00 Km². Luasnya wilayah tersebut terdapat masyarakat yang penduduknya tergolong banyak. Kemiskinan masih banyak melanda di kecamatan ini, dikarenakan masih terdapat masyarakat yang tidak mampu dalam ekonomi. Minimnya ekonomi atau kemiskinan

merupakan masalah ketidakmampuan yang dihadapi masyarakat kecamatan mentawa baru ketapang. Ketidakmampuan masyarakat ini berkaitan dengan kebutuhan untuk memenuhi kebutuhan pendidikan dan kesehatan dalam kesejahteraan masyarakat sosial. Untuk kesejahteraan masyarakat pemerintah menyalurkan bantuan berupa program harapan keluarga (PKH) (Ikhwan dan Aslami, 2020)

Program Keluarga Harapan (PKH) merupakan program pemerintah berupa bantuan sosial seperti uang tunai sesuai ketentuan yang sudah dibuat pemerintah untuk keluarga tidak mampu yang telah ditetapkan sebagai keluarga penerima bantuan PKH (Aprilawati brBarus dan Tarigan, 2019). PKH adalah program pemerintah untuk menanggulangi

*) Penulis korespondensi: wewarisa020@gmail.com

kemiskinan yang dibuat oleh pemerintah, adanya sistem pemberian bantuan secara tunai kepada keluarga kurang mampu yang tercantum pada Data (DTKS) yang sudah memenuhi syarat. Tujuan PKH yaitu untuk mengurangi ketidakmampuan masyarakat dan meningkatkan kualitas SDM terutama kelompok masyarakat yang tidak mampu. PKH ini menargetkan berupa bantuan ke beberapa golongan seperti, anak usia dini, anak sekolah, ibu hamil, penyandang disabilitas, dan lansia (Putri *et al.*, 2022). Oleh karena itu, dalam meningkatkan kesejahteraan masyarakat social serta Penanganan rakyat miskin di kecamatan mentawa baru ketapang maka pemerintah menyalurkan program keluarga harapan (PKH) sebagai bantuan perlindungan masyarakat sosial (Haliim, 2016).

Klasifikasi Dana Bantuan Pada Program Keluarga Harapan (PKH) Menggunakan Metode K-Means, penelitian ini melakukan klasifikasi berdasarkan Penelitian yang menggunakan klasifikasi Rumah Tangga mendekati Miskin, Rumah Tangga yang Miskin, dan Rumah Tangga paling Miskin dapat digunakan untuk mengukur tingkat kemiskinan di suatu wilayah atau negara. Jumlah cluster yang digunakan dalam penelitian tersebut akan bergantung pada jumlah sampel yang diambil dan jumlah kategori yang digunakan dalam klasifikasi kemiskinan. Cluster yang digunakan dalam penelitian ini akan membantu dalam menganalisis perbedaan dan kesamaan dalam tingkat kemiskinan di wilayah yang diteliti. diketahui bahwa sistem ini mampu mengelompokkan keluarga menjadi 3 cluster : Klaster 1 (C1) Rumah Tangga Hampir Miskin (RTHM) mewakili keluarga yang hampir miskin, Klaster 2 (C2) Rumah Tangga Miskin (RTM) mewakili keluarga miskin, dan Cluster 3 (C3) Rumah Tangga Sangat Miskin mewakili keluarga yang sangat miskin. Penting untuk dicatat bahwa klaster didasarkan pada kriteria penelitian dan belum tentu mencerminkan garis kemiskinan resmi di negara tersebut (Said *et al.*, 2021).

Monitoring Program Keluarga Harapan Berbasis Mobile GIS Menggunakan K-Means Clustering (Muhariya *et al.*, 2021), Dalam penelitian ini pengelompokan dilakukan berdasarkan beberapa variabel seperti anak SD, SMP, SMA, ibu hamil, usia dini, lanjut usia dan penyandang disabilitas. Penelitian menggunakan sampel sebanyak 21 keluarga untuk menguji validitas sistem dan ditemukan bahwa sistem mampu mengelompokkan keluarga menjadi 7 cluster dengan tingkat presisi 90,4%. Kesimpulannya adalah metode K-Means Clustering yang digunakan dalam penelitian memiliki akurasi yang tinggi untuk mengelompokkan keluarga miskin.

Masalah pada program bantuan di kecamatan mentawa baru ketapang seperti masih terdapat kuota bantuan sosial lebih sedikit daripada orang yang berhak menerima. Selain itu pemilihan penerima

bantuan sosial berdasarkan kedekatan, data penduduk yang banyak, penerimaan bantuan yang double, serta penggunaan data lama. Karena masalah inilah pemerintah sulit mengelompokkan data yang terdapat di kecamatan tersebut. Untuk menangani masalah tersebut maka penelitian ini melakukan klasterisasi menggunakan algoritma k-means (Fard *et al.*, 2020).

Pada penelitian (Nurahman *et al.*, 2022) menjelaskan klasterisasi pada fasilitas di seluruh sekolah kabupaten seruyan. Hasil penelitian pada Cluster0 terbentuk 178 sekolah, Cluster1 terdiri 3 sekolah, dan Cluster2 terdiri 43 sekolah. Kemudian hasil peringkat klaster diperoleh yaitu mulai dari Cluster1, Cluster0, dan terakhir Cluster2. Selanjutnya dalam pengujian performan algoritma K-Means 3Cluster memperoleh nilai DBI -0,695. Beberapa penelitian terdahulu di atas mempunyai perbedaan dengan penelitian yang dilakukan peneliti sekarang. Adapun pembaharuan pada penelitian ini yaitu menggunakan seleksi fitur yang bertipe gain ratio.

