

Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam

Febriansyah Febriansyah ^{(1)*}, Siti Muntari ⁽²⁾

Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam

e-mail : {febriansyahh1213,muntariaza}@gmail.com.

* Penulis korespondensi.

Artikel ini diajukan 26 September 2022, direvisi 24 Januari 2023, diterima 25 Januari 2023, dan dipublikasikan 30 Januari 2023.

Abstract

The purpose of this study was to obtain a poverty data cluster in Pagar Alam City. The data collection of beneficiaries of the Program Keluarga Harapan (PKH) is not correct, the provision of assistance only pays attention to the criteria for poverty in general, so there are still many poor people who feel more deserving of PKH assistance. To overcome the problem of PKH recipients, it is necessary to cluster the community into various levels, so that the government can know the level of poverty of the community and can provide PKH assistance appropriately. The methods used in this study are CRISP-DM and the K-Means clustering algorithm. The attributes used are Identity Number, Name, Family Family Card Number, Poverty Rate, Pregnant Women, Early Childhood, Elementary School, Junior High School, Senior High School, Elderly, and Family Hope Program Recipient Group. This clustering process produced three clusters, namely cluster_0 as many as 156 people, cluster_1 as many as 82 people, and cluster_2 as many as 233 people. Furthermore, it was developed into a system with the Rapid Application Development (RAD) system development method. Thus producing a K-Means algorithm system to classify the poor in Pagar Alam City. The system test method uses black box testing with the alpha method and obtained database test results with a value of 4, interfaces with a value of 4, functionality of 4.42, and algorithms with a value of 4. In the testing process with UAT, in the system aspect got 87% of users agreed, in the user aspect 86% agreed, and in the interaction aspect 87% of users agreed. So it can be concluded that this system is worth using.

Keywords: Data Mining, Poverty, K-Means, Clustering, Black Box

Abstrak

Tujuan dari penelitian ini adalah untuk mendapatkan klaster data kemiskinan di Kota Pagar Alam. Pendataan penerima bantuan Program Keluarga Harapan (PKH) belum tepat, pemberian bantuan hanya memperhatikan kriteria kemiskinan secara umum, sehingga masih banyak masyarakat miskin yang merasa lebih pantas dapat bantuan PKH. Untuk mengatasi masalah penerima PKH tersebut perlu adanya pengklasteran untuk membagi masyarakat ke dalam berbagai tingkatan, sehingga pemerintah dapat mengetahui tingkat kemiskinan masyarakat dan dapat memberikan bantuan PKH secara tepat. Metode yang digunakan dalam penelitian ini ialah CRISP-DM dan algoritma *clustering* K-Means. Atribut yang digunakan adalah Nomor Induk Kependudukan, Nama, Nomor Kartu Keluarga Keluarga, Tingkat Kemiskinan, Ibu Hamil, Usia Dini, Sekolah Dasar, Sekolah Menengah Pertama, Sekolah Menengah Akhir, Lansia, dan Kelompok Penerima Program Keluarga Harapan. Proses *Clustering* ini menghasilkan tiga *cluster* yaitu *cluster_0* sebanyak 156 orang, *cluster_1* sebanyak 82 orang, dan *cluster_2* sebanyak 233 orang. Selanjutnya dikembangkan menjadi sebuah sistem dengan metode pengembangan sistem Rapid Application Development (RAD). Sehingga menghasilkan sistem algoritma K-Means untuk mengklasifikasikan penduduk miskin di Kota Pagar Alam. Metode pengujian sistem menggunakan *black box testing* dengan metode alpha dan didapatkan hasil pengujian *database* dengan nilai 4, antarmuka dengan nilai 4, fungsionalitas 4,42, dan algoritma dengan nilai 4. Pada proses pengujian dengan UAT aspek sistem mendapatkan 87% pengguna menyatakan setuju, aspek pengguna bernilai 86% setuju, dan pada aspek interaksi sebesar 87% pengguna setuju. Sehingga dapat disimpulkan jika sistem ini layak digunakan.

Kata Kunci: Data Mining, Kemiskinan, K-Means, Klasterisasi, Black Box



1. PENDAHULUAN

Berdasarkan Undang-Undang No. 24 Tahun 2004, kemiskinan adalah kondisi sosial ekonomi seseorang atau sekelompok orang yang tidak terpenuhinya hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat. Kebutuhan dasar yang menjadi hak seseorang atau sekelompok orang meliputi kebutuhan pangan, kesehatan, pendidikan, pekerjaan, perumahan, air bersih, pertanahan, sumber daya alam, lingkungan hidup, rasa aman dari perlakuan atau ancaman tindak kekerasan, dan hak untuk berpartisipasi dalam penyelenggaraan kehidupan sosial dan politik.

Berdasarkan hasil jumlah dan persentase penduduk miskin di Kota Pagar Alam selama periode tahun 2018-2019. Jumlah penduduk miskin tahun 2018-2019 tidak mengalami banyak perubahan karena persentase penduduk miskin di Pagar Alam masih berada angka 8,77% dan 8,90%. Namun, jumlah penduduk miskin justru mengalami peningkatan pada tahun 2020 berjumlah 12,71 ribu jiwa dengan persentase penduduk miskin 9,07% pada Kota Pagar Alam hal ini menunjukkan bahwa masalah kemiskinan masih menjadi masalah yang cukup serius di kota Pagar Alam terutama pada Kecamatan Dempo Selatan. Dari 455 jumlah penerima (keluarga penerima manfaat) bahwa yang tergolong miskin adalah sebanyak 4,55%, Kecamatan Dempo Tengah dari 471 jumlah penerima bahwa yang tergolong miskin adalah sebanyak 4,71%, Kecamatan Dempo Utara dari 450 jumlah penerima bahwa yang tergolong miskin adalah sebanyak 4,5%, Kecamatan Pagar Alam Selatan dari 441 jumlah penerima bahwa yang tergolong miskin adalah sebanyak 4,41%, dan Kecamatan Pagar Alam Utara dari 455 jumlah penerima bahwa yang tergolong miskin adalah sebanyak 4,55%.

