

## **IMPLEMENTASI DATA MINING ALGORITMA K-MEANS UNTUK KLASIFIKASI PENDUDUK MISKIN BERDASARKAN TINGKAT KEMISKINAN DI JAWA BARAT**

**Erna Nurliana, Bambang Irawan, Agus Bahtiar**

Teknik Informatika, Stmik Ikmi Cirebon

Jalan Perjuangan No.10B

*ernanurliana31@gmail.com*

### **ABSTRAK**

Penelitian ini menggunakan algoritma K-Means untuk menganalisis data populasi penerima manfaat Program Keluarga Harapan (PKH) di Desa Inbanagara Raya. Tujuan penelitian adalah menerapkan teknik data mining untuk menentukan prioritas penerima manfaat PKH. Uji Davies-Bouldin dilakukan dengan uji cluster dari 2 hingga 10, dan hasilnya menunjukkan nilai terbaik pada cluster ke-8 dengan indeks 0,002. Data ini mencakup atribut dan kriteria penerima PKH yang memenuhi standar yang ditetapkan. Diharapkan hasilnya dapat mendukung distribusi bantuan yang lebih merata kepada masyarakat desa Inbanagara Raya, membantu mengurangi beban kemiskinan dan memutus rantai kemiskinan antargenerasi.

**Kata Kunci:** Kemiskinan, Warga, Jawa barat

### **1. PENDAHULUAN**

Program Keluarga Harapan (PKH) merupakan salah satu upaya pemerintah untuk membantu masyarakat kurang mampu untuk mengatasi permasalahan ini. Ini adalah program untuk mengurangi kemiskinan. Masyarakat Desa Palasah menerima bantuan berupa: Alat bantu menjahit. Menentukan penerima bantuan peralatan menjahit dari otoritas public Kota Palasah sebenarnya menggunakan strategi adat untuk mengumpulkan informasi penduduk atau calon penerima bantuan peralatan menjahit. Informasi dikumpulkan kemudian tahap pilihan selesai. Pilihan yang dilakukan Pemerintah Kota juga masih mengarah pada strategi manual namun tidak terkoordinasi, lebih spesifiknya dengan melihat dan mengisolasi secara individu, informasi mengenai penduduk atau calon penerima manfaat yang memenuhi aturan dan dengan siapa tidak memenuhi langkah-langkah tersebut.

Program Keluarga Harapan (PKH), juga dikenal sebagai Bantuan Tunai Bersyarat dalam istilah internasional, adalah program bantuan tunai bersyarat untuk keluarga berpenghasilan rendah (KM). Di Indonesia, pemberlakuan PKH dimulai pada tahun 2007, dengan tujuan untuk menciptakan sistem jaminan sosial bagi keluarga berpenghasilan rendah (KM) dalam rangka meningkatkan kualitas hidup mereka melalui perbaikan tindakan terhadap Pendidikan dan kesehatan, serta peningkatan kualitas hidup. mendorong tercapainya kesejahteraan sosial. PKH juga dimaksudkan untuk mengurangi beban keuangan keluarga (dampak konsumsi langsung) sekaligus meningkatkan tabungan untuk generasi mendatang dengan meningkatkan kesehatan dan pendidikan anak. PKH seharusnya memutus siklus kemiskinan antar generasi dalam jangka Panjang.

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 penyandang distabilitas berat. Akan tetapi, dalam proses penyaluran bantuan keluarga miskin melalui Program Keluarga Harapan (PKH) ini masih menemui banyak permasalahan dalam penyalurnya. 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).

Informasi kependudukan yang ada akan dipelajari terlebih dahulu, kemudian diselidiki dan Hasil investigasi ini digunakan untuk mengetahui siapa saja individu yang membutuhkan dan berhak mendapatkan bantuan sosial terbatas dari PKH, namun secara umum hal ini sangat besar Alat angkut bantuan PKH banyak/ada pula yang tidak sesuai jalur Hal tersebut.

Hal ini dikarenakan banyak masyarakat tidak mampu yang tidak mendapatkan bantuan PKH, salah satunya karena situasi dengan keluarga kurang mampu yang memenuhi syarat untuk mendapatkan bantuan masih belum menentu ideal.[1]

Maksud dari pemeriksaan ini adalah untuk membedakan susunan faktor-faktor yang sangat mempengaruhi kebutuhan informasi. Selain itu, hasil ini juga dapat memberikan informasi tambahan dalam meningkatkan sasaran kemiskinan di Jawa Barat bagi para pengaruh kepentingan, khususnya bagi pemerintah Jawa Barat, pemerintah pusat dan masyarakat untuk mempercepat sasaran kemiskinan.

