

KLASTERISASI ANGKATAN KERJA DI INDONESIA BERDASARKAN USIA MENGGUNAKAN METODE ALGORITMA K-MEANS

Ririn Restu Aria ¹

Program Studi Sistem Informasi ¹
Universitas Bina Sarana Informatika ¹
www.bsi.ac.id ¹
ririn.rra@bsi.ac.id ¹



Abstract—The concept of the population is divided into two groups, namely the working age population and the population not working age. Indonesia, which has 34 provinces, has an unequal distribution of labor force due to the level of economic growth that is still not evenly distributed in several sectors. Labor is the most important and influential element in managing and controlling the economic system. In this study the method used in the grouping of provinces was based on the workforce in 34 provinces using the K-Means algorithm. The purpose of grouping data is done to get a province grouping that has a workforce in Indonesia by grouping / clustering into 3 groups based on age groups using the K-Means algorithm. Based on the calculations, the results of cluster 0 were 6 provinces, cluster 1 as many as 3 provinces and cluster 2 were 25 provinces. The K-Means algorithm can be used to understand the workforce problems and make it easier to describe the characteristics or characteristics of each group. Based on these results, the local government can give more attention to the regions with the smallest workforce such as the Province of Central Sulawesi, East Kalimantan, Jambi so that economic growth in various sectors can be increased so that the welfare of the workforce, especially in terms of work in the field of work can be easily obtained.

Keywords: *k-means, labor force, rapidminer.*

Abstrak—Secara konsep penduduk terbagi menjadi dua kelompok yaitu penduduk usia kerja dan penduduk bukan usia kerja. Indonesia yang memiliki 34 provinsi memiliki penyebaran angkatan kerja yang tidak merata disebabkan oleh tingkat pertumbuhan ekonomi yang masih belum merata di beberapa sektor. Tenaga kerja merupakan unsur yang paling penting dan berpengaruh dalam mengelola dan mengendalikan sistem ekonomi. Dalam penelitian ini metode yang digunakan dalam pengelompokan provinsi berdasarkan angkatan kerja di 34 provinsi menggunakan algoritma K-Means. Tujuan dari pengelompokan data yang dilakukan untuk mendapatkan pengelompokan provinsi yang memiliki angkatan kerja di Indonesia dengan mengelompokkan atau *clustering* menjadi 3 kelompok berdasarkan kelompok usia menggunakan algoritma K-Means. Berdasarkan perhitungan yang dilakukan diperoleh hasil *cluster* 0 sebanyak 6 provinsi, *cluster* 1 sebanyak 3 provinsi dan *cluster* 2 sebanyak 25 provinsi. Algoritma K-Means bisa digunakan untuk memahami permasalahan angkatan kerja dan memudahkan untuk mendiskripsikan sifat – sifat atau karakteristik dari masing – masing kelompok. Berdasarkan hasil tersebut maka pemerintah daerah bisa lebih memberikan perhatian lebih terhadap daerah dengan angkatan kerja terkecil seperti provinsi Sulawesi Tengah, Kalimantan Timur, Jambi agar pertumbuhan ekonomi dalam berbagai sektor bisa ditngkatkan sehingga kesejahteraan bagi angkatan kerja terutama dalam hal bidang pekerjaan bisa mudah diperoleh.

Kata kunci: *k-means, angkatan kerja, rapidminer.*

PENDAHULUAN

Indonesia merupakan negara dengan 34 provinsi yang memiliki penduduk dengan usia kerja yang tercatat pada bulan Agustus 2023 sebanyak 212,59 juta orang, jumlah ini naik sebesar 3,17 juta orang atau sekitar 1,51 persen

jika dibandingkan dengan keadaan angkatan kerja yang ada di bulan Agustus 2022 dimana penduduk usia kerja itu terdiri dari angkatan kerja dan bukan angkatan kerja. Dari data 212,59 juta orang yang merupakan angkatan kerja di Indonesia untuk bulan Agustus 2023 mencapai 147,71 juta orang dan yang bukan angkatan kerja sebanyak 64,88

juta orang. Untuk jumlah angkatan kerja bertambah sebanyak 3,99 juta (2,77 persen) dan bukan angkatan kerja berkurang sekitar 0,82 juta orang (1,24 persen) dibandingkan jumlah yang ada di bulan Agustus 2022 (BPS, 2023).