Faktor-faktor pembagian kelompok penerima bantuan PKH dilihat dari keluarga miskin atau pra sejahtera, memiliki anggota keluarga dengan kriteria ibu hamil/menyusui, memiliki anak berusia 0 sampai dengan 6 tahun, memiliki anak dengan kategori pendidikan SD, SMP, atau SMA sederajat, memiliki keluarga lanjut usia minimal 60 tahun, dan penyandang distabilitas yang diutamakan peyandang distabilitas berat. Akan tetapi, dalam proses penyaluran bantuan keluarga miskin melalui Program Keluarga Harapan (PKH) ini masih menemui banyak permasalahan dalam penyalurannya. Di antaranya ialah pendataan dan penyaluran penerima bantuan PKH masih belum tepat, pemberian bantuan hanya memperhatikan kriteria kemiskinan secara umum, sehingga masih banyak masyarakat miskin yang merasa lebih pantas mendapatkan bantuan PKH.

Berdasarkan latar belakang tersebut dilakukanlah upaya penyelesaian yang salah satunya menggunakan proses *Data Mining*. *Data mining* merupakan proses analisa data untuk menemukan suatu pola dari kumpulan data. Tujuan dari *clustering* adalah mengelompokan item data ke dalam sejumlah kecil grup sedemikian rupa sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial (Bahauddin et al., 2021). *Data mining* dapat mengatasi masalah tersebut dengan proses pengklasteran untuk mengetahui yang mana termasuk *cluster* miskin tinggi, *cluster* miskin rendah, dan *cluster* miskin sedang, sehingga pemerintah dapat mengetahui tingkat kemiskinan masyarakat dan dapat memberikan bantuan PKH secara tepat.

Berdasarkan penelitian terdahulu tentang klasterisasi kemiskinan penduduk di provinsi oleh Nasution et al. (2020) menghasilkan pengelompokan kemiskinan sebanyak 8 provinsi dengan *cluster* tinggi dan 26 provinsi *cluster* rendah, hasil *clustering* selanjutnya dapat digunakan oleh pemerintah untuk memberikan perhatian lebih pada provinsi yang masih memiliki *cluster* kemiskinan tinggi. Metode yang pernah digunakan dalam penelitian sebelumnya yaitu metode algoritma K-Means sebagai metode *clustering* kemiskinan dan hasil evaluasi dapat divisualisasikan dalam bentuk peta sehingga mempermudah dalam melihat sebaran penduduk miskin (Astuti, 2017). Penelitian lainnya menggunakan metode K-Means *clustering* dilakukan oleh Paramitha et al. (2020) di mana teknik K-Means yaitu suatu metode penganalisaan data yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode *clustering* K-Means digunakan agar penelitian prioritas penduduk tidak mampu bisa lebih berkualitas dan efektif.



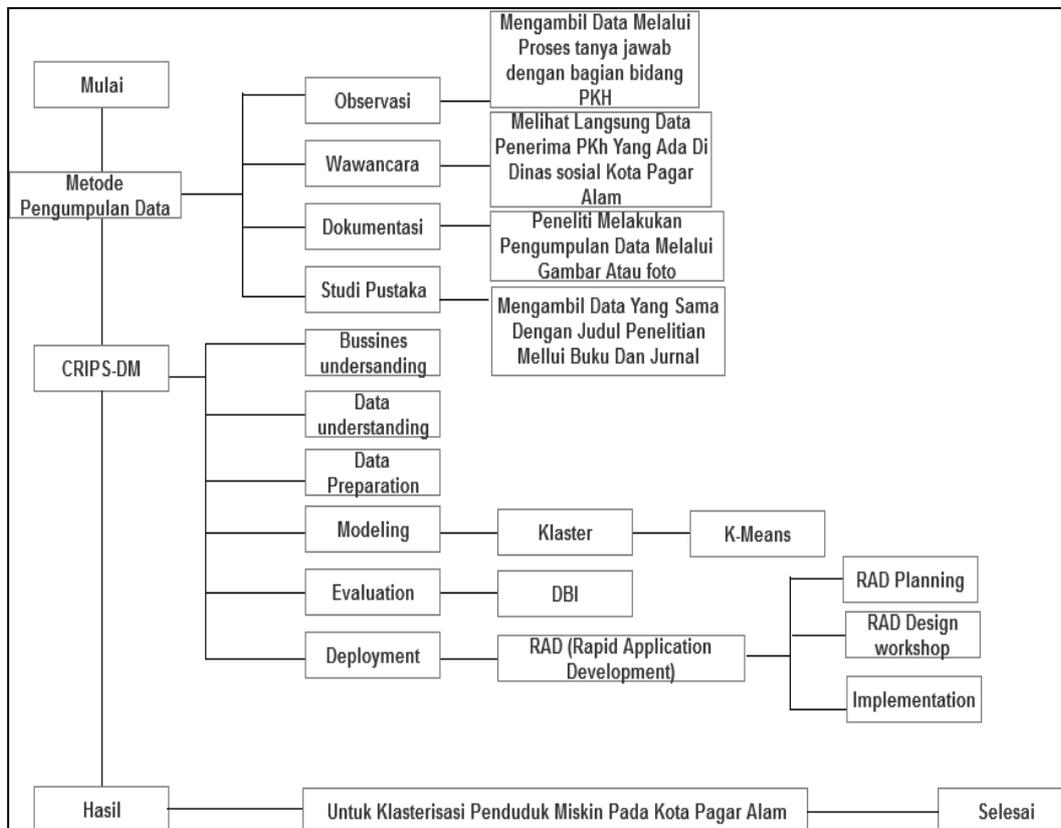
Beberapa penelitian lainnya mengenai penerapan *data mining* terhadap penanganan kemiskinan juga telah dilakukan oleh Jainuddin et al. (2018), Parjito & Permata (2021), dan Sudibyo et al. (2020). Sedangkan Fatmawati & Windarto (2018), Marzuki (2015), dan Sucipto (2019) melakukan penelitian mengenai implementasi algoritma K-Means.