Mengingat penjelasan dasar di atas, maka judul yang diambil dalam eksplorasi ini adalah pelaksanaan penambangan informasi ke sekelompok jaringan putus ada menggunakan perhitungan k-implies.

## 2. TINJAUAN PUSTAKA

### 2.1. Data Mining

*Data mining* adalah suatu metode pengolahan data untuk menemukan. Pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data mining adalah pengolahan data dengan skala besar, sehingga data mining memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi (Ong, 2013). Data mining juga dapat dilakukan pada berbagai jenis database dan penyimpanan informasi, namun jenis pola yang akan ditemukan ditentukan oleh berbagai fungsi data mining seperti deskripsi class/konsep, asosiasi, analisa korelasi, klasifikasi, prediksi, analisa cluster dan lain-lain (Dash et al., 2010).[2]

### 2.2. Pengelompokan Data (Clustering)

*Clustering* (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. *Clustering*

membagi kumpulan data ke dalam beberapa 8 kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar daripada diantara kelompok-kelompok (Rui Xu dan Donald. 2009). Gagasan mengenai pengelompokan data atau clustering, memiliki sifat yang sederhana dan dekat dengan cara berpikir manusia, kapanpun kepada kita dipresentasikan jumlah data besar ini ke dalam sejumlah kecil kelompok-kelompok atau kategori-kategori untuk memfasilitasi analisanya lebih lanjut. Selain dari itu sebagian besar data yang dikumpulkan dalam banyak masalah terlihat memiliki beberapa sifat yang melekat yang mengalami pengelompokan-pengelompokan natural (Hammuda dan Karay, 2003).[3]

### 2.3. Algoritma K-Means

*K-means* termasuk dalam partitioning *clustering* yaitu setiap data harus masuk dalam *cluster* tertentu dan memungkinkan bagi setiap data yang termasuk dalam cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* yang lain. *K-means* memisahkan data ke k daerah bagian yang terpisah, dimana k adalah bilangan integer positif. Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K cluster yang sudah diterapkan di awal. Algoritma *K-Means* sederhana untuk diimplementasikan dan dijalankan, relative cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-Means* menjadi salah satu algoritma yang paling penting dalam bidang data mining (Wu dan Kumar, 2009)[4]

Menurut Eko Prasetyo (2012) mengatakan bahwa metode K-Means ini mempartisi data kedalam kelompok sehingga data berkarakteristik sama dimasukan kedalam sat kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain.[5] Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk.
2. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara, namun yang paling sering dilakukan adalah dengan cara random yang diambil dari data yang ada.
3. Menghitung jarak setiap data input terhadap masing – masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid

Berikut adalah persamaan Euclidian Distance :  

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

Dimana :

$d$  : titik dokumen

$x_i$  : data kriteria

$\mu_j$  : centroid pada cluster ke- $j$

4. Mengklasifikasikan setiap data berdasarkan kedadatannya dengan *centroid* (jarak terkecil).
5. Memperbarui nilai *centroid*. Nilai *centroid* baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus:

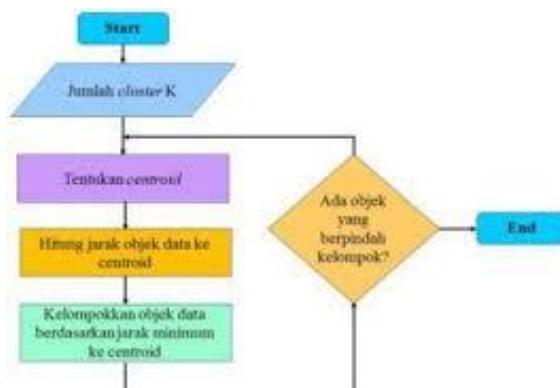
$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in sj} x_j$$

Dimana:

$\mu_j(t+1)$  : centroid baru pada iterasi ke  $(t+1)$   
 $N_{sj}$  : banyak data pada cluster  $s_j$

6. Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

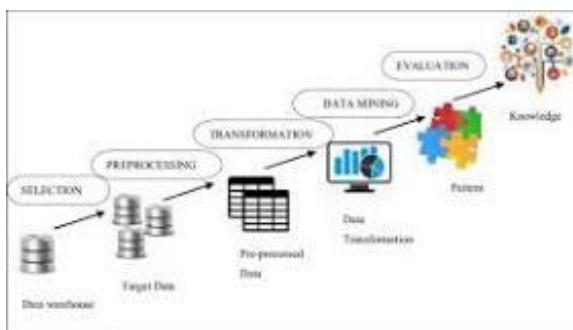
Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* ( $\mu_j$ ) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.