Angkatan kerja di Indonesia merupakan penduduk yang berusia 15 tahun keatas yang aktif melakukan kegiatan ekonomi yaitu penduduk yang bekerja yang memiliki pekerjaan tetapi sementara ini tidak bekerja dan tidak memiliki pekerjaan sama sekali namun aktif dalam mencari pekerjaan (Fajriati & Syafriandi, 2022); (Andriatno et al., 2021). Angkatan kerja yang dapat dibagi berdasarkan usia sangat memiliki dampak yang sangat erat terkait dengan masalah kependudukan yang akan berhubungan dengan ketenagakerjaan (Saputra et al., 2019). Permasalahan yang sering terjadi dalam angkatan kerja adalah kurangnya kebutuhan akan kompetensi tenaga kerja berdasarkan masing-masing provinsi.

Penelitian ini mengelompokkan provinsi yang ada di Indonesia berdasarkan jumlah angkatan kerja berdasarkan kelompok usia yang dimiliki penduduknya yang ada di perkotaan dan pedesaan. Metode yang akan digunakan dalam proses penelitian ini menggunakan metode K-Means yang merupakan salah satu metode *clustering* yang ada di Data Mining. Data mining sering disebut KDD (*knowledge discovery in database*) untuk mengumpulkan, mengolah dengan ukuran besar untuk menemukan aturan atau pola yang disimpan dalam sebuah database (Rohmah et al., 2021); (Kurnia et al., 2021). Data mining digunakan untuk mengekstraksi informasi secara *predictive* yang tidak diketahui disebuah database (Arifandi et al., 2021). *Clustering* adalah salah satu teknik dalam data mining yang efektif untuk mengelompokkan data (Akil, 2021). *Clustering* juga merupakan bagian dari bidang penelitian dan analisis (Gustientiedina et al., 2019). Dalam menggunakan metode K-Means dibutuhkan *clustering* yang merupakan proses dalam mengelompokkan data menjadi beberapa *cluster* atau kelompok sehingga data mudah untuk diproses (Handayanna, 2023). K-Means dalam perhitungannya melakukan dua proses yaitu proses mendeteksi lokasi pusat *cluster* dan mengelompokkan anggota kedalam tiap *cluster* dan terakhir akan membentuk *cluster* baru (Rahmati & Wijayanto, 2021).

Terdapat penelitian terdahulu terkait algoritma K-Means yaitu penelitian oleh (Damanik et al., 2020) menggunakan metode K-Means dalam perbandingan tingkat pengangguran dengan tenaga kerja berdasarkan provinsi dengan mengelompokkan jumlah provinsi yang memiliki pengangguran sebesar 57% daripada jumlah

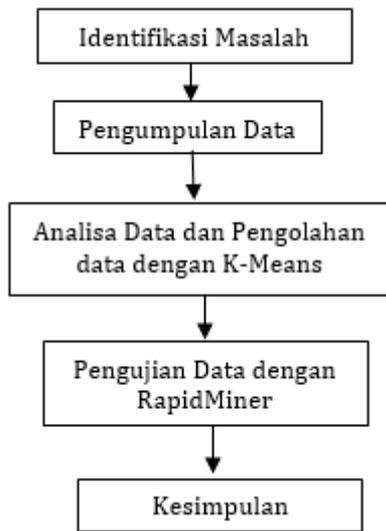
provinsi yang memiliki tenaga kerja sebesar 43%. Peneliti (Nurseptia et al., 2022) melakukan pemetaan lapangan kerja baru yang ada di provinsi Jawa Barat dengan 3 atribut yaitu tingkat pengangguran terbuka (TPT), tingkat partisipasi angkatan kerja (TPAK) dan produk domestik regional bruto (PDRB) dengan 3 cluster yang memiliki karakteristik yang berbeda. Peneliti (Aziz & Norhasanah, 2023) menggunakan metode K-Means yang menggunakan data pengangguran provinsi dari tahun 2019 hingga 2022 untuk melakukan klusterisasi tenaga kerja berdasarkan provinsi dengan membagi menjadi 2 *cluster* dengan hasil analisa 23 provinsi dengan nilai cluster pengangguran tinggi dan 11 provinsi dengan nilai cluster pengangguran rendah. Penelitian selanjutnya (Nagari & Inayati, 2020) menggunakan algoritma K-Means untuk proses pengelompokan status gizi anak dibawah usia 60 bulan menggunakan K-Means dengan hasil 4 kluster yang terbentuk terdiri dari 23 balita gizi buruk, 17 balita gizi kurang, 7 balita gizi baik dan 10 balita gizi lebih sehingga bisa dijadikan perhatian untuk mencegah kematian yang lebih tinggi.

