

## Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan

Adellia Meiriza<sup>1</sup>, Edwar Ali<sup>2</sup>, Rahmiati<sup>3</sup>, Agustin<sup>4</sup>

adelliameiriza@gmail.com<sup>1\*</sup>, edwarali@sar.ac.id<sup>2</sup>, rahmiati@sar.ac.id<sup>3</sup>, agustin@sar.ac.id<sup>4</sup>

<sup>1,2,3,4</sup> STMIK AMIK Riau Pekanbaru

---

### Informasi Artikel

Diterima : 10 Apr 2023  
Direview : 19 Apr 2023  
Disetujui : 27 Apr 2023

---

### Kata Kunci

Bukan Penerima Upah (BPU), Clustering, K-Means, K-Medoid, Perbandingan.

---

### Abstrak

BPJS Ketenagakerjaan bertugas menyelenggarakan program jaminan sosial bagi para pekerja di Indonesia, seperti Jaminan Kecelakaan Kerja, Jaminan Hari Tua, Jaminan Pensiun, Jaminan Kematian, dan Jaminan Pemeliharaan Kesehatan. Pengelompokan program bukan penerima upah dapat menggunakan metode *clustering*. Dalam penelitian ini, peneliti membandingkan dua algoritma *clustering* yaitu *K-Means* dan *K-Medoids* untuk mengelompokkan program bukan penerima upah berdasarkan karakteristik yang dimiliki. Data yang digunakan dalam penelitian ini diperoleh dari BPJS Ketenagakerjaan cabang Pekanbaru. Pengelompokan dilakukan dengan menggunakan jumlah *cluster* yang sama untuk kedua algoritma yaitu  $K = 3$ . Hasil dari penelitian menunjukkan bahwa *K-Medoids* menghasilkan kelompok yang lebih stabil dan robust dibandingkan dengan *K-Means*. Hasil nilai DBI menunjukkan bahwa *K-Medoid* lebih baik dari *K-Means*. Hasil ini dapat dijadikan rekomendasi kepada pendaftar yang akan mengambil program BPJS Ketenagakerjaan selain itu penggunaan *K-Medoids* sebagai algoritma *clustering* lebih efektif dibandingkan *K-Means* untuk pengelompokan program bukan penerima upah.

---

### Keywords

Not a Wage Earner, Clustering, Comparison, K-Means, K-Medoid.

---

### Abstrak

*BPJS Ketenagakerjaan is tasked with administering social security programs for workers in Indonesia, such as Work Accident Insurance, Old Age Insurance, Pension Benefits, Death Insurance, and Healthcare Benefits. Grouping programs that are not wage earners can use the clustering method. In this study, researchers compared two clustering algorithms, namely K-Means and K-Medoids to classify programs that are not wage earners based on their characteristics. The data used in this study were obtained from the Pekanbaru branch of the Employment BPJS. Clustering was carried out using the same number of clusters for both algorithms, namely  $K = 3$ . The results of the study show that K-Medoids produces groups that are more stable and robust than K-Means. The DBI value results show that K-Medoid is better than K-Means. These results can be used as a recommendation for registrants who will take the BPJS Employment program. In addition, the use of K-Medoids as a clustering algorithm is more effective than K-Means for grouping programs that are not wage earners.*

## A. Pendahuluan

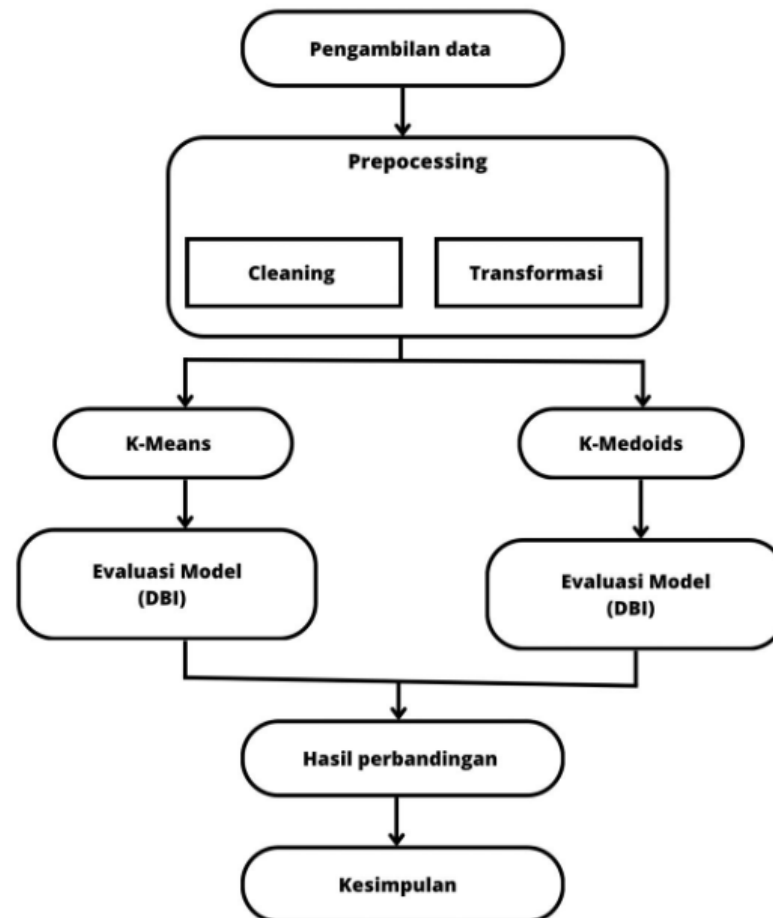
BPJS Ketenagakerjaan memberikan perlindungan sosial bagi tenaga kerja di Indonesia, terutama dalam hal kecelakaan kerja, kematian, pensiun, dan juga memberikan jaminan hari tua bagi peserta yang telah mencapai usia pensiun. BPJS Ketenagakerjaan merupakan badan hukum publik yang menyelenggarakan program jaminan sosial untuk tenaga kerja. Jaminan ini secara khusus untuk menanggulangi risiko yang terjadi dalam dunia pekerjaan dan meningkatkan kesejahteraan pekerja[1]. BPJS Ketenagakerjaan termasuk suatu jaminan sosial yang dimana sangat diperlukan terhadap masyarakat, terutama pada karyawan yang bekerja di perusahaan industri[2]. Terdapat 4 golongan penerima dan pemberi iuran pada BPJS Ketenagakerjaan yaitu pekerja penerima upah (PU), pekerja bukan penerima upah (BPU), pekerja jasa konstruksi (JAKON), dan pekerja migran Indonesia (PMI). Bukan Penerima Upah (BPU) adalah pekerja yang melakukan kegiatan atau usaha ekonomi secara mandiri untuk memperoleh penghasilan dari kegiatan atau usahanya[3]. Untuk memudahkan peserta baru yang akan mendaftar program BPJS Ketenagakerjaan maka penelitian ini akan mengelompokkan peserta untuk dijadikan rekomendasi kepada peserta baru menggunakan metode *clustering*.