Gambar 1. Flowchart Algoritma K-Means Clustering

### 3. METODE PENELITIAN

Menemukan informasi yang tersembunyi di dalam tumpukan record pada sebuah basis data memerlukan pendekatan tersendiri. Salah satu pendekatan tersebut adalah KDD (Knowledge Discovery in Database). KDD merupakan proses penggalian informasi dari sebuah basis data yang besar, yang mana proses ini terdiri dari beberapa tahapan. Adapun tahapan – tahapan dalam KDD dijelaskan sebagai berikut:



Gambar 2. Metode Penelitian

Sumber data yang digunakan dalam penelitian ini diperoleh dari desa, sebuah platform terkemuka dalam komunitas data science yang menyediakan akses kepada peneliti, ilmuwan data, dan praktisi di berbagai bidang. Data yang digunakan yakni data transaksi

penjualan. Dari data yang sudah ada dilanjutkan dengan proses preparasi data melalui tahapan data selection, data cleaning, dan data transformation.

#### 3.1. Selection

Penyeleksian atribut yang akan digunakan ada pada tahap ini. Tidak semua atribut dapat digunakan, hanya atribut yang relevan untuk dianalisis. Data hasil seleksi yang akan digunakan dalam proses data mining disimpan dalam suatu berkas yang terpisah dari basis data operasional.[6]

#### 3.2. Preprocessing

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan tahapan *preprocessing* pada data yang menjadi fokus KDD. Tahapan *preprocessing* mencakup antara lain membersihkan data dari noise, seperti duplikasi data, memeriksa data yang inkonsisten, memeriksa missing value, memperbaiki kesalahan pada data, seperti kesalahan cetak.

#### 3.3. Transformation

Data harus dilakukan transformasi sebelum dilakukan pengolahan menggunakan data mining. Hal ini bertujuan untuk menyesuaikan data yang diolah berdasarkan algoritma dan *software* yang digunakan di dalam pengolahan data.[7]

#### 3.4. Data Mining

Ini adalah proses menemukan pola dan informasi menarik dalam data terpilih dengan menggunakan teknik dan metode tertentu. Teknik, metode, dan algoritma penambangan data sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan Anda dan proses KDD secara keseluruhan.

#### 3.5. Evaluasi

Merupakan proses menginterpretasikan hasil *rule* yang didapatkan dari teknik data mining. Pada bagian ini diberikan rekomendasi strategi penjualan produk berdasarkan hasil *rule* yang didapat dari hasil pengolahan data

### 4. HASIL DAN PEMBAHASAN

#### 4.1. Hasil Penelitian

Implementasi algoritma *K-Means Clustering* ini dirancang dengan menggunakan *tools* RapidMiner. Pada penelitian ini, digunakan tahapan *Knowledge Discovery in Database* (KDD) yang terdiri dari beberapa tahap, sebagai berikut:

##### 4.1.1. Data Selection

Pada tahap ini, gunakan administrator Read Succeed untuk mengimpor informasi desain Succeed yang disimpan di PC klien ke dalam proses penanganan informasi berkelanjutan dalam perangkat lunak RapidMiner. Batas yang digunakan adalah batas default. adalah operator Baca Excel.



Gambar 3. Data Selection

Data pengelompokan Masyarakat kurang mampu selama ini tidak tersusun dengan baik sehingga data ini

tidak dimanfaatkan dengan baik untuk mengetahui Masyarakat yang layak dan tidak layak oleh karena itu data tersebut akan digunakan dalam penelitian ini dengan tema Masyarakat layak dan tidak layak menggunakan metode k-means dengan atribut, nama, jenis kelamin, pekerjaan, status perkawinan, Pendidikan terakhir dan penghasilan per bulan.

Berikut ini merupakan data yang diolah menggunakan rapidminer.