Penelitian ini dilakukan untuk mengetahui *clusterisasi* data provinsi berdasarkan angkatan kerja menggunakan data mining. Kebaruan penelitian ini yaitu melakukan pengelompokan provinsi dari angkatan kerja berdasarkan kelompok usia mulai dari 15 tahun sampai usia 60 tahun menjadi 3 cluster dari 34 provinsi di Indonesia sedangkan peneliti sebelumnya pengelompokkan jumlah penduduk Sumatera Barat berdasarkan angkatan kerja berdasarkan atribut yang berhubungan dengan karakteristik angkatan kerja dan karakteristik bukan angkatan kerja (Safira et al., 2020). Analisa yang dilakukan dalam penelitian ini menggunakan algoritma K-Means dan memiliki tujuan untuk pengelompokkan provinsi dengan angkatan kerja tinggi, sedang dan rendah. Hasil dari pengelompokkan ini diharapkan dapat memberikan hasil yang membantu pemerintah daerah agar bisa mengembangkan dan meningkatkan pertumbuhan ekonomi di provinsi masing - masing sehingga bisa meningkatkan kesejahteraan penduduk diwilayahnya dan juga diharapkan dapat membantu peneliti selanjutnya terkait algoritma K-Means.

BAHAN DAN METODE

Dalam penelitian yang dilakukan diperlukan beberapa tahapan yang harus dilakukan agar data yang diperoleh bisa diolah untuk menghasilkan sebuah klusterisasi agar bisa mengelompokkan

berdasarkan kriteria tertentu. Tahapan penelitian dapat dilihat pada gambar 1.



Sumber : (Aria, 2024)

Gambar 1. Tahapan penelitian

A. Identifikasi Masalah

Untuk mendapatkan hasil yang sesuai berdasarkan metode yang akan digunakan dalam sebuah penelitian maka identifikasi masalah merupakan suatu hal yang perlu mutlak harus ditentukan sesuai dengan permasalahan yang ada. Penelitian melakukan identifikasi berdasarkan angkatan kerja dilihat dari usia angkatan kerja yang ada di 34 provinsi di Indonesia yang memiliki jumlah yang berbeda dalam pengelompokan umur sehingga dapat mempengaruhi pengelompokan produktivitas suatu provinsi tertentu.

B. Pengumpulan Data

Data yang digunakan untuk penelitian ini diambil dari sumber sekunder yang berasal dari Badan Pusat Statistik yang terdiri 34 provinsi yang memberikan data keadaan angkatan kerja di Indonesia pada bulan Agustus 2023.

C. Analisa Data dan Pengolahan data dengan K-Means.

Untuk analisa data dilakukan dengan menganalisa permasalahan yang ada di 34 provinsi berdasarkan angkatan kerja berdasarkan setiap langkah yang digunakan dalam proses algoritma K-Means (Utomo, 2020) :

1. Tentukan nilai k yang akan digunakan sebagai jumlah *cluster* yang akan dibentuk.
2. Tentukan titik pusat awal dari setiap masing – masing *cluster* sesuai jumlah cluster yang dibentuk.

3. Hitung jarak setiap data masing-masing *centroid* menggunakan rumus jarak Euclidean sehingga ditemukan jarak terdekat dari setiap data dengan *centroid*.
4. Klasifikasikan data berdasarkan kedekatan dengan *centroid*.
5. Hitung kembali pusat *cluster* yang terbaru berdasarkan data yang sekarang.
6. Lakukan perhitungan kembali dari setiap objek dengan menggunakan nilai pusat *cluster* baru. Jika pusat cluster tidak berubah lagi maka proses selesai. Jika masih ada data yang berubah maka ulangi kembali langkah 3 sampai pusat *cluster* tidak mengalami perubahan.

D. Pengujian Data dengan RapidMiner

Dengan melakukan pengujian data yang sudah diolah dengan menggunakan aplikasi Rapidminer. Rapidminer adalah aplikasi yang digunakan untuk mengolah data mining (Fitriani et al., 2022) dan aplikasi ini berbasis GUI yang bisa digunakan dengan berbagai sistem operasi (Sudarsono et al., 2021). Tahapan dalam menggunakan Rapidminer (RapidMiner, 2023).

1. Lakukan proses import data pada aplikasi dengan memilih file yang akan digunakan dan tentukan tipe data yang akan digunakan dalam file yang sudah dipilih serta tentukan penyimpanan file tersebut ke dalam *local repository*.
2. Letakkan file ke *blank process* dan tentukan operator K-Means dan tentukan banyaknya *cluster* yang akan digunakan.
3. Hubungkan bagian *output* ke *result* dan kemudian klik *run process*.
4. Hasil pengolahan data Rapidminer dapat dilihat dalam tab *example set* yang akan ditampilkan dalam bentuk pengelompokan berdasarkan masing – masing *cluster* dan tampilan dalam bentuk grafis.