*Clustering* adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi[4], beberapa contoh algoritma yaitu *K-Means*, *K-Medoid*, *Hierarchical Clustering* dan *Density-Based Spatial Clustering of Applications with Noise*. *K-Means* adalah *K-Means* adalah salah satu metode data *clustering non hierarki* yang prosesnya membagi data ke dalam bentuk satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang persis dikelompokkan ke dalam satu cluster yang serupa[5]. *K-Medoid* adalah teknik partisi klasik *clustering* yang mengelompokkan indeks informasi dari objek  $n$  ke dalam  $k$  kumpulan yang dapat diketahui kesimpulannya[6]. *K-Medoid* adalah metode *K-Medoid* menggunakan object sebagai perwakilan pusat *cluster* untuk setiap *cluster*, sedangkan *K-Means* menggunakan nilai rata-rata (*Mean*) sebagai pusat *cluster* [7]. Algoritma *K-means* dan *K-Medoid* akan digunakan dalam penelitian ini untuk mengelompokkan peserta BPJS Ketenagakerjaan pada golongan Bukan Penerima Upah.

Penelitian sebelumnya pernah dilakukan oleh [8] menggunakan algoritma *K-means* untuk mengelompokkan peserta BPJS Ketenagakerjaan pada tanjung perak. Selain itu juga pernah dilakukan oleh [9] menggunakan algoritma *K-Means* untuk mengelompokkan peserta BPJS Ketenagakerjaan cabang langkat. Berdasarkan penelitian sebelumnya pengelompokkan BPJS Ketenagakerjaan sudah pernah dilakukan menggunakan satu algoritma, perbedaan penelitian ini dengan penelitian sebelumnya penelitian ini akan membandingkan dua algoritma *clustering* yaitu *K-Means* dan *K-Medoid* dalam pengelompokkan peserta BPJS Ketenagakerjaan menggunakan rapid miner kemudian diambil nilai terbaik berdasarkan nilai *davies-bouldien index*.

## B. Metode Penelitian

Alur diagram penelitian disajikan pada Gambar 1



**Gambar 1.** Alur Diagram Penelitian

Berdasarkan Gambar 1, penelitian ini dimulai dengan pengambilan data pada BPJS Ketenagakerjaan cabang Pekanbaru, setelah itu dilakukan data *preprocessing*, *cleaning* digunakan untuk mengecek jika terdapat data yang *missing value* dan *transformasi* digunakan untuk meng-*konversi* data yang berupa teks menjadi angka. Setelah itu dilanjutkan dengan menguji data dengan dua algoritma *K-Means* dan *K-Medoid* menggunakan rapid miner, evaluasi kinerja model akan membandingkan nilai *davies bouldien index* terbaik.

### a. Pengambilan data

Tahap pertama dimulai dengan pengambilan data pada BPJS Ketenagakerjaan pada cabang Pekanbaru. Data yang didapatkan berjumlah 1000 dengan periode data dari Januari – Desember 2022.

### b. Data Preprocessing

Menyiapkan atau mengolah data sebelum menggunakannya untuk analisis atau aplikasi pembelajaran mesin dikenal sebagai data *preprocessing*. Tujuannya adalah untuk menjamin bahwa data tersebut akurat, berkualitas tinggi, dan sesuai

untuk pemrosesan di masa mendatang. Proses ini memerlukan beberapa langkah, termasuk pembersihan data, penghapusan duplikasi data, pengisian data, dan *konversi* format ke format yang lebih mudah diproses. Karena kualitas data yang buruk dapat menghasilkan hasil analisis dan prediksi yang tidak akurat, proses ini sangat penting. Tahapan *preprocessing* pada penelitian ini adalah sebagai berikut

1. Data *Cleaning* adalah Identifikasi, penilaian, dan koreksi data yang salah, tidak mencukupi, tidak relevan, atau tidak terstruktur dalam kumpulan data yang sangat besar. Data *cleaning* dilakukan untuk memastikan informasi yang digunakan untuk analisis atau pengambilan keputusan akurat, andal, dan konsisten.
2. Data *Transformation* adalah Proses mengubah teks menjadi bentuk numerik yang dapat ditangani oleh sistem data mining. hal ini biasanya dilakukan saat menggunakan metode penambangan data seperti *clustering* klasifikasi, atau analisis asosiasi untuk mengevaluasi data teks.