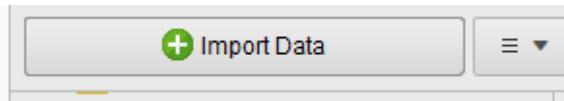
Tabel 1. Data set

no	Nama	Jenis kelamin	Pekerjaan	Status perkawinan	Pendidikan terakhir	Penghasilan perbulan
1	Nusa	L	Perdagangan	Kawin	SLTP/sederajad	1000.000/2000.000
2	Rini	P	IRT	Kawin	SD/sederajad	0
3	Rizkiadian	L	PNS	Belum kawin	SLTA/Sederajat	2.000.000-3.000.000
---	---	----	-----	-----	-----	-----
195	Passa ardiansyah	L	Tidak Bekerja	Belum Kawin	SLTP/Sederajat	0
196	Alfian	P	Tidak bekerja	Belum kawin	Belum tamat Sd	0

#### 4.1.2. Data Preprocessing

##### 1) Import Data

Pada gambar 4 data yang berekstensi xls atau excel dimasukan kedalam rapidminer untuk diolah dalam proses data mining.



Gambar 4. Import Data

Menampilkan proses setelah data di masukan kedalam rapid miner untuk dipilih cell rangenya.

Select the cells to import.										
Sheet: Sheet1 Cell range: B:M Select All Define header row: 1										
A	B	C	D	E	F	G	H	I	J	
1	NO	NAMA	JK	TEMPAT	TANGGA..	USIA	ALAMAT		PEKERJAAN	
2	1.000	NUSA	L	CIAMIS	Aug 15, ..	50.000	BLOK SA...	1.000	1.000	PERDAGANGAN...
3	2.000	RINI	P	CIAMIS	2004/19...	46.000	BLOK SA...	1.000	1.000	MENGURUS RU...
4	3.000	RIZKI AD...	L	CIAMIS	16/01/19...	27.000	BLOK SA...	1.000	1.000	PEGAWAI NEGE...
5	4.000	HAKAM ...	L	CIAMIS	16/05/20...	14.000	BLOK SA...	1.000	1.000	BELUM TIDAK B...
6	5.000	NANA S...	L	CIAMIS	15/07/19...	48.000	BLOK SA...	1.000	1.000	SOPIR(B1)
7	6.000	NURHAL...	P	CIAMIS	Oct 8, 19...	43.000	BLOK SA...	1.000	1.000	MENGURUS RU...
8	7.000	IDA PUR...	P	CIAMIS	18/08/19...	23.000	BLOK SA...	1.000	1.000	KARYAWAN SWA...
9	8.000	WILDAN ...	L	CIAMIS	Oct 8, 1979...	12.000:00 AM ICTA	BLOK SA...	1.000	1.000	BELUM TIDAK B...
10	9.000	RUSGANDA	L	CIAMIS	28/07/20...	74.000	BLOK SA...	1.000	1.000	WIRASWASTA(88)
11	10.000	SUNERAH	P	CIAMIS	17/12/19...	67.000	BLOK SA...	1.000	1.000	MENGURUS RU...
12	11.000	SUDARMA	L	CIAMIS	15/03/19...	62.000	BLOK SA...	1.000	1.000	BURUJ...

Gambar 5. Hasil Import Data

Pada gambar 5 menampilkan proses setelah pemilihan cell range dan yang akan diolah oleh operator.

Format your columns.						
<input type="checkbox"/> Replace errors with missing values ⓘ						
B	C	D	E	F	G	
polynomial	polynomial	polynomial	polynomial	polynomial	polynomial	
id	JK	TEMPAT	TANGGAL LAHIR	USA	ALAMAT	
1	NAMA	L	CIAMIS	26526	50	BLOK SAWAH
2	NUSA	P	CIAMIS	2004/19/76	46	BLOK SAWAH
3	RINI	L	CIAMIS	16/01/1996	27	BLOK SAWAH
4	RIZKI ADIANTO	L	CIAMIS	15/06/2000	14	BLOK SAWAH
5	HAKAM SYUKUR	L	CIAMIS	15/07/1974	48	BLOK SAWAH
6	NANA SUTISNA	L	CIAMIS	29/12/1999	23	BLOK SAWAH
7	NURHASANAH	P	CIAMIS	28/07/2003	19	BLOK SAWAH
8	IDA PURWATI	P	CIAMIS	18/07/1948	74	BLOK SAWAH
9	WILDAN PRATAMA	L	CIAMIS	17/12/1955	67	BLOK SAWAH
10	RUSGANDA	L	CIAMIS	16/03/1980	60	BLOK SAWAH
11	SUNERAH	P	CIAMIS			
12	SUDARMA	I	CIAMIS			