E. Pengambilan Keputusan

Dengan menarik kesimpulan dari hasil perhitungan dan pengolahan data yang telah dihitung secara manual dengan data hasil olah dari aplikasi RapidMiner.

HASIL DAN PEMBAHASAN

1. Perhitungan dengan algoritma K-Means

Data yang digunakan dalam penelitian ini adalah data angkatan kerja yang ada diperkotaan dan pedesaan yang terdapat di 34 provinsi di Indonesia. Data yang akan diolah terdapat di gambar 2 yang berisi 11 atribut yang terdiri dari provinsi, usia 15-19,usia 20-24,usia 25-29, usia 30-

34,usia 35-39,usia 40-44,usia 45-49,usia 50-54,usia 55-59 dan usia 60+.

Tabel 1. Data Angkatan Kerja

No	Provinsi	Usia 15-19	Usia 20-24	Usia 25-29	Usia 30-34	Usia 35-39	Usia 40-44	Usia 45-49	Usia 50-54	Usia 55-59	Usia 60+
1	Aceh	91.659	291.753	331.080	330.807	326.043	304.440	277.772	231.236	177.313	242.349
2	Sumatera Utara	466.460	971.695	1.021.438	958.430	933.719	874.903	806.289	686.565	543.730	758.393
3	Sumatera Barat	131.280	333.558	369.002	357.384	349.771	336.861	311.777	275.174	225.140	334.483
33	Papua Barat	21.770	61.853	82.143	82.143	82.170	72.579	60.933	48.449	37.071	45.912
34	Papua	122.869	261.462	331.704	331.704	324.431	303.485	261.745	211.668	160.195	217.636

Sumber : (BPS,2023)

A. Menentukan jumlah cluster

Dari angkatan kerja akan dikelompokkan menjadi 3 cluster dimana cluster 0 adalah tingkat angkatan kerja tinggi, cluster 1 adalah cluster tingkat angkatan kerja sedang dan cluster 2 adalah tingkat angkatan kerja rendah.

B. Menentukan nilai centroid

Pemilihan nilai centroid yang akan diambil secara acak dari 34 provinsi yang ada. Berdasarkan data angkatan kerja yang akan diambil sebagai berikut :

Data ke 1 sebagai pusat cluster ke 1

Data ke 10 sebagai pusat cluster ke 2

Data ke 15 sebagai pusat cluster ke 3

Tabel 2. Data Pusat Centroid Awal

No	Provinsi	Usia 15-19	Usia 20-24	Usia 25-29	Usia 30-34	Usia 35-39	Usia 40-44	Usia 45-49	Usia 50-54	Usia 55-59	Usia 60+
1	Aceh	91.659	291.753	331.080	330.807	326.043	304.440	277.772	231.236	177.313	242.349
2	Kep. Bangka Belitung	36.413	88.538	98.987	97.523	95392	89.683	82.574	70.824	53.791	71.902
3	Sumatera Barat	787.505	2229315	2484058	2529416	2578974	2612846	2594439	2359385	2058217	3634607

Sumber : (Aria, 2024)

C. Menentukan jarak cluster

Menentukan letak cluster dengan cara membandingkan nilai antara 3 cluster dimana nilai cluster yang paling kecil merupakan nilai yang akan

menjadi pilihan. Jika ada yang paling kecil maka nilai tersebut bisa dimasukkan kedalam cluster tersebut. Hasil perhitungan jarak cluster dapat dilihat di tabel 3.

Tabel 3. Jarak Cluster Iterasi 1

No	Provinsi	C1	C2	C3	Jarak Terdekat
1	Aceh	0	180106,3	278841,1	0
2	Sumatera Utara	1747918,0	1576030,4	1911107,5	1576030,4
3	Sumatera Barat	144703,2	118476,6	323721,1	118476,6
33	Papua Barat	658377,6	829949,8	563153,9	563153,9
34	Papua	58650,2	208900,3	286428,2	58650,2

Sumber : (Aria, 2024)

D. Kelompokkan objek berdasarkan jarak minimum

Setelah menentukan posisi data dari angkatan kerja 34 provinsi di Indonesia

berdasarkan jarak minimum data terhadap *cluster*. Kemudian data tersebut dikelompokkan berdasarkan posisi data dari setiap *cluster* tersebut. Adapun data tersebut ditampilkan pada tabel 4.