Tujuan dalam penelitian ini untuk mengelompokkan data penerima bantuan PKH yang menggunakan algoritma K-means clustering pada data tersebut di kecamatan mentawa baru ketapang. Penelitian ini juga bertujuan untuk melihat akurasi dari pengklasteran yang telah dilakukan. Pengklasteran dilakukan dengan 2 tahapan yaitu pengklasteran pertama menggunakan dataset asli dan pengklasteran kedua menggunakan data yang sudah dilakukan penyeleksian fitur yang bertujuan melihat performa dari masing-masing pengklasteran.

2. Kerangka Teori

2.1. Davies Boildin Index

Pengujian dalam performa algoritma K-Means yaitu menggunakan Davies Boildin Index (DBI) dalam proses data mining. Pencarian nilai DBI dilakukan setelah proses klasterisasi menggunakan K-Means berdasarkan nilai *cluster* yang sudah ditentukan. Klasterisasi yang terbentuk kemudian di uji menggunakan dengan pengukuran DBI. Pengukuran DBI menggunakan cara pemaksimalan jarak *inter Cluster* juga meminimalkan jarak *intra cluster*. dinyatakan bahwa semakin kecil nilai DBI maka hasil menunjukkan skema klaster yang paling baik (Wijaya *et al.*, 2021).

Adapun rumus dari pencarian nilai DBI yaitu sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \{R_i\}$$

dan

$$R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}, R_{ij} = \frac{S_1 + S_j}{d_{ij}} \quad (1)$$

Dengan

$$S_1 = \left[\frac{1}{n} \sum_{x \in n_i} d^2(x, v_i) \right]^{\frac{1}{2}}$$

2.2. Algoritma K-means

K-Means merupakan metode yang mengelompokkan data nonhierarki yang mempartisi dataset kedalam beberapa kelompok. metode tersebut akan memberi batasan pemisahan data dalam suatu Klaster dimana data yang memiliki karakteristik sama dan kemudian dimasukkan kedalam suatu klaster yang sama sedangkan dataset dengan karakteristik berbeda akan digolongkan kedalam kelompok yang lain (Dinata *et al.*, 2020).

2.3. Seleksi Fitur Gain Ratio

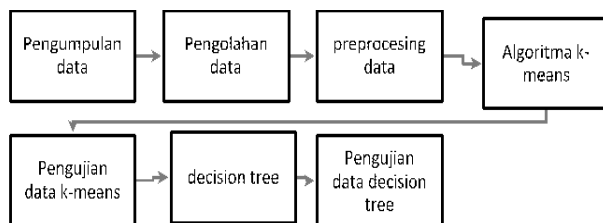
Gain ratio merupakan pembobotan yang dapat memperbaiki dataset yang tidak stabil, cocok untuk data yang bertipe numerik dua kelas, sederhana sehingga komputasi menjadi lebih cepat. Dalam perhitungan gain ratio memerlukan pemisahan Informasi-informasi digunakan untuk persamaan (Kurniabudi *et al.*, 2021).

2.4. Program Keluarga Harapan (PKH)

PKH yaitu suatu program perlindungan sosial yang memberikan bantuan berupa uang tunai kepada keluarga yang tidak mampu dengan syarat tertentu. Tujuan program harapan keluarga adalah mengurangi angka kemiskinan, meningkatkan SDM, maupun merubah perilaku yang tidak mendukung kesejahteraan dalam golongan yang tidak mampu. anggota PKH wajib memenuhi persyaratan yang terkait sebagai upaya peningkatan SDM (Rahmawati dan Kisworo, 2017).

3. Metode

Metodologi Penelitian adalah tahap-tahap penelitian yang dilakukan pada penelitian ini meliputi Tahap penelitian yang telah dilakukan untuk menganalisa masalah, mengumpulkan data, pengolahan data, algoritma k-means, pengujian data (Safitri *et al.*, 2019). Adapun tahap-tahap penelitian terdapat pada Gambar :



Gambar 1. Tahap penelitian

3.1 Pengumpulan Data

Proses pengumpulan Data primer adalah data yang diperoleh langsung dari sumber yang bersangkutan melalui metode seperti wawancara, observasi, atau pengukuran. Data ini tidak pernah digunakan atau diolah sebelumnya dan digunakan untuk tujuan penelitian yang spesifik. data ini merupakan bantuan program harapan keluarga(PKH) di kecamatan mentawa baru ketapang. data masyarakat yang mendapatkan bantuan sebanyak 845 orang.

3.2 Pengolahan Data

Pengolahan data dilakukan menggunakan algoritma k-means untuk pengelompokan/cluster data penerima PKH.

3.3 Preprocessing Data

Pada langkah-langkah dataset tersebut yang berjumlah 845 data penduduk kemudian dilakukan proses Preprocessing Data. Tahapan ini dilakukan preprocessing data menjadi dua tahapan yaitu tahapan Clearing dan tahapan penyeleksian atribut.