Gambar 6. Hasil

Format your columns.						
<input type="checkbox"/> Replace errors with missing values ⓘ						
H	I	J	K	L	M	
integer	integer	polynomial	polynomial	polynomial	polynomial	
1	?	?	PEKERJAAN	STATUS PERKA...	PENDIDIKAN TE...	PENGHASILAN P...
2	1	1	PERDAGANGAN...	KAWIN(2)	SLTP/SEDERAJAT	1.000.000-2.000...
3	1	1	MENGURUS RU...	KAWIN(2)	TAMAT SD / SED...	0
4	1	1	PEGAWAI NEGE...	BELUM KAWIN(1)	SLTA / SEDERA...	2.000.000-3.000...
5	1	1	BELUM/TIDAK B...	BELUM KAWIN(1)	TAMAT SD / SED...	0
6	1	1	SOPIR(B1)	KAWIN(2)	TAMAT SD / SED...	1.000.000-2.000...
7	1	1	MENGURUS RU...	KAWIN(2)	TAMAT SD / SED...	0
8	1	1	KARYAWAN SWA...	BELUM KAWIN(1)	SLTA / SEDERA...	2.000.000-3.000...
9	1	1	BELUM/TIDAK B...	BELUM KAWIN(1)	SLTA / SEDERA...	0
10	1	1	WIRASWASTA(88)	KAWIN(2)	DIPLOMA IV / STR...	2.000.000-3.000...
11	1	1	MENGURUS RU...	KAWIN(2)	TAMAT SD / SED...	0
12	1	1	RUPI(MARDIAM)	CERAI MATI(4)	TIDAK JADI U...	<1.000.000.000

Gambar 7. Hasil

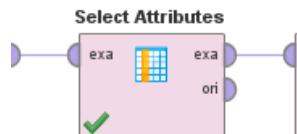
2) Pemanggilan Operator

Pada gambar 8 data yang berekstensi xls atau excel dimasukan kedalam rapidminer menggunakan operator read excel untuk diolah dalam proses data mining.



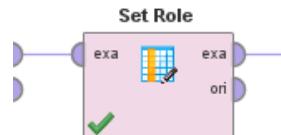
Gambar 8. Read Excel

Setelah menggunakan operator read excel laangkah selanjutnya menambah operator select atribut



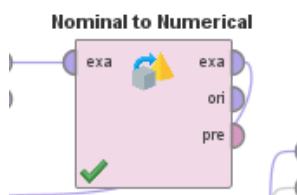
Gambar 9. Select Attributes

Operator selanjutnya yang akan dimasukan ke dalam kerja rapid miner adalah set role, set role digunakan untuk melabelkan data yang akan diolah menggunakan rapid miner.



Gambar10. Set Role

Setelah setrol operator yang digunakan selanjutnya yaitu nominal to numerical yang digunakan untuk mengubah tipe atribut nonnumerik menjadi tipe numerik.



Gambar 11. Nominal to Numerical

Gambar di bawah ini merupakan replace missing value yang berfungsi sebagai salah satu cara untuk menyelesaikan masalah missing value dengan mengisi nilai baru atau dikenal dengan metode imputasi.



Gambar 12. Replace Missing Values

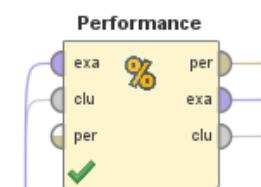
Setelah menetapkan beberapa operator di atas atau menampilkan data yang ingin diolah kemudian operator selanjutnya yang akan dipanggil yaitu

clustering k-means yang digunakan untuk mengelompokan atau membuat cluster.



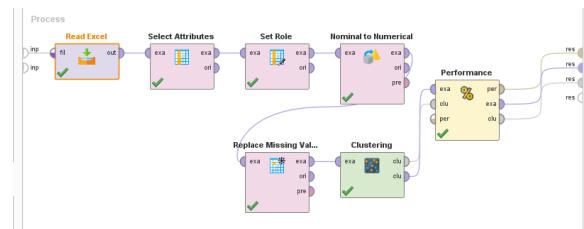
Gambar 1 Clustering

Gambar di bawah merupakan operator terakhir yaitu operator cluster distance performance untuk melihat hasil cluster dengan grafik.