### c. Clustering

Suatu pendekatan analisis data yang disebut *clustering* atau pengelompokan membagi sejumlah besar data menjadi beberapa pengelompokan (*cluster*) berdasarkan kesamaan fiturnya. Menemukan pola atau struktur dalam data yang tidak langsung terlihat adalah tujuan dasar pengelompokan. Setiap potongan data yang digunakan untuk pengelompokan direpresentasikan sebagai titik atau vektor dalam ruang multidimensi, dengan setiap dimensi sesuai dengan atribut atau fitur data. Prosedur pengelompokan kemudian membagi titik-titik ini menjadi kelompok-kelompok sesuai dengan seberapa jauh mereka berada di ruang itu. Ada beberapa algoritma pengelompokan yang banyak digunakan, termasuk *K-means*, *K-medoid*, *Hierarchical clustering* dan *Density-based clustering*.

### d. K-Means

*K-means* adalah salah satu algoritma *clustering* atau pengelompokan data yang umum digunakan dalam analisis data. Tujuan dari algoritma ini adalah untuk membagi suatu set data menjadi  $k$  kelompok atau *cluster*, di mana setiap data masuk ke dalam kelompok yang paling dekat dengan data tersebut berdasarkan jarak *Euclidean*. Algoritma *k-means* bekerja dengan cara memilih  $k$  titik acak dari data sebagai pusat *cluster* awal. Kemudian setiap data diberi label sesuai dengan pusat *cluster* terdekat, dan pusat *cluster* dihitung ulang berdasarkan rata-rata data di setiap *cluster*. Proses ini diulang sampai tidak ada lagi perubahan dalam label *cluster*, atau hingga mencapai batas iterasi yang ditentukan.

*K-means* memiliki beberapa kelebihan, antara lain mudah diimplementasikan, skalabilitas yang baik pada data yang besar, dan hasil *clustering* yang relatif cepat dan efisien, tetapi *k-means* juga memiliki beberapa kelemahan, seperti sensitivitas terhadap inisialisasi awal pusat *cluster*, dan tidak dapat menangani *cluster* yang memiliki ukuran atau bentuk yang berbeda. Formula yang digunakan pada algoritma *K-Means* adalah sebagai berikut [10]

1. Pilih nilai  $k$  sebagai pusat *cluster* awal secara acak.
2. Setiap data dalam dataset dibagi kedalam beberapa kelompok  $k$  *cluster* antara setiap titik dan pusat *cluster* yang didapatkan

berdasarkan pada jarak *Euclidean* . Adapun rumus perhitungan *Euclidean Distance* ditunjukkan pada persamaan (1)

$$d(x,y) = ||x - y||^2 \quad (1)$$

d= jarak data ke pusat *cluster*

x= data ke pada atribut ke

y= titik pusat ke, pada atribut

3. Setiap pusat *cluster* dihitung ulang berdasarkan dari nilai rata - rata dalam *cluster* yang didapatkan.
4. Langkah 2 dan 3 ulangi sampai kelompok tersebut sesuai. Perulangan dapat didefinisikan secara berbeda tergantung pada implementasi, tetapi biasanya pada proses langkah 2 dan 3 dapat diulangi apabila kelompok *cluster* masih mengalami perubahan dan proses akan terhenti apabila tidak ada perubahan antar material dalam *cluster*

#### e. K-Medoid

*K-medoid* adalah suatu metode *clustering* yang menggunakan *medoid* sebagai pusat klaster. *Medoid* adalah data yang dianggap sebagai representasi atau pusat dari suatu klaster yang dihitung dengan mengambil rata-rata jarak atau nilai kesalahan dari semua data dalam klaster tersebut. Berbeda dengan *centroid*, *medoid* adalah salah satu data pada klaster yang sudah ada. Metode *k-medoid*, jumlah klaster yang dihasilkan harus ditentukan terlebih dahulu, dan titik *medoid* diinisialisasi secara acak. Selanjutnya, data pada klaster yang sudah ada akan diperbarui dengan memperbarui *medoid* yang mewakili setiap klaster. Iterasi ini dilakukan sampai *medoid* tidak lagi berubah atau sampai batas iterasi tertentu tercapai.

Metode *k-medoid* sering digunakan dalam berbagai aplikasi, seperti analisis data pelanggan, pengelompokan data genetik, dan segmentasi data pasar. Kelebihan dari metode *k-medoid* adalah lebih stabil terhadap data *outlier* atau pencilan daripada metode *k-means*, dan lebih fleksibel dalam menangani berbagai jenis data. Formula *K-Medoid* adalah sebagai berikut [10]

1. Inisialisasi pusat *cluster* sebanyak jumlah cluster (k) .
2. Setiap data atau objek dialokasikan ke cluster terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan rumus persamaan (1).
3. Pilih objek pada masing - masing *cluster* secara acak sebagai kandidat *medoid* baru.
4. Hitung jarak setiap objek yang terdapat pada masing - masing *cluster* dengan calon *medoid* baru.
5. Hitung total simpangan (S) dengan menghitung nilai total jarak baru – total jarak lama. Jika didapatkan  $S < 0$ , tukarlah objek dengan data *cluster* untuk membuat sekumpulan k objek baru sebagai *medoid*
6. Ulangi langkah 3 sampai dengan 5 hingga tidak terjadi perubahan *medoid* , sehingga di peroleh *cluster* serta anggota *cluster* masing - masing.