3.3.1 Clearing

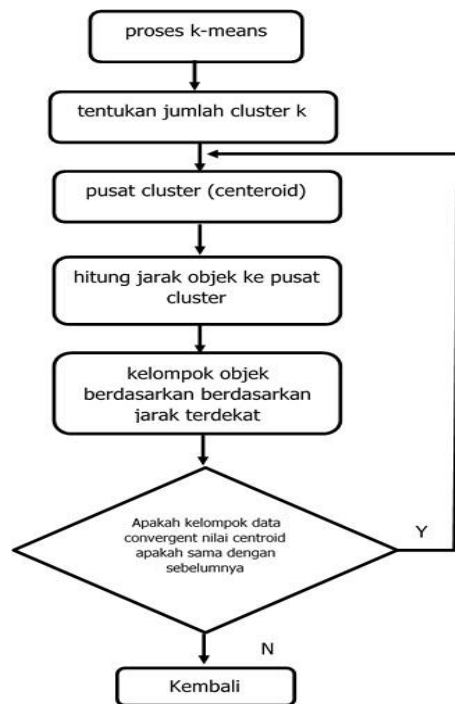
Pengolahan pertama yang dilakukan yaitu tahapan penghapusan data. Dalam penghapusan data ini dilakukan penghapusan kolom atau atribut yang tidak digunakan, atribut yang tidak digunakan ialah sebagai berikut nomor peserta, nama pengurus, tempat lahir, tanggal lahir, alamat, nama provinsi, nama kabupaten, nama kecamatan, nama kelurahan, nama pendamping, dan nama penanggunguhan (Nurahman dan Prihandoko, 2019).

3.3.2 Penyeleksian atribut

Proses penyeleksian atribut dilakukan menggunakan metode decision tree berbasis gain ratio. Adapun dalam proses ini menggunakan rapidminer dan menghasilkan atribut terbaik yang disarankan dari penyeleksian fitur tersebut (Putra dan Anggrawan, 2021).

3.4 Algoritma K-Means

Algoritma k-means yaitu metode yang bertujuan mengelompokkan objek berdasarkan kemiripan karakter dengan jumlah cluster yang telah ditetapkan lebih dulu. Algoritma k-means adalah suatu metode non hirarki (Riza *et al.*, 2021).



Gambar 2. Tahap k-means

Pada Gambar 2 menjelaskan tahapan algoritma K-means yaitu sebagai berikut :

1. Menetapkan nilai k menjadi jumlah cluster yang akan dibentuk
2. Menentukan k centroid pada titik pusat cluster pertama secara acak.
3. Menghitung jarak data ke tiap-tiap centroid menerapkan rumus Euclidean (Nasyuha *et al.*, 2022).

$$d(x_i - \mu_j) = \sqrt{x_i} = \mu_j^2 \quad (2)$$

Keterangan ;

- d : objek x ke-i
- μ_j : data j ke-i
- x_i : banyaknya objek

4. Menggolongkan tiap data berdasarkan jarak yang paling dekat antara data dan centroidnya.
5. Menetapkan letak centroid terbaru

$$c_k = \left(\frac{1}{n_k}\right) \sum d_i \quad (3)$$

Keterangan ;

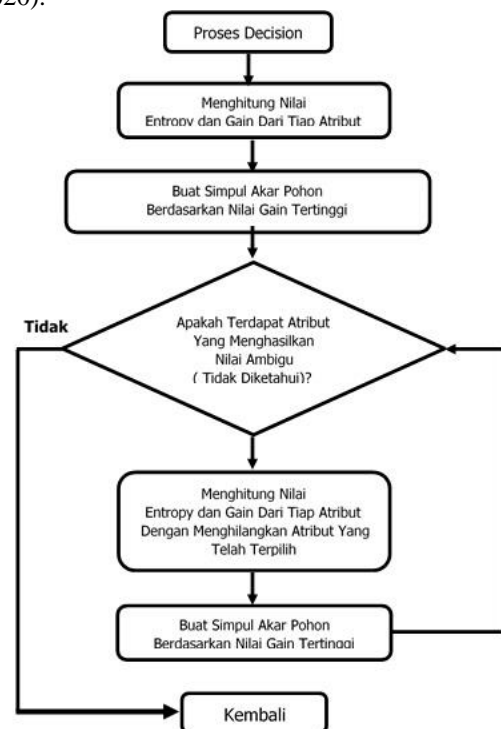
- n_k = jumlah data dalam cluster k
 - d_i = jumlah dari nilai jarak yang masuk dalam masing-masing cluster
6. Kembali ke tahap 3 jika letak centroid terbaru dengan centroid yang lama berbeda.

3.5 Pengujian Data K-Means

Pengujian data akurasi pada algoritma K-Means dilakukan pencarian nilai DBI (Davies Bouldin Index). Dalam Pencarian nilai DBI dilakukan setelah cluster dengan Algoritma K-Means, menurut nilai K yang sudah ditentukan. Cluster yang sudah ditemukan di uji menggunakan pengukuran pada DBI. Pengujian DBI menggunakan prinsip maksimal jarak inter Cluster dan minimal jarak intra cluster. Nilai DBI terbaik adalah nilai yang paling rendah mendekati -0, karena semakin rendah nilai DBI, semakin baik kualitas clustering pada data mining. DBI merupakan singkatan dari Dunn Index, yang mengukur rasio antara jarak antar-cluster dan jarak intra-cluster pada data mining. Semakin tinggi rasio ini, semakin buruk kualitas clustering, dan semakin sulit untuk membedakan antara cluster-cluster yang berbeda. Sebaliknya, semakin rendah rasio ini, semakin baik kualitas clustering, dan semakin mudah untuk membedakan antara cluster-cluster yang berbeda. Oleh karena itu, nilai DBI yang paling rendah menunjukkan bahwa clustering pada data mining memiliki kualitas yang baik dan cluster-cluster yang terbentuk berbeda secara signifikan satu sama lain.