Gambar 13. Performance

Tampilan keseluruhan operator pada lembar kerja rapid miner.



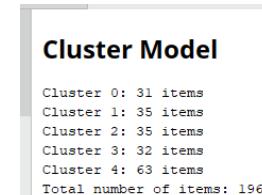
Gambar 2 Proses Eksperimen

### 3) Penampilan Hasil Menjalankan Rapid Miner

Setelah operator dimasukan kedalam lembar kerja rapid miner, Langkah selanjutnya adalah menjalankan projek yang telah kita susun operatornya. Setelah menjalankan projek maka akan tampil hasil dari rapid miner. Menampilkan gambar running atau menjalankan projek yang kita buat.



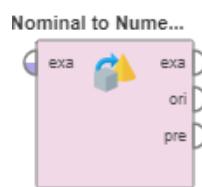
Gambar 3 Ikon Play



Gambar 4 Hasil Cluster

#### 4.1.3. Data Transformation

Operator *Nominal to Numeric* seperti pada Gambar 17, ini digunakan untuk mengubah tipe data atribut non-numerik menjadi numerik. Pada dataset, atribut "Status" akan diubah menjadi tipe numerik.

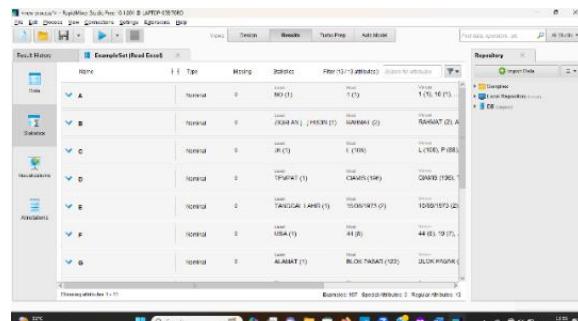


Gambar 17. Nominal to Numeric

Parameter pada operator *Nominal to Numerical* yang digunakan

Tabel 1. Parameter Nominal to Numerical

No	Parameter	Isi
1	<i>Attribute Filter Type</i>	<i>Subset</i>
2	<i>Selected Attributes</i>	<i>Status</i>
3	<i>Coding Type</i>	<i>Unique Integers</i>



Gambar 5 Hasil Statistik

#### 4.1.4. Data Mining

Tahapan data mining pada penelitian ini menggunakan metode clustering k-means. Untuk membantu proses penggalian data penelitian ini menggunakan perangkat lunak rapidminer.

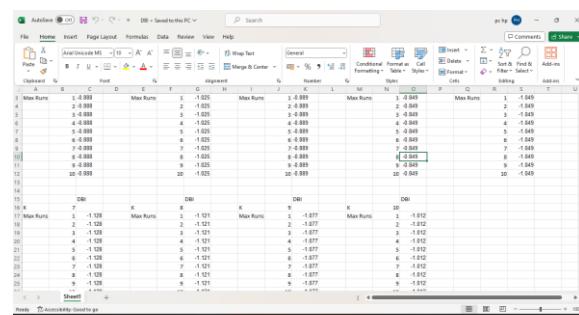
#### 4.2. Pembahasan

Program Perwalian Keluarga (PKH) dapat memanfaatkan perhitungan k-implies untuk memecah informasi dari jaringan yang tertindas. Informasi tersebut ditangani menggunakan ekskavator cepat, dan administrator "Eksekusi Jarak Banyak" digunakan untuk menilai eksekusi kelompok. Perhitungan k-implies adalah strategi pembagian yang mengelompokkan informasi berdasarkan kualitas dan kemiripan. Perhitungan ini memerlukan k info batasan dari klien dan mengumpulkan dataset yang berisi N objek ke dalam k kumpulan. Produk akhirnya adalah tingkat kemiripan yang tinggi di dalam pertemuan (intrakluster) dan rendahnya kemiripan antar pertemuan (intercluster). Beberapa batasan perhitungan K-Means, misalnya anggapan bahwa penyampaian setiap sifat berbentuk lingkaran, semua kredit mempunyai fluktiasi yang sama.