#### f. Rapid Miner

*Rapid Miner* adalah sebuah *platform* analisis data yang populer, berbasis *open-source*, dan mudah digunakan. *Platform* ini menyediakan berbagai macam alat dan teknik analisis data, seperti pemrosesan data, visualisasi data, pembuatan model, dan penambangan data. *RapidMiner* juga mendukung integrasi dengan berbagai sumber data, termasuk database, file teks, dan berbagai jenis data lainnya. Pengguna dapat membangun alur kerja analisis data dengan menarik dan menjatuhkan berbagai modul yang tersedia. Setiap modul memiliki fungsi spesifik dan dapat diatur dengan berbagai parameter untuk memproses data dan menghasilkan *output* yang diinginkan, misalnya modul pemrosesan data dapat digunakan untuk menghapus data yang tidak relevan, sementara modul pembuatan model dapat digunakan untuk membangun model prediksi berdasarkan data yang ada. Selain itu *RapidMiner* juga mendukung pengembangan ekstensi dan integrasi dengan berbagai *platform* analisis data lainnya, sehingga memungkinkan pengguna untuk memperluas fungsionalitas *platform* sesuai dengan kebutuhan mereka. *Platform* ini digunakan dalam berbagai macam industri, termasuk keuangan, kesehatan, pemasaran, dan penelitian akademis.

#### g. Evaluasi Model Davies Bouldien Index

*Davies Bouldin Index* (DBI) adalah salah satu metode evaluasi yang digunakan untuk mengukur kualitas pemodelan *clustering* (pengelompokan) pada data yang diberikan. Metode ini dikembangkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 dan menjadi salah satu metode evaluasi *clustering* yang paling banyak digunakan hingga saat ini. Tujuan utama dari evaluasi *clustering* adalah untuk mengukur seberapa baik data yang diberikan dapat dikelompokkan menjadi beberapa kelompok yang berbeda berdasarkan karakteristik yang dimiliki oleh masing-masing data. Dalam hal ini, DBI digunakan untuk mengukur seberapa baik setiap kelompok dalam *clustering* dipisahkan dari kelompok lainnya berdasarkan jarak dan karakteristik data.

DBI dihitung dengan membandingkan antara kelompok dalam (*intra-cluster*) dan kelompok luar (*inter-cluster*) dengan menggunakan rasio jarak antar kelompok (*inter-cluster distance*) dan rasio jarak antara data dalam kelompok (*intra-cluster distance*). DBI menghitung rata-rata dari seluruh kelompok dalam *clustering* dan memberikan nilai skor untuk setiap kelompok yang menunjukkan seberapa baik kelompok tersebut terisolasi dari kelompok lain.

### C. Hasil dan Pembahasan

#### a. Pengambilan dan Preprocessing Data

Data diambil langsung pada BPJS Ketenagakerjaan cabang Pekanbaru. Hasil pengambilan data disajikan pada Tabel 1

**Tabel 1.** Pengambilan Data

Nama	Usia	Pekerjaan	Upah	Program
A	42	PEDAGANG	1000000	JKK, JKM, JHT
B	54	SOPIR	2017292	JKK, JKM, JHT
C	28	KULI BANGUNAN	2950000	JKK, JKM, JHT

D	48	PENJAHIT	1000000	JKK, JKM
E	24	PETANI	1000000	JKK, JKM

Berdasarkan Tabel 1 data memiliki 5 variabel yaitu nama, usia, pekerjaan, upah dan program. Tahap selanjutnya melakukan preprocessing data, tahap yang dilakukan adalah sebagai berikut.

1. *Data Cleaning*

Tahap data *cleaning* bertujuan untuk memeriksa jika terdapat

2. *Data Transformasi*

Tahap data *transformasi* akan meng-*konversi* data yang berupa teks pada variabel pekerjaan dan program menjadi angka. Tahap *konversi* teks menjadi angka pada kolom pekerjaan dapat dilihat pada Tabel 2

**Tabel 2.** Konversi Data Pekerjaan

Pekerjaan	Konversi
PEDAGANG	1
GOJEK	2
PETANI	3
ASISTEN RUMAH TANGGA	4
KURIR	5
PENJAHIT	6
SOPIR	7
JURU MASAK	8
KULI BANGUNAN	9
JURU PARKIR	10
MONTIR	11
DOKTER	12
PEMBANTU	13

Berdasarkan Tabel 2 data pekerjaan yang awalnya teks telah di-*konversi* kan menjadi angka. Selanjutnya data yang berupa teks pada variabel program juga akan di konversikan menjadi teks berdasarkan kode ASCII, proses tersebut dapat dilihat pada Tabel 3

**Tabel 3.** Konversi Data Program menjadi ASCII

Program	Konversi	Total
JKK	$74 + 75 + 75$	224
JKM	$74 + 75 + 77$	226
JHT	$74 + 72 + 84$	230
JKK, JKM	$224 + 226$	450
JKK, JKM, JHT	$224 + 226 + 230$	680

Berdasarkan Tabel 3 data program yang awalnya teks telah di-*konversi* kan menjadi angka berdasarkan kode ASCII dari tiap-tiap kata yang kemudian dijumlahkan. Setelah data tersebut di-*konversi* menjadi kode ASCII kemudian akan diproses menggunakan algoritma *clustering*.

**b. K-Means**

Pengujian pertama dilakukan dengan algoritma *K-Means*, diambil 10 data sample untuk dilakukan perhitungan manual, 10 data tersebut disajikan pada Tabel 4.

**Tabel 4.** Data Sampel

No	Nama	Umur	Pekerjaan	Upah	Program
1	A	29	11	1350000	680
2	B	51	1	2950000	680
3	C	34	2	1600000	680
4	D	54	1	2017292	680
5	E	47	3	3450000	680
6	F	39	8	2950000	680
7	G	32	1	2000000	450
8	H	44	4	1000000	450
9	I	54	1	2950000	450
10	J	49	7	2950000	450

Langkah-Langkah perhitungan manual K-Means adalah sebagai berikut

1. Tentukan pusat *cluster* yaitu  $K = 3$ , terdapat pada data ke 2, 4 dan 8. Penentuan data tersebut disajikan pada Tabel 5.