Dalam penelitian ini akan mengelompokkan data kemiskinan kota Pagar Alam diharapkan dapat memberikan masukan kepada pemerintah agar dapat menjadikan hasil *clustering* sebagai bahan pertimbangan dalam penentuan penerima PKH secara tepat. Dalam penelitian sebelumnya hasil dari *cluster* dapat dijadikan masukan bagi pemerintah agar provinsi yang masuk ke dalam *cluster* tinggi mendapat perhatian lebih (Rofiqo et al., 2018) selain itu juga pengklasteran untuk membantu Dinas Sosial dalam pengelompokan keluarga miskin sehingga bantuan dapat tersalurkan dengan tepat (Paramitha et al., 2020). Pengujian pada penelitian ini menggunakan *black box testing* dengan metode alpha, yaitu memastikan aplikasi dapat berjalan dengan lancar tanpa gangguan (Masripah & Ramayanti, 2020).

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Pelaksanaan penelitian dilakukan dengan beberapa tahapan, dimulai dari proses pengumpulan data dengan metode observasi, wawancara, dokumentasi, dan studi pustaka. Berdasarkan data yang telah ada maka dilanjutkan proses klasterisasi dengan metode *CRISP-DM* dengan algoritma K-Means. Di mana pada tahapan ini terdapat proses *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Hasil dari proses klasterisasi selanjutnya akan diimplementasikan ke dalam sistem berbasis *website* yang dapat mengklaster penduduk miskin di Kota Pagar Alam. Adapun tahapan penelitian secara lengkap dapat dilihat pada Gambar 1.

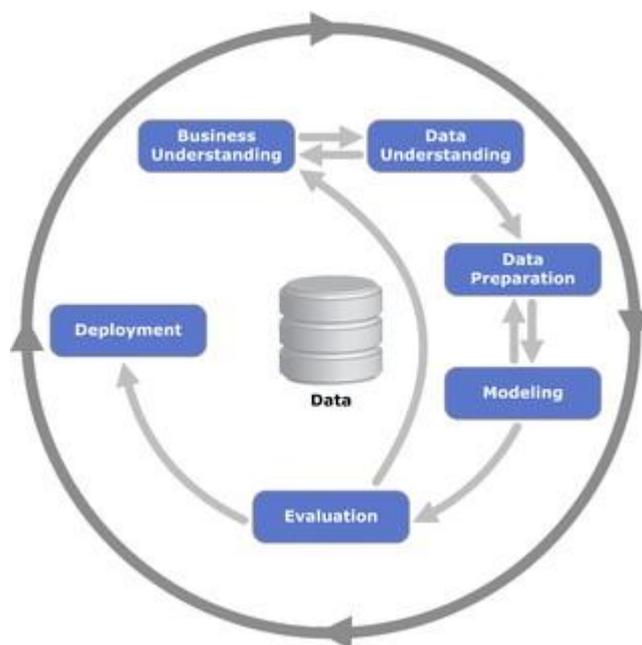


Gambar 1 Tahapan Penelitian



2.2 Cross Industry Standard Process untuk Data Mining (CRISP-DM)

Cross industry standard process untuk *data mining* atau CRISP-DM dikembangkan tahun 1996 oleh analisis dari beberapa industri seperti Daimler Chrysler, SPSS, dan NCR. CRISP-DM merupakan standarisasi proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM sebuah proyek *data mining* memiliki siklus hidup yang terbagi dan enam fase (Feblian, 2021). CRISP-DM bukan merupakan satu-satunya standar dalam *data mining* namun merupakan yang terpopuler saat ini. CRISP-DM merupakan metode yang menggunakan model proses pengembangan data yang banyak digunakan para ahli untuk memecahkan masalah terbukti 3 sampai 4 kali lebih banyak digunakan dibanding dengan standar lain yang digunakan. Gambar 2 merupakan gambaran secara umum mengenai siklus hidup dalam CRISP-DM.



Gambar 2 Siklus Hidup dalam CRISP-DM (Astuti, 2017)

2.3 Algoritma K-Means

K-Means merupakan salah satu algoritma *clustering* yang digunakan untuk mempartisi data ke dalam beberapa *cluster*. Di mana data yang memiliki tingkat kemiripan yang tinggi dikelompokkan dalam satu *cluster* sedangkan data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang berbeda (Butarbutar et al., 2017; Rahayu et al., 2019).