3.6 Decision Tree

Decision Tree merupakan processing data yang memprediksi data, dengan cara membentuk klasifikasi model pohon keputusan yang menggunakan aturan dalam membuat keputusan struktur pohon yang kemungkinan hasil, sumber daya, resiko, konsekuensi dan utilitas (Sa'Adah *et al.*, 2020).



Gambar 3. Tahap Decision tree

Pada Gambar 3 menunjukkan Rumus decision tree sebagai berikut :

1. Rumus Entropy (Swe, 2019)

$$C_i = \sum_{i=1}^n - p_i * \log_2 p_i \tag{4}$$

Keterangan:

S= Himpunan pada Kasus

n = jumlah S

pi = proporsi pada Si untuk S

2. Rumus Gain

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \tag{5}$$

Keterangan:

S = Himpunan Kasus

A = Atribut

n = jumlah partisi atribut A

|Si|= jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S P

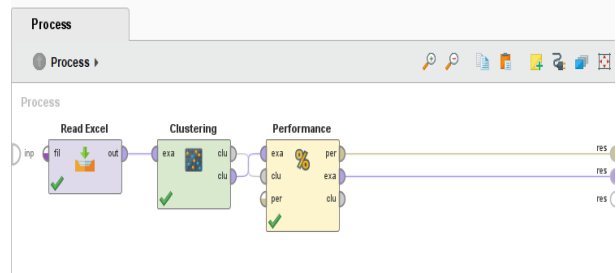
4. Hasil dan Pembahasan

Pada penelitian ini menggunakan data berasal dari kecamatan mentawa baru ketapang yang disediakan dinas sosia tahun 2022. Data yang digunakan ialah data bantuan program harapan keluarga (PKH). Jumlah data yang diperoleh dalam penelitian yaitu 845 data penduduk. Data yang diperoleh diproses menggunakan software atau tool pada rapidminer.

Tabel 1. Jangkauan Nilai Dataset

| Atribut | Nilai |
|--------------------|-------|
| Jumlah anak SD | 0-2 |
| Jumlah anak SMP | 0-2 |
| Jumlah anak SMA | 0-2 |
| Jumlah Bumil | 0 |
| Jumlah Usia Dini | 0-2 |
| Jumlah Lansia | 0-2 |
| Jumlah Disabilitas | 0-1 |
| Jumlah ART | 0-7 |

Tabel 1 menunjukkan jangkauan setiap nilai dari setiap atribut dalam bentuk numerik. Nilai numerik tersebut akan dianalisis sesuai dengan langkah-langkah yang terdapat dalam metode penelitian. Informasi tersebut penting untuk mengidentifikasi karakteristik dari dataset yang digunakan dalam penelitian.