**Tabel 5.** Penentuan Pusat *Cluster*

s	No	Nama	Umur	Pekerjaan	Upah	Program
1	2	B	51	1	2950000	680
2	4	D	54	1	2017292	680
3	8	H	44	4	1000000	450

2. Hitung pusat kelompok dari data masing-masing cluster menggunakan persamaan berikut

$$\sqrt{\frac{((Cluster\ 1 - data\ kolom\ umur) + (Cluster\ 1 - data\ kolom\ pekerjaan))}{(Cluster\ 1 - data\ kolom\ upah) (Cluster\ 1 - data\ kolom\ program)}} \quad (2)$$

$$\sqrt{\frac{((Cluster\ 2 - data\ kolom\ umur) + (Cluster\ 2 - data\ kolom\ pekerjaan))}{(Cluster\ 2 - data\ kolom\ upah) (Cluster\ 2 - data\ kolom\ program)}} \quad (3)$$

$$\sqrt{\frac{((Cluster\ 3 - data\ kolom\ umur) + (Cluster\ 3 - data\ kolom\ pekerjaan))}{(Cluster\ 3 - data\ kolom\ upah) (Cluster\ 3 - data\ kolom\ program)}} \quad (4)$$

Persamaan (2), (3), (4) data dihitung berdasarkan *cluster* kemudian dijumlahkan dengan masing-masing data pada tiap variabel umur, pekerjaan, upah dan program, kemudian ditentukan jarak terdekat antar *cluster*. Hasil perhitungan untuk Iterasi 1 disajikan pada Tabel 6

**Tabel 6.** Iterasi 1

No	C1	C2	C3	Jarak terdekat	Cluster
1	1600000	667292	402900	402900	C3
2	0	932708	2002900	0	C1
3	1350000	417292	652900	417292	C2
4	932708	0	1070192	0	C2



5	500000	1432708	2502900	500000	C1
6	13.89244	932708	2002900	13.89244	C1
7	1002900	70192.01	1000000	70192.01	C2
8	2002900	1070192	0	0	C3
9	52903	985608	1950000	52903	C1
10	52906.32	985608	1950000	52906.32	C1

Berdasarkan Tabel 6, *Cluster 1* terdapat pada data 2, 5, 6, 9 dan 10. *Cluster 2* terdapat pada data 3, 4, dan 7, dan *Cluster 3* terdapat pada data 1 dan 8.

3. Tentukan *cluster* baru harus berdasarkan *average* dari masing-masing *cluster* menggunakan persamaan berikut

$$\text{Cluster baru 1} = \frac{\text{data ke 2,5,6,9,10 pada tiap - tiap variabel}}{5} \quad (5)$$

$$\text{Cluster baru 2} = \frac{\text{data ke 3,4,7 pada tiap - tiap variabel}}{3} \quad (6)$$

$$\text{Cluster baru 2} = \frac{\text{data ke 1,8 pada tiap - tiap variabel}}{2} \quad (7)$$

Persamaan (5), (6), (7) digunakan untuk menentukan *cluster* baru sebagai penentuan *cluster* pada iterasi 2. Hasil penentuan *cluster* baru tersebut disajikan pada Tabel 7

Tabel 7. Penentuan *Cluster* Baru *K-Means*

Cluster	Umur	Pekerjaan	Upah	Program
baru 1	48	4	3050000	588
baru 2	40	1.33	1872430	603
baru 3	36.5	7.5	1175000	565

4. Selanjutnya hitung iterasi kedua, menggunakan persamaan berikut

$$\sqrt{\frac{((\text{Cluster baru 1} - \text{data kolom umur}) + (\text{Cluster baru 1} - \text{data kolom pekerjaan}))}{(\text{Cluster baru 1} - \text{data kolom upah}) (\text{Cluster baru 1} - \text{data kolom program})}} \quad (8)$$

$$\sqrt{\frac{((\text{Cluster baru 2} - \text{data kolom umur}) + (\text{Cluster baru 2} - \text{data kolom pekerjaan}))}{(\text{Cluster baru 2} - \text{data kolom upah}) (\text{Cluster baru 2} - \text{data kolom program})}} \quad (9)$$

$$\sqrt{\frac{((\text{Cluster baru 3} - \text{data kolom umur}) + (\text{Cluster baru 3} - \text{data kolom pekerjaan}))}{(\text{Cluster baru 3} - \text{data kolom upah}) (\text{Cluster baru 3} - \text{data kolom program})}} \quad (10)$$

Persamaan (8), (9), (10) data dihitung berdasarkan *cluster* baru kemudian dijumlahkan dengan masing-masing data pada tiap variabel umur, pekerjaan, upah dan program, kemudian ditentukan jarak terdekat antar *cluster*. Hasil perhitungan untuk Iterasi 1 disajikan pada Tabel 8

**Tabel 8.** Iterasi 2 *K-Means*

No	C1	C2	C3	Jarak terdekat	Cluster
1	1708464	528308.4	188225	188225	C3
2	108464	1083447	1788225	108464	C1
3	1458464	278308.4	438225	278308.4	C2
4	1041172	150739.1	855517	150739.1	C2
5	408464	1583447	2288225	408464	C1
6	108464	1083447	1788225	108464	C1
7	1069044	151080.4	838225	151080.4	C2
8	2069044	895941.8	188225	188225	C3
9	119044	1101080	1788225	119044	C1
10	119044	1101080	1788225	119044	C1

Berdasarkan Tabel 8 setelah dilakukan perhitungan pada Iterasi 2, data yang terdapat pada masing-masing *cluster* akan dibandingkan dengan Iterasi 1. Perbandingan data pada tiap *cluster* dari masing-masing Iterasi disajikan pada Tabel 9

**Tabel 9.** Perbandingan *cluster* Iterasi 1 dan 2

No	Jarak Terdekat Iterasi 1	Cluster Iterasi 1	Jarak Terdekat Iterasi 2	Cluster Iterasi 2	Kesimpulan
1	402900	C3	188225	C3	Tidak Berubah
2	0	C1	108464	C1	Tidak Berubah
3	417292	C2	278308.4	C2	Tidak Berubah
4	0	C2	150739.1	C2	Tidak Berubah
5	500000	C1	408464	C1	Tidak Berubah
6	13.89244	C1	108464	C1	Tidak Berubah
7	70192.01	C2	151080.4	C2	Tidak Berubah
8	0	C3	188225	C3	Tidak Berubah
9	52903	C1	119044	C1	Tidak Berubah
10	52906.32	C1	119044	C1	Tidak Berubah

Berdasarkan Tabel 9 *cluster* pada masing-masing data tidak berubah, sehingga proses perhitungan dihentikan.