Dalam penelitian ini peneliti menggunakan data *sample* sebanyak 471 data keluarga dengan jumlah dalam keluarga dengan tipe *integer* berjumlah 8.306 orang. Cara pengelompokan menggunakan K-Means yaitu:

- 1) Menentukan banyaknya *cluster* yang dibentuk ada 3 *cluster* ($k=3$) hal ini didasarkan pada perhitungan *Euclidean Distance* yang telah dilakukan di mana ketika perhitungan nilai *Euclidean Distance* tidak lagi berubah pada bentukan 3 *cluster*. Penentuan *cluster* harus lebih kecil dari pada banyaknya data ($k < n$).
- 2) Menentukan nilai secara manual atau random untuk pusat *cluster* awal sebanyak *cluster* yang ditentukan.
- 3) Untuk menghitung jarak data dengan *centroid* menggunakan rumus *Euclidean Distance*. Persamaan *Euclidean Distance* ditunjukkan pada Pers. (1).



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

Di mana $d(x, y)$ adalah jarak antara data pusat x ke data pusat y , dengan x_i sebagai *data testing* ke i dan y_i sebagai *data training* ke i .

- 4) Mengecek setiap data berdasarkan kedekatannya dengan jarak terkecil.
- 5) *Centroid* baru dihitung dengan menghitung nilai rata-rata data pada setiap *cluster*.
- 6) Melakukan perulangan. Jika perhitungan iterasi baru berbeda dengan iterasi sebelumnya, maka proses dilanjutkan ke langkah perulangan selanjutnya. Namun jika iterasi yang baru dihitung sama dengan iterasi sebelumnya, maka proses *clustering* selesai. Dengan demikian, nilai pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

2.4 Klasterisasi

Klasterisasi adalah salah satu teknik yang digunakan dalam *data mining*. Pengertian klasterisasi dalam *data mining* menurut Sucipto (2019) adalah suatu teknik untuk mengelompokkan data ke dalam suatu klaster tertentu yang memungkinkan data dalam klaster tersebut memiliki kesamaan dan memiliki perbedaan yang jelas dengan data pada klaster lainnya. Sedangkan Marzuki (2015) berpendapat klasterisasi adalah suatu kumpulan objek atau data yang memiliki kesamaan di antara mereka dan data yang tidak memiliki kesamaan dimasukkan ke dalam klaster lain, sedangkan klasterisasi proses pengelompokan objek atau data ke dalam grup yang anggotanya memiliki kesamaan tertentu.

Maka dapat disimpulkan bahwa klasterisasi adalah metode pengelompokan data ke dalam suatu kesamaan tertentu. Hal ini berbeda dengan klasifikasi yaitu proses pengelompokan data baru berdasarkan kelompok atau klasifikasi yang sudah ada, *clustering* akan mengelompokkan data baru berdasarkan atribut dengan karakteristik yang sama yang dalam hal ini lebih cocok digunakan pada data penelitian ini.

3. HASIL DAN PEMBAHASAN

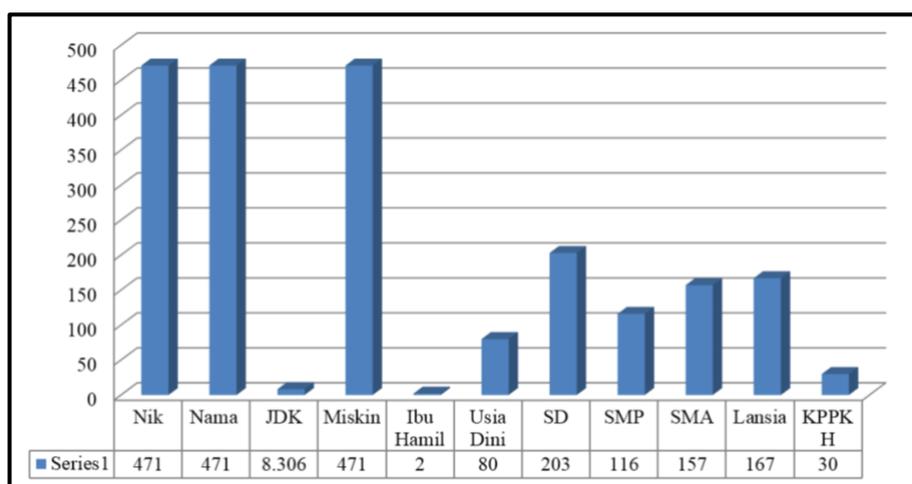
3.1 Pemahaman Bisnis (*Business Understanding*)

Pada pemahaman bisnis dilakukan tahapan menentukan tujuan penelitian dan ruang lingkup penelitian. Di mana pada tujuan penelitian ini adalah untuk menghasilkan *cluster* yang tinggi dengan menggunakan data PKH, sehingga nantinya model yang menghasilkan nilai *cluster* yang tinggi dapat digunakan untuk melakukan klasterisasi pada penerima PKH. Adapun ruang lingkup dalam penelitian ini ialah menggunakan metode *clustering* dengan algoritma K-Means.

3.2 Pemahaman Data (*Data Understanding*)

Pada fase pemahaman data ini, data didapat dari Dinas Sosial Kota Pagar Alam. Data yang diambil yaitu data Kecamatan Dempo Tengah pada tahun 2020. Terdapat 471 keluarga dengan jumlah dalam keluarga berjumlah 8.306 *record*. Atribut dalam data tersebut yaitu Jdk, Miskin, Ibu Hamil, Usia Dini, SD, SMP, SMA, Lansia dan KPPKH, Nik, Nama, dan Jumlah dalam Keluarga. Kategori data yang diterima dalam bentuk *excel* dan tidak terdapat *missing data* di dalamnya. Dalam penelitian ini atribut yang digunakan dapat dilihat pada Gambar 3.