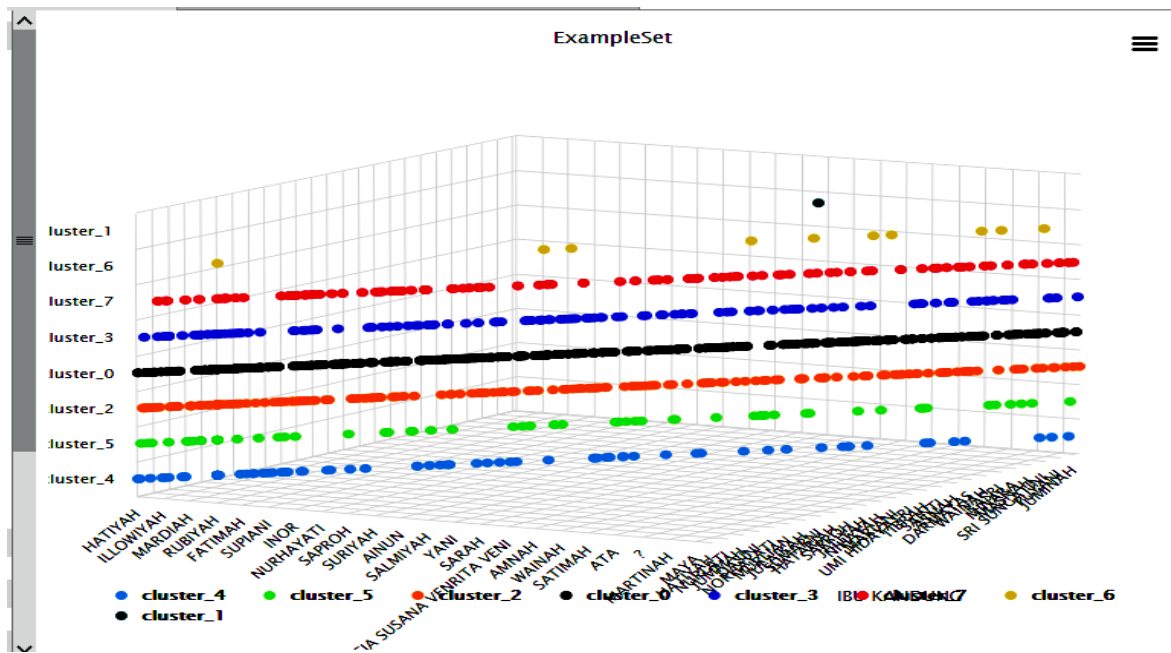
Gambar 4. Proses clustering

Pada Gambar 4 Proses clustering dalam penelitian ini menggunakan perangkat lunak RapidMiner untuk mengkluster dataset dan memrosesnya melalui operator Read Excel. Kemudian, dilakukan Operator Clustering pada data yang telah dimasukkan sebelumnya. Evaluasi kinerja dilakukan untuk mendapatkan hasil dari data clustering Bantuan Program Keluarga Harapan (PKH) di Kecamatan Mentawa Baru Ketapang.

| PerformanceVector | Cluster Model |
|---|--------------------------|
| PerformanceVector: | Cluster 0: 258 items |
| Avg. within centroid distance: -0.725 | Cluster 1: 49 items |
| Avg. within centroid distance_cluster_0: -0.640 | Cluster 2: 60 items |
| Avg. within centroid distance_cluster_1: -0.000 | Cluster 3: 52 items |
| Avg. within centroid distance_cluster_2: -0.613 | Cluster 4: 223 items |
| Avg. within centroid distance_cluster_3: -0.810 | Cluster 5: 14 items |
| Avg. within centroid distance_cluster_4: -0.377 | Cluster 6: 104 items |
| Avg. within centroid distance_cluster_5: -1.858 | Cluster 7: 21 items |
| Avg. within centroid distance_cluster_6: -0.000 | Cluster 8: 63 items |
| Avg. within centroid distance_cluster_7: -0.737 | Total number of items: 8 |
| Davies Bouldin: -0.994 | |

Gambar 5. Description Performancevector

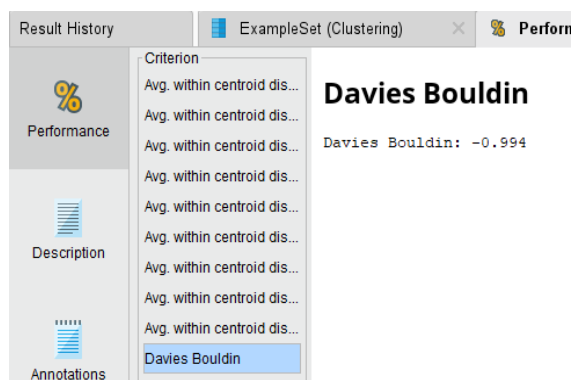
Gambar 5 menunjukkan daftar anggota cluster. Cluster_0 : -0.640, Cluster_1 : -0.000, Cluster_2 : -0.613, Cluster_3 : -0.810, Cluster_4 : -0.377, Cluster_5 : -1.858, Cluster_6 : -0.000, Cluster_7: -0.737. cluster model penerima bantuan dari masing-masing kelompok. Cluster 0 terdiri dari 258 items artinya pada cluster ini terdapat jumlah penerima bantuan 258 orang. Cluster 1 terdiri dari 49 items artinya pada cluster ini terdapat jumlah penerima bantuan 49 orang. Cluster 2 terdiri dari 60 items artinya pada cluster ini terdapat jumlah penerima bantuan 60 orang. Selanjutnya Cluster 3 terdiri dari 52 items artinya pada cluster ini terdapat jumlah penerima bantuan 52 orang. Cluster 4 terdiri dari 223 items artinya pada cluster ini terdapat jumlah penerima bantuan 223 orang. Cluster 5 terdiri dari 14 items artinya pada cluster ini terdapat jumlah penerima bantuan 14 orang. Cluster 6 terdiri dari 104 items artinya pada cluster ini terdapat jumlah penerima bantuan 104 orang. Cluster 7 terdiri dari 21 items artinya pada cluster ini terdapat jumlah penerima bantuan 21 orang. Cluster 8 terdiri dari 63 items artinya pada cluster ini terdapat jumlah penerima bantuan 63 orang.