### c. K-Medoid

Pengujian kedua dilakukan menggunakan algoritma *K-Medoid*, ambil 10 data sample untuk dilakukan perhitungan manual, 10 data tersebut disajikan pada Tabel 4. Langkah-langkah perhitungan manual adalah sebagai berikut

1. Tentukan pusat *cluster* yaitu  $K = 3$ , terdapat pada data ke 2, 4 dan 8. Penentuan data tersebut disajikan pada Tabel 5

2. Hitung pusat kelompok dari data masing-masing *cluster* menggunakan persamaan (2), (3) dan (4)
3. Hasil dari perhitungan untuk iterasi 1 disajikan pada Tabel 6
4. Jumlahkan data pada variabel jarak terdekat yang terdapat pada iterasi 1 menggunakan persamaan berikut

$Sum(\text{variabel jarak terdekat pada iterasi 1})$  (11)  
 Hasil penjumlahan tersebut mendapatkan nilai 1496207.232

$$402900.0004 + 0 + 417292.0005 + 0 + 500000 + 13.89244399 + 70192.01399 + 0 + 52903 + 52906.32456 = 1496207.232$$

5. Kemudian tentukan *cluster* baru, peneliti menetapkan untuk *cluster* baru terdapat pada data 7, 8 dan 10. *Cluster* baru tersebut disajikan pada Tabel 10

**Tabel 10.** Penentuan *Cluster* Baru *K-Medoid*

Cluster baru	Umur	Pekerjaan	Upah	Program
baru 1	32	1	2000000	450
baru 2	44	4	1000000	450
baru 3	49	7	2950000	450

6. Setelah itu hitung kembali untuk iterasi 2 menggunakan persamaan (8), (9) dan (10). Hasil perhitungan tersebut disajikan pada Tabel 11

**Tabel 11.** Iterasi 2 *K-Medoid*

No	C1	C2	C3	Jarak terdekat	Cluster
1	702900.0001	402900.0004	1652900	402900.0004	C2
2	1002900	2002900	52906.32456	52906.32456	C3
3	452900	652900.0001	1402900	452900	C1
4	70192.01399	1070192	985608	70192.01399	C1
5	1502900	2502900	552900	552900	C3
6	1002900	2002900	52910.04988	52910.04988	C3
7	0	1000000	950000.0002	0	C1
8	1000000	0	1950000	0	C2
9	950000.0003	1950000	7.810249676	7.810249676	C3
10	950000.0002	1950000	0	0	C3

7. Jumlahkan kembali data pada variabel jarak terdekat yang terdapat pada iterasi 2 menggunakan persamaan berikut

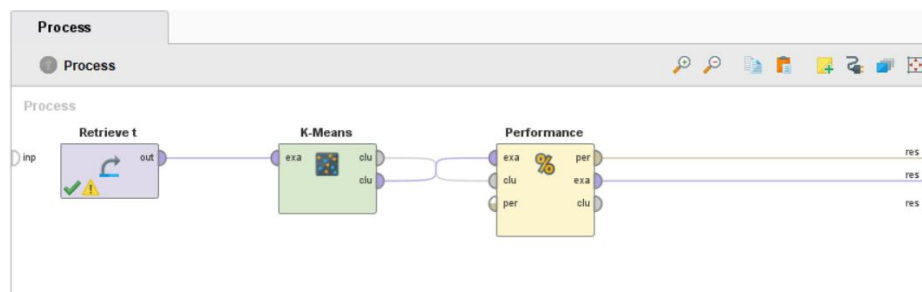
$Sum(\text{variabel jarak terdekat pada iterasi 2})$  (12)

Hasil penjumlahan tersebut mendapatkan nilai 1584716.199  
 $402900.0004 + 52906.32456 + 452900 + 70192.01399 + 552900 + 52910.04988 + 0 + 0 + 7.810249676 + 0 = 1584716.199$

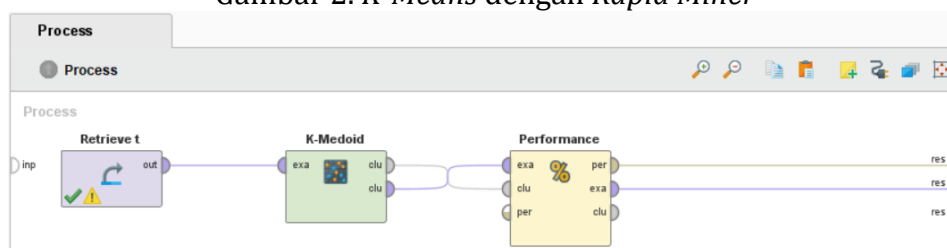
Berdasarkan perhitungan Iterasi 1 dan Iterasi 2 pada variabel jarak terdekat disimpulkan bahwa jarak terdekat pada Iterasi 2 lebih besar dari Iterasi 1 yaitu  $1584716.199 > 1496207.232$  sehingga perhitungan dihentikan