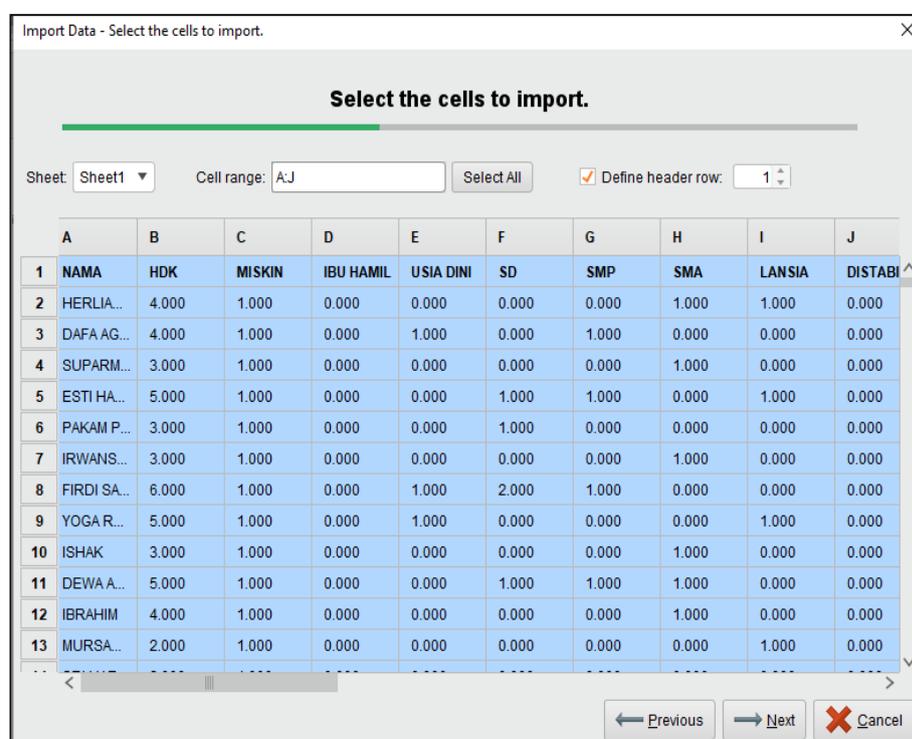
Gambar 3 Grafik Atribut

3.3 Pengolahan Data (*Data Preparation*)

Tahapan pengolahan data langsung diimplementasikan pada RapidMiner dengan tiga tahapan sebagai berikut.

3.3.1 *Data Selection*

Pada tahapan ini ada sepuluh atribut yang didapat dari Dinas Sosial Kota Pagar Alam dengan 471 data. Namun jika ingin menyeleksi data bisa menggunakan *cell range* yang telah disediakan pada *software* RapidMiner. Semua atribut pada data penelitian ini dapat digunakan sehingga peneliti tidak melakukan *selection data* seperti yang ditunjukkan pada Gambar 4.



Gambar 4 *Data Selection*



3.3.2 Data Processing

Pada tahap *data processing* memastikan bahwa tidak ada lagi *missing value* atau data yang kosong. Seperti yang terlihat pada Gambar 5 seluruh atribut yang digunakan tidak terdapat *missing value*.

Name	Type	Missing	Statistics		
NAMA ART	Polynomial	0	Least ZEPIANA LETISIA (1)	Most ARDIANSYAH (2)	Values ARDIANSYAH (2), IBRAHIM (2), ... [463 more]
cluster	Nominal	0	Least cluster_0 (225)	Most cluster_1 (246)	Values cluster_1 (246), cluster_0 (225)
HDK	Integer	0	Min 1	Max 7	Average 3.473
MISKIN	Integer	0	Min 1	Max 2	Average 1.004
IBU HAMIL	Integer	0	Min 0	Max 1	Average 0.006
USIA DINI	Integer	0	Min 0	Max 2	Average 0.170
SD	Integer	0	Min 0	Max 2	Average 0.420
SMP	Integer	0	Min 0	Max 2	Average 0.278