Gambar 6. Peringkat kluster

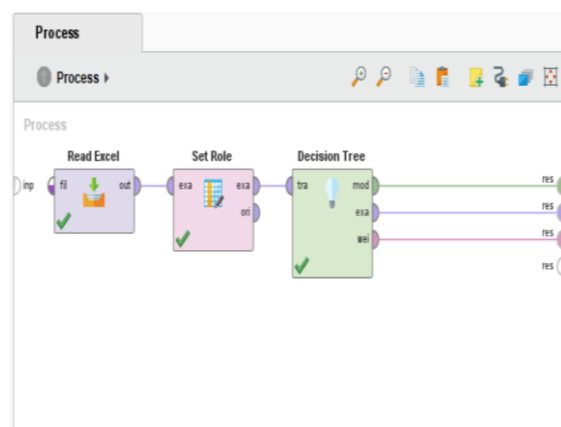
Gambar 6 menampilkan hasil visualisasi dari peringkat cluster yang dibentuk dari data yang telah dianalisis. Terdapat tujuh cluster yang terlihat dengan warna-warna yang berbeda, yaitu dark green untuk cluster 1, orange untuk cluster 2, blue untuk cluster 3, light blue untuk cluster 4, green untuk cluster 5, yellow untuk cluster 6, dan red untuk cluster 7. Pada Gambar 6, peringkat dari masing-masing cluster ditunjukkan dengan urutan warna yang terlihat pada visualisasi. Terdapat tujuh cluster yang dihasilkan dari data yang dianalisis, dan masing-masing cluster ditunjukkan dengan warna yang berbeda. Hasil visualisasi pada Gambar 6 menampilkan peringkat cluster yang terbentuk dari data yang telah dianalisis. Terdapat tujuh cluster yang terlihat dengan warna yang berbeda, dan peringkat dari masing-masing cluster ditunjukkan dengan urutan warna pada visualisasi tersebut.

kualitas dan keakuratan hasil clustering atau pengelompokan data. Semakin rendah nilai DBI, semakin baik performa clustering-nya. penggunaan algoritma dari suatu hasil yang dapat disediakan dalam berbagai macam cara. Penelitian ini menghasilkan performance dengan memberikan nilai DBI. Davies Bouldin nilai yang dihasilkan $-0,994$ dengan menggunakan $K=8$. Nilai DBI yang dihasilkan dapat digunakan untuk membandingkan kualitas clustering dengan menggunakan algoritma dan parameter yang berbeda. Dalam penggunaan K-means, pemilihan nilai K yang optimal juga sangat penting untuk menghasilkan clustering yang akurat dan optimal. Namun, pemilihan nilai K yang optimal bukanlah suatu hal yang mudah dan dapat memakan waktu yang cukup lama. Oleh karena itu, pemilihan nilai K yang tepat dan penyesuaian parameter yang optimal perlu dilakukan dengan hati-hati dan secara teliti untuk mendapatkan hasil clustering yang baik dan akurat.



Gambar 7. Davies Bouilding

Pada Gambar 7 Seperti yang telah saya jelaskan sebelumnya, Davies-Bouldin Index (DBI) adalah sebuah metrik yang digunakan untuk mengukur



Gambar 8. Desain proses decision tree

Pada Gambar 8 merupakan hasil rancangan fitur selection menggunakan decision tree, Gambar ini juga menampilkan Atribut yang disarankan dari fitur seleksi. Hasil atribut yang disarankan akan dilanjutkan menggunakan algoritma k-means kembali. K-means merupakan salah satu metode clustering yang sering digunakan dalam machine learning. Metode ini bertujuan untuk mengelompokkan data yang memiliki kemiripan atau kesamaan ke dalam beberapa kelompok atau cluster. Dalam penggunaannya, k-means membutuhkan beberapa parameter seperti jumlah cluster (k) yang diinginkan dan fitur-fitur yang digunakan sebagai input. Dengan menggunakan hasil rancangan fitur selection dari decision tree, kita dapat mengambil atribut-atribut yang disarankan untuk digunakan sebagai input dalam proses k-means. Hal ini diharapkan dapat meningkatkan akurasi dan performa dari model machine learning yang akan dihasilkan.

| attribute | weight |
|--------------------|--------|
| NAMA KELURAHAN | 0.006 |
| JUMLAH USIA DINI | 0.012 |
| JUMLAH ANAK SMA | 0.032 |
| TANGGAL LAHIR | 0.800 |
| JUMLAH ANAK SMP | 0.032 |
| JUMLAH LANSIA | 0.022 |
| JUMLAH ANAK SD | 0.029 |
| JUMLAH ART | 0.063 |
| JUMLAH DISABILITAS | 0.004 |

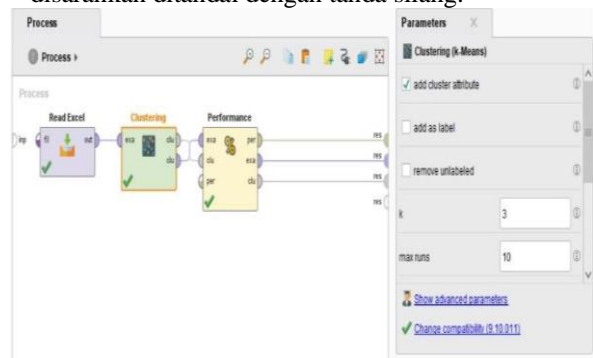
Gambar 9. Desain proses gain ratio

Gambar 9 menunjukkan tabel hasil dari proses seleksi fitur menggunakan metode gain ratio. Tabel tersebut menampilkan nilai terbaik dari setiap fitur yang diproses dalam seleksi fitur tersebut. Metode seleksi fitur dalam machine learning yang digunakan pada Gambar 9 adalah gain ratio. Metode ini membantu untuk menentukan fitur mana yang paling informatif atau memberikan kontribusi terbesar dalam membedakan kelas atau label pada dataset. Dengan menggunakan metode gain ratio, proses seleksi fitur pada Gambar 9 berhasil menghasilkan tabel yang menunjukkan nilai terbaik dari setiap fitur yang diproses. Tabel tersebut dapat membantu untuk memilih fitur-fitur yang paling penting dalam membedakan kelas atau label pada dataset yang digunakan.