Tabel 4. Hasil *Cluster* Iterasi 1

No	Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
1	Aceh	0	180106,3	278841,1	0	0
2	Sumatera Utara	1747918,0	1576030,4	1911107,5	1576030,4	1
3	Sumatera Barat	144703,2	118476,6	323721,1	118476,6	1
33	Papua Barat	658377,6	829949,8	563153,9	563153,9	2
34	Papua	58650,2	208900,3	286428,2	58650,2	0

Sumber : (Aria, 2024)

E. Menentukan nilai centroid baru iterasi ke 2

Untuk menghitung nilai centroid baru maka kita harus menghitung berapa banyak *cluster* yang dipilih dan kemudian membaginya berdasarkan

jumlah *cluster* yang dipilih. Perhitungan ini berdasarkan hasil jarak yang didapatkan pada iterasi pertama digunakan untuk iterasi kedua. Nilai *centroid* baru ada di tabel 5.

Tabel 5. Data Pusat *Centroid* Awal Iterasi 2

	Usia 15-19	Usia 20-24	Usia 25-29	Usia 30-34	Usia 35-39	Usia 40-44	Usia 45-49	Usia 50-54	Usia 55-59	Usia 60+
C1	97.604	1.291.824	292.929	290.573	287.921	278.422	255.162	216.473	169.237	242.456
C2	364.085	913.306	985.277	977.703	972.420	949.226	899.840	788.673	647.874	1.000.187
C3	50.214	117.830	136.477	134.989	132.191	127.551	118.449	100.948	79.956	121.274

Sumber : (Aria, 2024)

Kemudian tentukan nilai terkecil dari *cluster* 1, *cluster* 2 dan *cluster* 3 untuk mendapatkan hasil secara keseluruhan dari data iterasi 2 untuk mencari jarak terdekat dan posisi masing – masing *cluster*. Hasil pengelompokkan dari iterasi 2 dapat dilihat pada tabel 6. Berdasarkan jarak kedekatan yang telah diperoleh maka langkah selanjutnya mencari nilai dari setiap *cluster* yang ada sehingga bisa terlihat masing – masing dari setiap provinsi

yang ada. Berdasarkan hasil dari tabel 6 dimana posisi *cluster* yang diperoleh masih tidak sama dengan hasil iterasi 1 maka harus dilakukan klasterisasi kembali untuk iterasi 3 agar data yang diperoleh tidak mengalami perubahan dari iterasi sebelumnya. Berdasarkan tabel 6 didapatkan :
Cluster 0 = 3 provinsi
Cluster 1 = 6 provinsi
Cluster 2 = 25 provinsi

Tabel 6. Hasil *Cluster* Iterasi 2

No	Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
1	Aceh	1003083,4	1906303,9	493504,3	493504,3	2
2	Sumatera Utara	1708597,3	333284,7	2236429,6	333284,7	1
3	Sumatera Barat	977314,9	1774526,4	618107,0	618107,0	2
4	Riau	936595,7	1743430,0	666371,2	666371,2	2
5	Jambi	1102850,6	2144669,9	250362,8	250362,8	2
33	Papua Barat	1346555,7	2560527,8	169688,5	169688,5	2
34	Papua	1033539,9	1937716,6	468129,1	468129,1	2

Sumber : (Aria, 2024)

F. Menentukan nilai centroid baru iterasi ke 3

Tabel 7 adalah nilai *centroid* baru yang akan digunakan di iterasi 3 sebagai nilai untuk menentukan jarak terdekat dan posisi *cluster* yang

ada. Penentuan nilai *centroid* dilakukan berdasarkan nilai yang sudah diperoleh dari iterasi 2.

Tabel 7. Data Pusat *Centroid* Awal Iterasi 3

	Usia 15-19	Usia 20-24	Usia 25-29	Usia 30-34	Usia 35-39	Usia 40-44	Usia 45-49	Usia 50-54	Usia 55-59	Usia 60+
C1	225.049	518.437	565.073	557.070	552.005	531.130	501.596	429.039	338.174	511.611
C2	619.471	1.591.900	1.703.960	1.694.623	1.693.756	1.663.149	1.588.592	1.403.089	1.170.311	1.828.438
C3	77.441	189.114	215.103	212.397	207.966	199.859	183.718	156.521	122.774	181.151

Sumber : (Aria, 2024)

Tabel 8 mengelompokkan provinsi berdasarkan jarak terdekat dan cluster. Setelah mengetahui nilai kedekatan dari masing – masing provinsi maka selanjutnya menentukan nilai *cluster* yang ada di setiap provinsi sehingga provinsi yang memiliki nilai yang sama akan dikelompokkan dalam 1 cluster yang sama juga. Untuk hasil pengelompokkan *cluster* dapat dilihat di tabel 8.

Dari tabel 8 diperoleh :

Cluster 0 : 6 provinsi

Cluster 1 : 3 provinsi

Cluster 2 : 25 provinsi

pengelompokkan yang didapatkan masih ada provinsi yang mengalami perubahan di setiap kelompok *cluster* dibandingkan dengan hasil yang sudah diperoleh di iterasi 3 sehingga harus melakukan pengelompokkan kembali ke iterasi ke 4.