Gambar 5 Data Processing

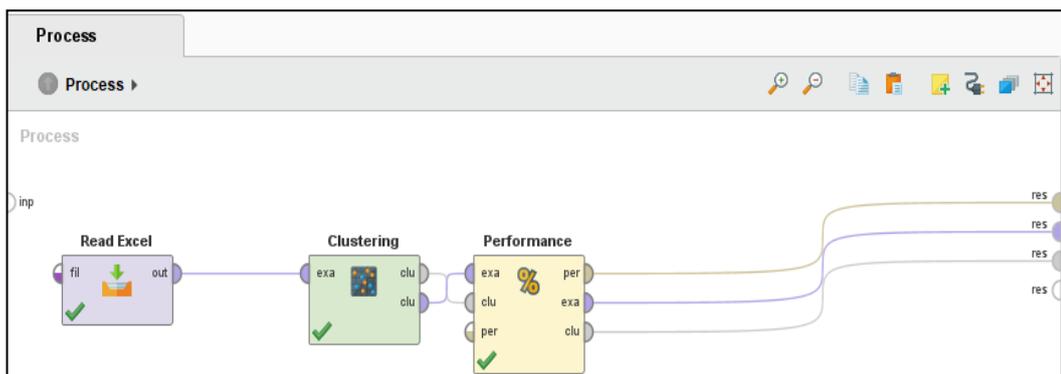
3.3.3 Data Transformation

Pada tahap *data transformation* data yang sudah diproses akan menampilkan atribut yang dipilih dan disatukan dalam RapidMiner. Atribut yang pertama yaitu nama penerima dengan tipe *polinomial* berjumlah 471 nama penerima. Atribut yang kedua yaitu jumlah dalam keluarga dengan tipe *integer* dengan jumlah 8.306 orang. Atribut yang ketiga yaitu miskin dengan tipe *integer* dengan jumlah 471 orang. Atribut yang keempat yaitu ibu hamil dengan tipe *integer* dengan jumlah 2 orang. Atribut yang kelima yaitu usia dini dengan tipe *integer* berjumlah 80 orang. Atribut yang keenam yaitu SD dengan tipe *integer* dengan jumlah 203 orang. Atribut yang ketujuh yaitu SMP dengan tipe *integer* dengan jumlah 116 orang. Atribut yang kedelapan yaitu SMA dengan tipe *integer* dengan jumlah 375 orang. Atribut yang kesembilan yaitu lansia dengan tipe *integer* dengan jumlah 167 orang. Atribut yang kesepuluh yaitu disabilitas dengan tipe *integer* dengan jumlah 30 orang.

3.4 Pemodelan (Modeling)

Tahapan pemodelan dilakukan menggunakan teknik klasterisasi dengan algoritma yang digunakan yaitu K-means. *Clustering* operator ini mengambil objek dari *port input* dan mengirimkan salinannya ke *port output*. Setiap *port* yang terhubung membuat salinan yang independen (tidak terikat). Jadi, ketika mengubah suatu salinan tidak berpengaruh pada salinan yang lainnya sehingga dapat dihubungkan dengan *performance* yaitu untuk mengetahui suatu model algoritma K-means. Hasil pemodelan yaitu berupa pola informasi yang dapat memudahkan pihak yang berkepentingan seperti yang terlihat pada Gambar 6.





Gambar 6 Model Algoritma K-Means

1) Tentukan jumlah *cluster*

Untuk menentukan jumlah *cluster* dilakukan percobaan jumlah *cluster* 3. Percobaan 3 *cluster* itu ada C0, C1 dan C2 dengan atribut Jdk, Miskin, Ibu hamil, Usia dini, SD, SMP, SMA, Lansia, dan KPPKH. Dengan pengukuran *performance vector* rata-rata dalam *centroid distance* yang didapat dengan nilai 0,154 kemudian rata-rata *cluster_0* bernilai 0,187, rata-rata dalam *cluster_1* bernilai 0,129, dan rata-rata dalam *centroid cluster_2* bernilai 0,141 dengan nilai Davies Bouldin Index 0,160 seperti yang ditunjukkan pada Gambar 7.

Attribute	cluster_0	cluster_1	cluster_2
HDK	4.474	1.817	3.386
MISKIN	1.013	1	1
IBU HAMIL	0	0	0.013
USIA DINI	0.346	0	0.112
SD	1.051	0.037	0.133
SMP	0.327	0.073	0.318
SMA	0.308	0.061	0.425
LANSIA	0.154	1	0.236
DISTABILITAS	0.051	0.073	0.069

Gambar 7 Percobaan 3 Cluster

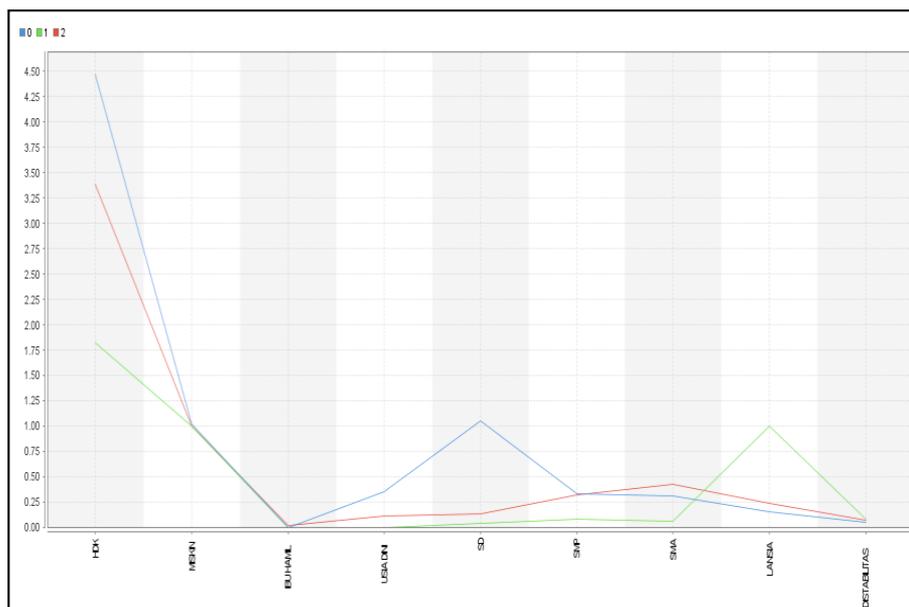
2) Menentukan titik *centroid* secara acak

Titik *centeroid* secara acak ditentukan berdasarkan beberapa percobaan *cluster* di RapidMiner. Dari 3 percobaan tersebut dipilih titik *centeroid* terkecil untuk mendapatkan hasil terbaik.

3) Hitung jarak data ke *centroid*

Setelah memasukkan *dataset* dan algoritma K-Means pada RapidMiner maka proses data dengan klik run untuk melihat jarak data ke *centeroid*. Grafik jarak data ke *centroid* dapat dilihat pada Gambar 8.