Tabel 2. Sesudah dan sebelum seleksi fitur

| Atribut sebelum seleksi fitur | Atribut setelah seleksi fitur |
|-------------------------------|-------------------------------|
| Kelurahan | √ |
| Ibu kandung | x |
| Jumlah ART | √ |
| Jumlah anak SD | √ |
| Jumlah anak SMP | √ |
| Jumlah anak SMA | √ |
| Jumlah bumil | x |
| Jumlah usia dini | x |
| Jumlah lansia | √ |
| Jumlah disabilitas | x |

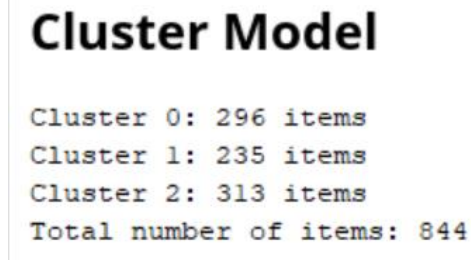
Proses seleksi fitur dilakukan dengan memilih fitur yang memiliki korelasi tertinggi, sehingga hanya fitur yang disarankan untuk penelitian yang dipilih. Tabel 2 menampilkan hasil seleksi fitur yang dilakukan, baik sebelum maupun setelah dilakukan seleksi. Pada tabel tersebut, atribut yang termasuk dalam seleksi fitur yang disarankan ditandai dengan tanda centang, sedangkan atribut yang tidak terseleksi atau tidak disarankan ditandai dengan tanda silang. Seleksi fitur dilakukan dengan memilih fitur yang memiliki korelasi tertinggi untuk penelitian. Tabel 2 memperlihatkan hasil seleksi fitur yang dilakukan, di mana atribut yang termasuk dalam seleksi fitur yang disarankan ditandai dengan tanda centang dan atribut yang tidak terseleksi atau tidak disarankan ditandai dengan tanda silang. Tabel tersebut membantu memudahkan pemilihan fitur yang disarankan. Dalam proses seleksi fitur, dipilih fitur yang memiliki korelasi tertinggi untuk penelitian sehingga hanya fitur yang disarankan yang dipilih. Tabel 2 menunjukkan hasil seleksi fitur sebelum dan setelah dilakukan seleksi. Pada tabel tersebut, atribut yang termasuk dalam seleksi fitur yang disarankan ditandai dengan tanda centang, sedangkan atribut yang tidak terseleksi atau tidak disarankan ditandai dengan tanda silang.



Gambar 10. Hasil clustering setelah fitur seleksi

Gambar 10 menampilkan hasil clustering setelah dilakukan seleksi fitur dan pengolahan data menggunakan algoritma K-Means dengan tujuan membandingkan performa DBI. Pada Gambar 10,

dapat dilihat rancangan algoritma K-Means yang digunakan serta hasil performa DBI sebesar -0,865, yang dihasilkan dengan menggunakan K=3. Seleksi fitur dan pengolahan data yang dilakukan sebelumnya mempengaruhi hasil clustering pada Gambar 10, yang ditampilkan dengan nilai performa DBI yang dihasilkan.



Gambar 11. Cluster Model

Gambar 11 Menampilkan cluster model k-means yang di proses setelah fitur seleksi. Adapun hasil yaitu penerima bantuan dari masing-masing kelompok terdiri dari 3 klaster antara lain Cluster 0 sebanyak 296 items yang artinya pada cluster ini terdapat jumlah penerima bantuan 296 orang. Kemudian Cluster 1 sebanyak 235 items yang artinya pada cluster ini terdapat jumlah penerima bantuan 235 orang, Dan terakhir Cluster 2 sebanyak 313 items yang artinya pada cluster ini terdapat jumlah penerima bantuan 313 orang.

Dari proses pengkalsteran yang telah dilakukan maka hasil yang di ambil adalah pengelompokan menggunakan k-means setelah proses seleksi fitur gain ratio yaitu K-3 dengan nilai DBI -0.865. Adapun hasil yang diperoleh dari pembahasan sebelumnya maka penelitian ini memberikan kemudahan dalam pemeliharaan penerima bantuan PKH berdasarkan masalah atau kasus yang di angkat pada penelitian ini.

5. Kesimpulan

Penelitian memiliki data yang berjumlah 845 record. Data memerlukan tahapan preprosesing untuk dijadikan sebagai data yang siap masuk pada tahapan pengolahan data. Preprosesing dilakukan dua tahap yaitu cleaning dan penyeleksian atribut. Selanjutnya dalam pengolahan data menggunakan algoritma k-means yang menghasilkan performa DBI awal sebelum dilakukan proses seleksi fitur gain ratio dengan nilai sebesar -0.994 dengan k=8 cluster. Kemudian performa Algoritma K-means setelah dilakukan seleksi fitur menggunakan gain ratio memiliki DBI senilai -0.865 dengan k=3 cluster. Pada penelitian yang dilakukan serta disimpulkan bahwa teknik seleksi fitur menggunakan metode decision tree berbasis gain ratio dapat lebih meningkatkan nilai performa DBI dibandingkan sebelum seleksi fitur.