Tabel 8. Hasil *Cluster* Iterasi 3

No	Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
1	Aceh	682037,5	4002961,9	290234,7	290234,7	2
2	Sumatera Utara	1076697,8	2297363,0	2032699,4	1076697,8	0
3	Sumatera Barat	551168,9	3871060,8	414505,1	414505,1	2
4	Riau	520776,5	3839522,3	463858,3	463858,3	2
5	Jambi	920259,2	4241223,6	50183,3	50183,3	2
33	Papua Barat	1336470,7	4656846,6	371959,7	371959,7	2
34	Papua	712893,2	4034362,0	267004,3	267004,3	2

Sumber : (Aria, 2024)

G. Menentukan nilai centroid baru iterasi ke 4

Selanjutnya menentukan nilai *centroid* iterasi ke 4 berdasarkan hasil pengelompokkan yang ada di *cluster* iterasi 3 kemudian membaginya dengan banyaknya data yang ada. Nilai pusat

centroid iterasi 4 ada di tabel 9. Dari setiap nilai yang ada dikelompok usia sesuai data yang dimiliki oleh setiap provinsi sehingga mendapatkan jarak terdekat dan kelompok dari setiap *cluster*.

Tabel 9. Data pusat *centroid* awal iterasi 4

	Usia 15-19	Usia 20-24	Usia 25-29	Usia 30-34	Usia 35-39	Usia 40-44	Usia 45-49	Usia 50-54	Usia 55-59	Usia 60+
C1	263.614	634.748	693.255	675.482	667.536	636.882	594.140	503.478	390.208	540.907
C2	936.762	2.432.741	2.586.484	2.595.351	2.604.446	2.583.664	2.490.500	2.228.262	1.898.379	3.086.675
C3	77.858	189.554	215.468	213.064	208.712	200.890	184.558	157.294	123.175	181.142

Sumber : (Aria, 2024)

Selanjutnya dari nilai pusat *centroid*, lakukan kembali menentukan jarak terdekat dan kelompok *cluster* di iterasi 4 ini. Hasil jarak terdekat dan hasil pengelompokan terdapat di tabel 10. Berdasarkan hasil yang diperoleh di iterasi ke 4, pengelompokan *cluster* yang didapatkan sama dengan hasil pengelompokan di iterasi ke 3 maka proses algoritma K-Means sudah mendapatkan hasil yang optimal dan hasil akhir

tersebut akan dijadikan sebagai acuan dalam mengelompokkan tingkat angkatan kerja berdasarkan masing-masing provinsi. Hasil dari iterasi ke 4 mendapatkan pengelompokan *cluster* yaitu

- Cluster 0 = 6 provinsi
- Cluster 1 = 3 provinsi
- Cluster 2 = 25 provinsi

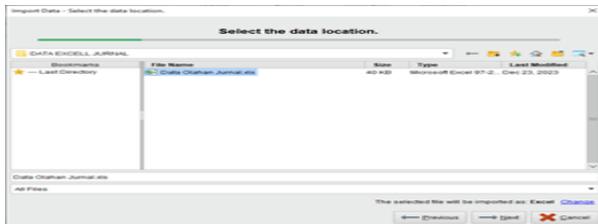
Tabel 10. Hasil Cluster Iterasi 4

No	Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
1	Aceh	965911,9	6770196,3	288422,1	288422,1	2
2	Sumatera Utara	786986,9	5066091,3	2030912,5	786986,9	0
3	Sumatera Barat	840981,9	6636901,1	412757,2	412757,2	2
33	Papua Barat	1623078,4	7421507,5	373734,2	373734,2	2
34	Papua	994976,2	6802158,6	265172,3	265172,3	2

Sumber : (Aria, 2024)

2. Pengujian data dengan Rapidminer

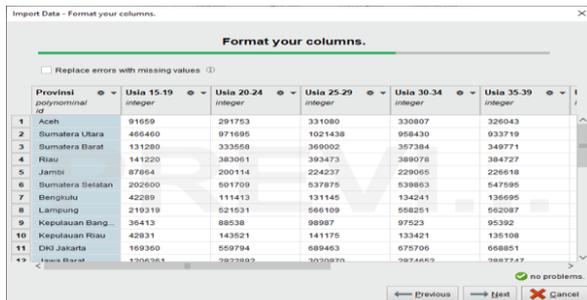
Dalam proses pengolahan dengan rapid miner dimulai dengan terlebih dahulu mengimport data dalam bentuk excel ke dalam aplikasi tersebut.