#### d. Evaluasi Model (Davies-Bouldin Index)

Mengukur kinerja dari model *K-Means* dan *K-Medoid* menggunakan bantuan *rapid miner* dengan melihat nilai DBI. DBI adalah salah satu metode untuk mengevaluasi kualitas *clustering* pada suatu model, DBI mengukur kepadatan dan jarak antar *cluster* pada dataset yang dikelompokkan. DBI mencari rasio antara rata-rata kepadatan *cluster* dan jarak antara *cluster*, semakin rendah DBI, semakin baik model klastering yang dihasilkan. Sebaliknya, semakin tinggi nilai DBI, semakin buruk klastering yang dihasilkan. Penggunaan Rapid Miner untuk mendapatkan nilai DBI pada algoritma *K-Means* dan *K-Medoid* disajikan pada Gambar 2 dan 3



Gambar 2. *K-Means* dengan *Rapid Miner*



Gambar 3. *K-Medoid* dengan *Rapid Miner*

Hasil evaluasi untuk perbandingan model dari *K-Means* dan *K-Medoid* disajikan pada Tabel 12

Tabel 12. Perbandingan Nilai DBI

Algoritma	DBI
<b>K-Means</b>	-0.27
<b>K-Medoid</b>	-51.9

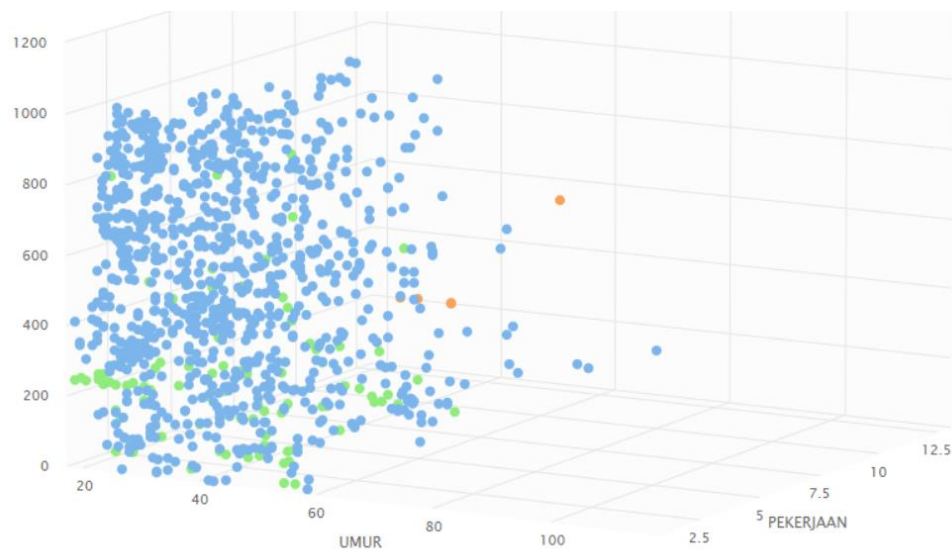
Berdasarkan Tabel 12 perbandingan nilai DBI dari kedua algoritma dapat disimpulkan bahwa *K-Medoid* mendapatkan nilai yang lebih kecil dari *K-Means* yang berarti *K-Medoid* memiliki performa yang lebih bagus dalam penelitian ini

#### e. Analisa Hasil

Setelah didapatkan hasil perbandingan dari kedua algoritma, selanjutnya akan dilakukan analisa hasil dari perhitungan manual dan penggunaan *tools rapid miner*. Hasil dari analisa dari kedua algoritma adalah sebagai berikut

### 1. K-Means

Analisa hasil dari perhitungan manual pada algoritma *K-Means* disimpulkan bahwa pada *cluster* 1 dengan data "Umur = 48", "Pekerjaan = 4", "Upah = 3050000" cenderung mengambil 3 program yaitu JKK, JKM dan JHT. Selanjutnya *cluster* 2 dengan data "Umur = 40", "Pekerjaan = 1", "Upah = 1872430" cenderung mengambil 3 prgoram yaitu JKK, JKM dan JHT. Selanjutnya *cluster* 3 dengan data "Umur = 37", "Pekerjaan = 8", "Upah = 1175000" cenderung mengambil 2 program yaitu JKK dan JKM. Penyebaran data menggunakan algoritma *K-Means* menggunakan bantuan *Rapid Miner* disajikan pada Gambar 4

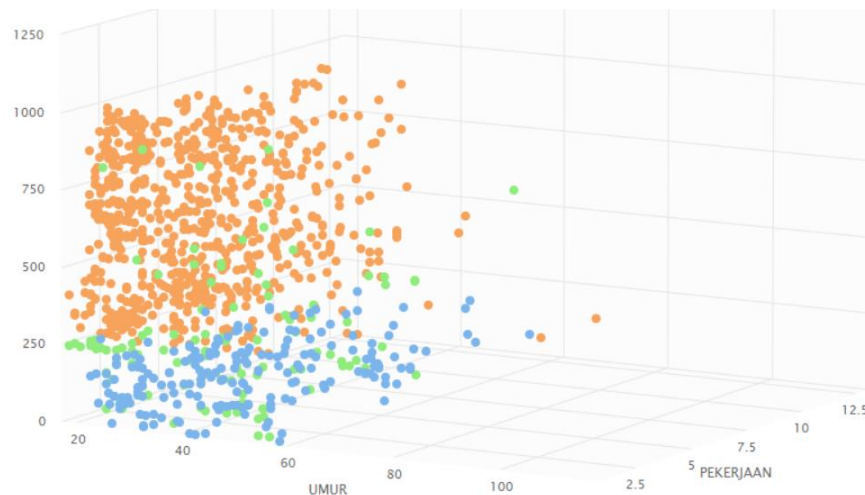


Gambar 4. Penyebaran Data pada *K-Means*

Berdasarkan Gambar 4 total dari penyebaran data menggunakan *K-Means* yaitu *cluster* 1 berjumlah 905 data , *cluster* 2 berjumlah 5 data dan *cluster* 3 berjumlah 90 data.