Gambar 8 Jarak Data ke *Centroid*

4) Perbarui nilai titik *centroid*

Setelah dilakukan beberapa percobaan, titik *centroid* pada 3 *cluster* tidak mengalami perubahan. Apabila *centeroid* berubah maka dilakukan perulangan iterasi namun jika tidak berubah maka alhasil pengulangan dihentikan dan telah didapatkan masing-masing kelompok.

Berdasarkan perhitungan RapidMiner maka diperoleh pola yang nantinya akan diimplementasikan pada sistem. Pola yang digunakan untuk mengelompokkan/*cluster* data berdasarkan hasil perhitungan jarak adalah:

- a) Jika $C_0 < C_1$ dan $C_0 < C_2$ maka *cluster* 0 dengan keterangan sedang.
- b) Jika $C_1 < C_0$ dan $C_1 < C_2$ maka *cluster* 1 dengan keterangan rendah.
- c) Jika $C_2 < C_0$ dan $C_2 < C_1$ maka *cluster* 2 dengan keterangan tinggi.

Maka dari pola yang didapatkan dari RapidMiner yang digunakan pada sistem dengan metode *clustering* K-Means diperoleh bahwa *cluster_0* memiliki tingkat kesejahteraan sedang dengan jumlah 156, *cluster_1* memiliki tingkat kesejahteraan rendah dengan jumlah 82, dan *cluster_2* memiliki tingkat kesejahteraan tinggi dengan jumlah 233. Setelah dilakukan proses *clustering* maka dapat diketahui *cluster* penerima berdasarkan tingkat kesejahteraaannya dan yang lebih direkomendasikan untuk mendapatkan PKH adalah dengan tingkat kesejahteraan rendah yang berada pada *cluster_1*.

3.5 Evaluation

Pengujian dilakukan dengan menggunakan Davies Bouldin Index (DBI) terhadap 3 *cluster* dengan keseluruhan data 471 *record*. *Performance vector* rata-rata dalam *centroid* yang didapat bernilai 0,154, kemudian rata-rata *centroid_0* bernilai 0,187, rata-rata dalam *centroid_1* bernilai 0,129, dan rata-rata dalam *centroid_2* berniali 0,141, kemudian nilai pengukuran jarak antar titik dalam *cluster* didapat nilai DBI 0,160 seperti yang ditunjukkan pada Gambar 9.



```

PerformanceVector

PerformanceVector:
Avg. within centroid distance: 0.154
Avg. within centroid distance_cluster_0: 0.187
Avg. within centroid distance_cluster_1: 0.129
Avg. within centroid distance_cluster_2: 0.141
Davies Bouldin: 0.160
    
```

Gambar 9 Nilai DBI

3.6 Deployment

Pada tahapan *deployment* ini ialah tahapan terakhir yaitu menerapkan model algoritma *clustering* yang dihasilkan pada bahasa pemrograman *framework php* dengan metode pengembangan *Rapid Application Development (RAD)* untuk dapat mengelompokkan data penerima PKH. Gambar 10 merupakan halaman Data Penerima yang terdapat informasi berupa data penerima PKH.

No	Nik	Nama	JDK	Miskin	Ibu Hamil	Usia Dini	Sd	Smp	Sma	Lansia	KPPKH	Hasil Cluster	Action
1	1672010912750001	HERLIANTO	4	1	0	0	0	0	1	1	0	Cluster_2	cluster Update Delete
2	1672011001760002	DAFA AGUSTIAN	4	1	0	1	0	1	0	0	0	Cluster_2	cluster Update Delete
3	1672012505500034	SUPARMAN	3	1	0	0	0	0	1	0	0	Cluster_2	cluster Update Delete
4	1672051204740001	ESTI HARIYATI	5	1	0	0	1	1	0	1	0	cluster_0	cluster Update Delete
5	1672034107520013	PAKAM PUTRA SEMIDANG	3	1	0	0	1	0	0	0	0	Cluster_2	cluster Update Delete
6	1672040708050003	IRWANSYAH	3	1	0	0	0	0	1	0	0	Cluster_2	cluster Update Delete
7	1672035206790002	FIRDI SAFUTRA	6	1	0	1	2	1	0	0	0	cluster_0	cluster Update Delete
8	1672051408150001	YOGA RAMADHAN	5	1	0	1	0	0	0	1	0	cluster_0	cluster Update Delete
9	1672016010860002	ISHAK	3	1	0	0	0	0	1	0	0	Cluster_2	cluster Update Delete
10	1672051009850001	DEWA ANGGA SAPUTRA	5	1	0	0	1	1	1	0	0	cluster_0	cluster Update Delete
11	1672015209070001	IRRAHIM	4	1	0	0	0	0	1	0	0	Cluster_2	cluster Update Delete