Sumber : (Aria,2024)

Gambar 2. Proses mengimport dataset

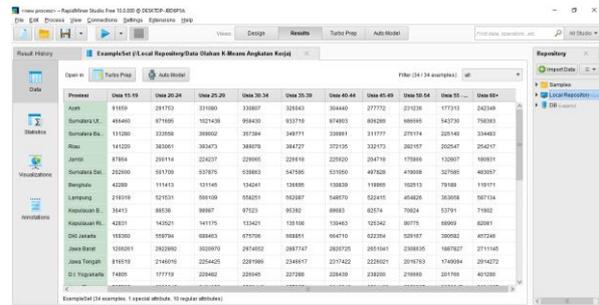
Setelah memilih file yang akan digunakan ke dalam Rapidminer maka langkah selanjutnya adalah melakukan format terhadap data yang ada dalam dataset dengan menentukan tipe data yang akan digunakan dalam setiap kolom.



Sumber : (Aria,2024)

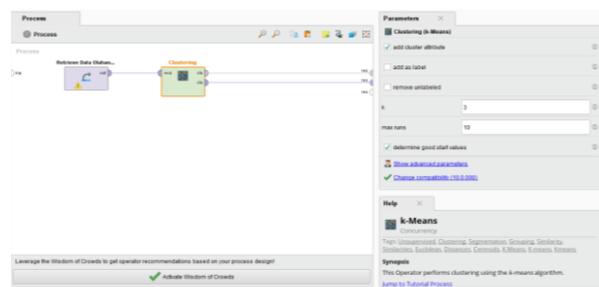
Gambar 3. Kolom dalam dataset yang akan di import ke aplikasi RapidMiner

Setelah proses mengimport dataset telah sesuai maka dalam aplikasi Rapidminer kemudian pilih *local repository* untuk menyimpan dataset tersebut dalam Rapidminer dan hasil dari dataset yang telah berhasil diimport dapat dilihat di gambar 4 yang selanjutnya data tersebut telah siap untuk digunakan dan dianalisa sesuai dengan kriteria algoritma K-means dalam Rapidminer.



Sumber : (Aria,2024)

Gambar 4 . Dataset yang akan diolah



Sumber : (Aria,2024)

Gambar 5. Proses mengimport dataset

Data yang telah disimpan dalam *local repository* Rapidminer selanjutnya akan diolah dengan menggunakan *clustering* dengan terlebih dahulu meletakkan dataset di *layout* kerja dari Rapidminer. Dalam penelitian ini cluster yang digunakan adalah $K=3$ sesuai dengan pengelompokan dari dataset yang memiliki usia angkatan kerja mulai dari usia 15 tahun – usia 60+ tahun dari 34 provinsi yang ada di Indonesia.



Sumber : (Aria,2024)

Gambar 6. Proses mengimport dataset

Berdasarkan data dari gambar 6 maka diperoleh pengelompokan berdasarkan *cluster* model yaitu *Cluster 0* memiliki 6 provinsi yaitu Sumatera Utara, Sumatera Selatan, Lampung, DKI Jakarta, Banten, Sulawesi Selatan. *Cluster 1* memiliki 3 provinsi yaitu Jawa Barat, Jawa Tengah, Jawa Timur dan *Cluster 2* memiliki 25 provinsi yaitu Aceh, Sumatera Barat, Riau, Jambi, Bengkulu, Kep.Bangka Belitung, Kep.Riau, D.I Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Berdasarkan data tersebut peneliti mengetahui provinsi mana saja yang termasuk provinsi untuk tingkat angkatan kerja tinggi, sedang dan rendah.

KESIMPULAN

Hasil penelitian yang telah dilakukan mendapatkan hasil pengelompokan usia angkatan kerja yang ada di 34 provinsi menjadi 3 *cluster* dengan menggunakan algoritma K-Means dan analisa dari aplikasi Rapidminer. Dari 34 provinsi terdapat 6 provinsi yang termasuk kedalam *cluster 0* yang termasuk kedalam provinsi yang memiliki usia angkatan kerja yang tinggi dan 3 provinsi dalam *cluster 1* dengan tingkat usia angkatan kerja yang sedang dan 25 provinsi masuk ke *cluster 2* dengan angkatan kerja yang rendah. Dari hasil pengelompokan tersebut terlihat bahwa angkatan kerja rendah lebih banyak berada diluar pulau

Jawa sehingga hal ini bisa dijadikan sebagai dasar pertimbangan untuk lebih meningkatkan lagi pertumbuhan ekonomi sehingga bisa memperluas peluang kerja dan menumbuhkan minat para investor untuk bisa menanamkan investasi mereka. Peneliti berharap untuk penelitian selanjutnya bisa lebih diperluas lagi data yang digunakan dengan kategori yang berbeda dan analisa faktor yang berbeda juga sehingga bisa memberikan hasil penelitian yang bermanfaat bagi pemerintah daerah dan peneliti lain.