##### 4.2.1. Penentuan Nilai K Optimal

Penentuan jumlah cluster yang dilakukan menggunakan operator cluster distance performance untuk menghasilkan nilai perbandingan dari nilai davis bouldin. Hasil operasi tersebut menghasilkan nilai –

0.888 untuk cluster 2, untuk cluster 3 menghasilkan nilai -1.025, untuk cluster 4 menghasilkan nilai -0.889, untuk cluster 5 menghasilkan nilai -0.849, untuk cluster 6 menghasilkan nilai -1.049, untuk cluster 7 menghasilkan nilai -1.128, untuk cluster 8 menghasilkan nilai -1.121, untuk cluster 9 menghasilkan nilai -1.077, dan untuk cluster 10 menghasilkan nilai -1.012, dihasilkannya cluster terbaik yaitu pada cluster ke 5 dengan nilai DBI -0.849. Maka dapat ditentukan status sangat layak, layak, menengah, tidak layak, dan sangat tidak layak



Gambar 19. Hasil K Optimal

#### 4.2.2. Hasil Penelitian

Implementasi yang dilakukan adalah implementasi data mining untuk mengelompokkan Masyarakat yang kurang mampu menggunakan algoritma k-means dengan jumlah data yang diolah sebanyak 198.data ini dikategorikan menjadi Masyarakat yang mampu dan Masyarakat kurang mampu. PPKH dapat memanfaatkan penerapan Algoritma K-Means dalam implementasi masyarakat kurang mampu, algoritma K-Means dapat digunakan untuk melakukan analisis pada PPKH.

Dengan menggunakan algoritma ini, PPKH dapat mengelompokkan masyarakat yang tidak mampu berdasarkan kesamaan karakteristik. Selain itu, dengan menggunakan metode Algoritma K-Means, Dalam konteks penggunaan RapidMiner, PPKH dapat menggunakan operator "Cluster Distance Performance" untuk mengevaluasi performa klaster berbasis centroid, termasuk klaster yang dihasilkan oleh algoritma K-Means. Operator ini memberikan kriteria performa seperti rata-rata jarak dalam klaster dan indeks Davies-Bouldin.

#### 5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini menunjukkan bahwa algoritma K-Means efektif digunakan untuk menganalisis masyarakat kurang mampu penerima manfaat Program Keluarga Harapan (PKH), memungkinkan identifikasi tingkat kelayakan penerima manfaat. Saran untuk penelitian lebih lanjut meliputi penggunaan metode clustering alternatif, seperti algoritma hierarkis, untuk perbandingan dengan K-Means. Proses preprocessing data dapat ditingkatkan dengan pembuatan kamus kata baku, daftar stop word yang lebih lengkap, dan pendataan terhadap karakteristik khusus yang perlu dihilangkan.

Penggunaan algoritma K-Nearest Neighbour dan penelitian terhadap algoritma klasifikasi dan clustering yang berkelanjutan, seperti penggunaan K-Means untuk mengelompokkan data secara lebih spesifik, juga dapat menjadi fokus penelitian selanjutnya.

## DAFTAR PUSTAKA

- [1] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes," *Respati*, vol. 16, no. 1, p. 32, 2021, doi: 10.35842/jtir.v16i1.386.
- [2] M. C. Algorithm, "Teori K Means," no. 2010, pp. 7–24, 2012.
- [3] D. Hidayatullah, "Chemical Information and Modeling," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 8–24, 2018.
- [4] Wahyu Baskoro, "BAB II Tinjauan Pustaka BAB II TINJAUAN PUSTAKA 2.1. 1–64," *Gastron. ecuatoriana y Tur. local.*, vol. 1, no. 69, pp. 5–24, 2018.
- [5] S. Ningsih, "Penerapan Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan," *J. Sist. Inf.*, vol. 3, no. 2, pp. 73–82, 2022, [Online]. Available: <http://ojs.itb-ad.ac.id/index.php/JUSIN/article/view/1961%0Ahttps://ojs.itb-ad.ac.id/index.php/JUSIN/article/download/1961/435>
- [6] S. K. Gusti, "Analisis Sebaran Puskesmas Kesehatan Dengan Metode Fuzzy C- Means," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2012, [Online]. Available: [http://sir.stikom.edu/1062/5/BAB\\_II.pdf](http://sir.stikom.edu/1062/5/BAB_II.pdf)
- [7] Ansori, "Tahapan Data Mining," *Pap. Knowl. . Towar. a Media Hist. Doc.*, vol. 3, no. April, pp. 49–58, 2015.