Daftar Pustaka

- Aprilawati brBarus, R., Tarigan, P., 2019. Implementasi data mining untuk menentukan keluarga yang layak mendapat kartu PKH (program keluarga harapan) dengan metode k-means clustering. *Pelita Informatika: Informasi Dan Informatika* 7(3).
- Dinata, R. K., Novriando, H., Hasdyna, N., Retno, S., 2020. Reduksi atribut menggunakan information gain untuk optimasi cluster algoritma k-means. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)* 6(1). <https://doi.org/10.26418/jp.v6i1.37606>
- Haliim, W., 2016. Poverty reduction for extremely poor households of malang city by the implementation of program keluarga harapan. *Jurnal Bina Praja* 8(2). <https://doi.org/10.21787/jbp.08.2016.331-340>
- Ikhwan, A., Aslami, N., 2020. Implementasi data mining untuk manajemen bantuan sosial menggunakan algoritma k-means. *Jurnal Teknologi Informasi* 4(2), 208–217. <https://doi.org/10.36294/jurti.v4i2.2103>
- Kurniabudi, K., Harris, A., Mintaria, A. E., 2021. Komparasi information gain, gain ratio, cfs-bestfirst dan cfs-pso search terhadap performa deteksi anomali. *Jurnal Media Informatika Budidarma*, 5(1), 332. <https://doi.org/10.30865/mib.v5i1.2258>
- Fard, M. M., Thonet, T., Gaussier, E., 2020. Deep k-means: jointly clustering with k-means and learning representations. *Pattern Recognition Letters* 138, 185-192. <https://doi.org/10.1016/j.patrec.2020.07.028>
- Muhariya, A., Widada, B., Siswanti, S., 2021. Monitoring program keluarga harapan berbasis mobile GIS menggunakan k-means clustering. *Techno.Com* 20(4). <https://doi.org/10.33633/tc.v20i4.4463>
- Nasyuha, A. H., Zulham, Rusydi, I., 2022. Implementation of K-means algorithm in data analysis. *Telkomnika (Telecommunication Computing Electronics and Control)* 20(2). <https://doi.org/10.12928/TELKOMNIKA.v20i2.21986>
- Nurahman, N., Prihandoko, P., 2019. Perbandingan hasil analisis teknik data mining “metode decision tree, naive bayes, smo dan part” untuk mendiagnosa penyakit diabetes mellitus. *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi* 4(1). <https://doi.org/10.25139/inform.v4i1.1403>
- Nurahman, N., Purwanto, A., Mulyanto, S., 2022. Klasterisasi sekolah menggunakan algoritma k-means berdasarkan fasilitas, pendidik, dan tenaga pendidik. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer* 21(2), 337–350. <https://doi.org/10.30812/matrik.v21i2.1411>

- Putri, F. I., Damayanti, R., Kismiantini., 2022. Penerapan algoritma k-means untuk mengelompokan kecamatan di kabupaten gunungkidul berdasarkan program keluarga harapan. *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya* 2, 408–418.
- Putra, L. G. R., Anggrawan, A., 2021. Pengelompokan penerima bantuan sosial masyarakat dengan metode k-means. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer* 21(1), 205-214. <https://doi.org/10.30812/matrik.v21i1.1554>
- Rahmawati, E., Kisworo, B., 2017., Peran pendamping dalam pemberdayaan masyarakat miskin melalui program keluarga harapan. *Journal of Nonformal Education and Community Empowerment* 1(2), 161–169. <https://doi.org/10.15294/pls.v1i2.16271>
- Riza, L. S., Rosdiyana, R. A., Wahyudin, A., Pérez, A. R., 2021. The k-means algorithm for generating sets of items in educational assessment. *Indonesian Journal of Science and Technology* 6(1). <https://doi.org/10.17509/ijost.v6i1.31523>
- Sa'Adah, U., Rochayani, M. Y., Astuti, A. B., 2020. Knowledge discovery from gene expression dataset using bagging lasso decision tree. *Indonesian Journal of Electrical Engineering and Computer Science* 21(2). <https://doi.org/10.11591/ijeecs.v21.i2.pp1151-1159>
- Safitri, E., Sumertajaya, I. M., Nur Aidi, M., 2019. TSClust approach using k-means method to forecast vegetable food commodities inflation in dki jakarta. *International Journal of Scientific and Research Publications (IJSRP)* 9(9). <https://doi.org/10.29322/ijsrp.9.09.2019.p9357>
- Said, A. A., Defit, S., Yunus, Y., 2021. Klasterisasi dana bantuan pada program keluarga harapan (PKH) menggunakan metode k-means. *Jurnal Informatika Ekonomi Bisnis* 3(2), 53-59.
- Sari, D. P., Rosadi, D., Effendie, A. R., Danardono., 2019. K-means and bayesian networks to determine building damage levels. *Telkomnika (Telecommunication Computing Electronics and Control)* 17(2). <https://doi.org/10.12928/TELKOMNIKA.V17I2.11756>
- Swe, T. T., 2019. Analysis of tree based supervised learning algorithms on medical data. *International Journal of Scientific and Research Publications (IJSRP)* 9(4). <https://doi.org/10.29322/ijsrp.9.04.2019.p8817>
- Wijaya, Y. A., Kurniady, D. A., Setyanto, E., Tarihoran, W. S., Rusmana, D., Rahim, R., 2021. Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities. *TEM Journal* 10(3). <https://doi.org/10.18421/TEM103-13>