### 2. K-Medoid

Analisa hasil dari perhitungan manual pada algoritma *K-Medoid* disimpulkan bahwa pada *cluster* 1 dengan data "Umur = 32", "Pekerjaan = 1", "Upah = 2000000" cenderung mengambil 2 program yaitu JKK dan JKM. Selanjutnya *cluster* 2 dengan data "Umur = 44", "Pekerjaan = 4", "Upah = 1000000" cenderung mengambil 2 prgoram yaitu JKK dan JKM. Selanjutnya *cluster* 3 dengan data "Umur = 49", "Pekerjaan = 7", "Upah = 2950000" cenderung mengambil 2 program yaitu JKK dan JKM. Penyebaran data menggunakan algoritma *K-Medoid* menggunakan bantuan *Rapid Miner* disajikan pada Gambar 5



Gambar 5. Penyebaran Data pada *K-Medoid*

Berdasarkan Gambar 5 total dari penyebaran data menggunakan *K-Medoid* yaitu *cluster* 1 berjumlah 701 data, *cluster* 2 berjumlah 103 data dan *cluster* 3 berjumlah 196 data.

#### D. Simpulan

*K-Means* dan *K-Medoids* adalah dua algoritma yang digunakan untuk pengelompokan data tanpa label, dalam konteks pengelompokan program bukan penerima upah, kedua algoritma ini dapat digunakan untuk mengelompokkan program-program tersebut berdasarkan fitur-fitur tertentu, seperti umur, pekerjaan, upah dan program. *K-Means* adalah algoritma yang bekerja dengan menghitung jarak antara setiap data point dengan *centroid* terdekat, kemudian mengelompokkan data point ke dalam kelompok yang memiliki *centroid* terdekat tersebut, sedangkan *K-Medoids* adalah algoritma yang mirip dengan *K-Means*, namun menggunakan *medoid* sebagai pusat kelompok, yang merupakan data point yang paling *representatif* dalam kelompok tersebut. Sehingga, dalam memilih antara kedua algoritma ini, perlu dipertimbangkan ketersediaan sumber daya dan tujuan dari pengelompokan data tersebut, jika ketersediaan sumber daya tidak menjadi masalah dan tujuan pengelompokan adalah untuk mencapai hasil yang lebih akurat, maka *K-Medoids* dapat menjadi pilihan yang lebih baik. Namun, jika waktu komputasi menjadi pertimbangan utama, *K-Means* dapat menjadi pilihan yang lebih cepat dan cukup akurat untuk pengelompokan data tanpa label. Peneliti memberikan saran kepada peneliti selanjutnya untuk menambahkan jumlah dataset serta mencoba dengan variasi parameter sehingga dapat memengaruhi kinerja algoritma *clustering*.

#### E. Ucapan Terima Kasih

Terima kasih kepada dosen STMIK AMIK Riau yang telah membantu dalam penelitian ini.

**F. Referensi**

- [1] F. Azani, L. Febriyanti, I. Syahdilla, and F. P. Gurning, "Analisis sistem klaim jaminan hari tua dan perhitungannya pada BPJS Ketenagakerjaan cabang Medan Utara," *J. Ilm. Kesehat.*, vol. 1, no. 2, pp. 53–58, 2022.
- [2] A. Wulandari and Y. Oktavia, "Analisis Kebijakan Bpjs Ketenagakerjaan Terhadap Karyawan Di Suatu Perusahaan(Kota Batam)," *J. Ilmu Sos. Dan Ilmu Polit.*, vol. 4, no. 1, pp. 852–866, 2022, doi: 10.56552/jisipol.v4i1.87.
- [3] F. Baskoro, "Analisis Strategi Pemasaran Dalam Meningkatkan Jumlah Kepesertaan Bukan Penerima Upah Pada BPJS Ketenagakerjaan kantor Cabang Perintis Ponorogo," *etheses.iainponorogo.ac.id*, 2021.
- [4] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia 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, doi: 10.33060/jik/2020/vol9.iss2.181.
- [5] K. S. H. Kusuma Al Atros, A. R. Padri, O. Nurdiawan, A. Faqih, and S. Anwar, "Model Klasifikasi Analisis Kepuasan Pengguna Perpustakaan Online Menggunakan K-Means dan Decission Tree," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 323, 2021, doi: 10.30865/jurikom.v8i6.3680.
- [6] Rizal, H. A. Aidilof, Mukhlis, and K. Nur, "Karakteristik Peserta Pedesaan Dan Perkotaan 2 Pake," *J. Tekno Kompak*, vol. 16, no. 2, pp. 85–97, 2022.
- [7] R. Alawiyah, A. Mahmudi, and S. Achmadi, "Penerapan Metode K-Medoid Pada Analisis Respon Emosi Marah Wanita Pendalungan Berbasis Web," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 1, pp. 126–133, 2022, doi: 10.36040/jati.v6i1.4508.
- [8] A. F. Ayutrisula and A. Fanani, "Customer Profiling dengan Menggunakan Metode K-Means Euclidean Distance di BPJS Ketenagakerjaan Tanjung Perak," *J. Mhs. Mat. Algebr.*, vol. 1, no. 1, pp. 157–168, 2020.
- [9] M. Murniati, "Pengelompokan Data Jaminan Kematian Berdasarkan Kecamatan Dengan Metode Clustering," *J. Tek. Komputer, Agroteknologi Dan Sains*, vol. 1, no. 1, pp. 103–113, 2022, doi: 10.56248/marostek.v1i1.16.
- [10] I. KaAmila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.