REFERENSI

- Akil, I. (2021). Penilaian Kinerja Tim Evaluasi Perijinanpembukaan Program Studidengan Teknik Clustering K-Means. *INTI Nusa Mandiri*, 16(1), 1–6.
- Andriatno, A., Pasha, A. I., Nasida, F. K., & Alfi, N. (2021). Klasterisasi Provinsi Di Indonesia Menurut Karakteristik Ketenagakerjaan Tahun 2019 Menggunakan Metode Fuzzy C-Means Clustering. *Jurnal Manajemen Dan Kebijakan Publik*, 6(2), 124–136.
- Arifandi, M., Hermawan, A., & Avianto, D. (2021). Implementasi Algoritma K-Medoids Untuk Clustering Wilayah Terinfeksi Kasus Covid19 Di Dki Jakarta. *JTT (Jurnal Teknologi Terapan)*, 7(2), 120–128.
- Aziz, A., & Norhasanah, S. (2023). Klastrisasi Data Tenaga Kerja Terbuka Menurut Provinsi Dengan Penggunaan Algoritma K-Means. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(3), 444–456.
- BPS. (2023). *Keadaan Angkatan Kerja di Indonesia*. Badan Pusat Statistik. <http://bps.go.id>
- Damanik, A. R., Defit, S., Hartama, Dedy, P. P. A. N. . F. I., & Zer, R. . (2020). Implementasi Metode K-Means Dalam Perbandingan Tingkat Pengangguran Dengan Tenaga Kerja Berdasarkan Provinsi. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS) 2020*, 24–31.
- Fajriati, Y. R., & Syafriandi. (2022). Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar Menggunakan Metode CART dan Metode CHAID. *Journal of Mathematics UNP*, 7(3), 25–33.
- Fitriani, M., Nama, G. F., & Mardiana. (2022). Implementasi Association Rule Dengan Algoritma Apriori Pada Data Peminjaman Buku Upt Perpustakaan Universitas Lampung Menggunakan Metodologi CRISP-DM. *Jurnal Informatika Dan Teknik Elektro Terapan (JITET)*, 10(1), 41–49.
- Gustientiedina, Adiya, M. H., & Desnelita, Y. (2019).

- Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17-25.
- Handayanna, F. (2023). Penerapan Algoritma K-Means Untuk Klasterisasi Penduduk Miskin Di Provinsi Banten. *INTI Nusa Mandiri*, 18(1), 93-99.
- Kurnia, H., Zahrotun, L., & Linarti, U. (2021). Pengelompokan Mahasiswa Berdasarkan Data Akademik Sebelum Kuliah dan Masa Studi Menggunakan K-Medoids. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, 5(2), 265-272.
- Nagari, S. S., & Inayati, L. (2020). Implementation of Clustering Using k-Means Method To Determine Nutritional Status. *Jurnal Biometrika Dan Kependudukan*, 9(1), 62-68.
- Nurseptia, B., Voutama, A., & Haeryana, N. (2022). Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru. *Jurnal Teknologi Informasi*, 6(2), 181-186.
- Rahmati, R., & Wijayanto, A. W. (2021). ANALISIS Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering (Studi Kasus: Indeks Pembangunan Manusia tahun 2019). *JIKO (Jurnal Informatika Dan Komputer)*, 5(2), 73-80.
- RapidMiner. (2023). *Welcome to RapidMiner Documentation!*
<https://docs.rapidminer.com/>
- Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspim Gegerbitung). *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 291-298.
- Safira, D., Mustakim, Lestari, E. D., Iffa, M., & Annisa, S. (2020). Pengelompokan Jumlah Penduduk Sumatera Barat Berdasarkan Angkatan Kerja Menggunakan Algoritma K-Means. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 6(1), 26-31.
- Saputra, I. S., Zulfanetti, & Edi, J. K. (2019). Analisis faktor-faktor yang mempengaruhi Tingkat Partisipasi Angkatan Kerja (TPAK) di Provinsi Jambi. *Ekonomi Sumberdaya Dan Lingkungan*, 8(2), 68-81.
- Sudarsono, B. G., Leo, marcell I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *Journal of Business and Audit Information Systems*, 4(1), 13-21.
- Utomo, W. (2020). Penerapan Algoritma K-Means Untuk Pengelompokan Penyakit Kronis pada Warga Lansia (Studi Kasus Pada: Posyandu Lansia RW 07). *Jurnal Media Informatika Budidarma*, 4(4), 1153-1161.