Gambar 10 Halaman Data Penerima Admin

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan menghasilkan klusterisasi penduduk miskin pada Kota Pagar Alam. Hasil dari proses *clustering* data dalam menentukan penerima PKH melalui aplikasi RapidMiner sama dengan hasil yang diterapkan di sistem yang telah dibangun dengan jumlah *cluster* yang terdiri dari tiga *cluster*. *Cluster* dimulai dari *cluster_0*, *cluster_1*, dan *cluster_2*. Data yang berada pada *cluster_0* berjumlah 156 data, *cluster_1* berjumlah 82 data, dan *cluster_2* berjumlah 233 data. *Performance vector* rata-rata dalam *centroid* yang didapat bernilai 0,154, kemudian rata-rata *centroid_0* bernilai 0,187, rata-rata dalam *cluster_1* bernilai 0,129, dan rata-rata dalam *centroid cluster_2* berniali 0,141, kemudian nilai pengukuran jarak antar titik dalam *cluster* didapat nilai DBI 0,160. Hasil pengujian *black box testing* berupa pengujian *alpha* menggunakan kuesioner yang diisi oleh pakar menghasilkan nilai pengujian *database* sebesar 4,



pengujian antarmuka bernilai 4, pengujian fungsionalitas bernilai 4,42, dan untuk algoritma bernilai 4. Dengan demikian, *alpha* memperoleh nilai kelayakan rata-rata nilai 4 sehingga menyimpulkan bahwa sistem layak digunakan. Pada proses pengujian dengan UAT aspek sistem 87% pengguna menyatakan setuju, aspek pengguna sebanyak 86% setuju, dan pada aspek interaksi sebesar 87% pengguna setuju. Berdasarkan hasil pengujian yang telah dilakukan tersebut maka sistem layak digunakan.

Penelitian ini menghasilkan *cluster* pada data penduduk miskin penerima bantuan PKH, dengan adanya hasil *cluster* ini pemerintah dapat lebih bijak dalam menentukan penerima bantuan ataupun membuat kebijakan. Sistem *clustering* yang dibuat akan mengelompokan data penerimaan bantuan secara tepat dengan algoritma K-Means yang teruji berdasarkan data sebelumnya.

DAFTAR PUSTAKA

- Astuti, F. D. (2017). Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma Hard C-Means. *Data Manajemen Dan Teknologi Informasi (DASI)*, 18(1), 64–69. <https://ojs.amikom.ac.id/index.php/dasi/article/view/1836>
- Bahauddin, A., Fatmawati, A., & Sari, F. P. (2021). Analisis Clustering Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 4(1), 1–8. <https://doi.org/10.36595/misi.v4i1.216>
- Butarbutar, N., Windarto, A. P., Hartama, D., & Solikhun, S. (2017). Komparasi Kinerja Algoritma Fuzzy C-Means dan K-Means dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 1(1), 46. <https://doi.org/10.30645/jurasik.v1i1.8>
- Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer dengan K-Means Cluster pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi. *Computer Engineering, Science and System Journal*, 3(2), 173. <https://doi.org/10.24114/cess.v3i2.9661>
- Feblian, D. (2021). Implementasi Model CRISP-DM untuk Menentukan Sales Pipeline pada PT. X [Universitas Trisakti]. In *THESIS-2016*. http://repository.trisakti.ac.id/usaktiana/index.php/home/detail/detail_koleksi/0/THE/judul/000000000000105201/
- Jainuddin, J., Agus, F., & Astuti, I. F. (2018). Sistem Informasi Data Kriteria Rakyat Miskin Desa Liang Ilir Kecamatan Kota Bangun. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 13(1), 39. <https://doi.org/10.30872/jim.v13i1.1004>
- Marzuki, I. (2015). Temu Kembali Informasi Big Data Menggunakan K-Means Clustering. *SMATIKA JURNAL : STIKI Informatika Jurnal*, 5(02), 01–07. <https://doi.org/10.32664/SMATIKA.V5I02.75>
- Masripah, S., & Ramayanti, L. (2020). Penerapan Pengujian Alpha dan Beta pada Aplikasi Penerimaan Siswa Baru. *Swabumi (Suara Wawasan Sukabumi) : Ilmu Komputer, Manajemen, Dan Sosial*, 8(1), 100–105. <https://doi.org/10.31294/SWABUMI.V8I1.7448>
- Nasution, I., Windarto, A. P., & Fauzan, M. (2020). Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi. *Building of Informatics, Technology and Science (BITS)*, 2(2), 76–83. <https://doi.org/10.47065/bits.v2i2.492>
- Paramitha, I. A. S. D., Sasmita, G. M. A., & Raharja, I. M. S. (2020). Analisis Data Log IDS Snort dengan Algoritma Clustering Fuzzy C-Means. *Majalah Ilmiah Teknologi Elektro*, 19(1), 95. <https://doi.org/10.24843/MITE.2020.v19i01.P14>
- Parjito, P., & Permata, P. (2021). Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Metode K-Means. *Ainet: Jurnal Informatika*, 3(1), 31–37. <https://doi.org/10.26618/AINET.V3I1.5878>
- Rahayu, A. E., Hikmah, K., Yustia, N., & Fauzan, Abd. C. (2019). Penerapan K-Means Clustering Untuk Penentuan Klasterisasi Beasiswa Bidikmisi Mahasiswa. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 82–86. <https://doi.org/10.28926/ilkomnika.v1i2.23>



- Rofiqo, N., Windarto, A. P., & Hartama, D. (2018). Penerapan Clustering pada Penduduk yang Mempunyai Keluhan Kesehatan dengan Datamining K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2(1). <https://doi.org/10.30865/komik.v2i1.929>
- Sucipto, A. (2019). Klasterisasi Calon Mahasiswa Baru Menggunakan Algoritma K-Means. *Science Tech: Jurnal Ilmu Pengetahuan Dan Teknologi*, 5(2), 50–56. <https://doi.org/10.30738/jst.v5i2.5829>
- Sudibyo, N. A., Iswardani, A., Sari, K., & Suprihatiningsih, S. (2020). Penerapan Data Mining pada Jumlah Penduduk Miskin di Indonesia. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(3), 199–207. <https://doi.org/10.46306/lb.v1i3.42>

