

Klastrisasi Data Tenaga Kerja Terbuka Menurut Provinsi Dengan Penggunaan Algoritma K-Means

Abdul Aziz^{*1}, Siti Norhasanah²

^{1,2}Universitas Darwan Ali; Jln. Batu Berlian No.10, Mentawa Baru Hulu, Kec. Mentawa Baru
Ketapang, Kabupaten Kotawaringin Timur, Kalimantan Tengah

³Jurusan Sistem Informasi, Universitas Darwan Ali Sampit

e-mail: ^{*1}abdul.aziz@unda.ac.id, ²sitinorhasanah234@gmail.com

Abstrak

Penelitian ini bertujuan memberikan informasi yang valid dan akurat kepada pemerintah untuk mengambil keputusan dan kebijakan yang tepat dalam mengatasi pengangguran terbuka di provinsi-provinsi yang ada di Indonesia dengan tingkat pengangguran tinggi. Dengan alokasi sumber daya yang efektif dan efisien, diharapkan pembangunan ekonomi dan kesejahteraan masyarakat dapat meningkat. Penambangan informasi adalah siklus berulang dan cerdas untuk menemukan contoh atau model baru, yang benar-benar membantu untuk memahami kumpulan data yang sangat besar. Penambangan informasi yang berisi pencarian pola atau contoh ideal dalam kumpulan data yang sangat besar untuk membantu menentukan pilihan di kemudian hari. Untuk menguji informasi yang telah dipecah. Berikut penjelasan pengujian menggunakan pemrograman rapidminer 9.9. Keterkaitan yang diperoleh antara perhitungan perhitungan manual dan hasil yang akan ditunjukkan oleh instrumen rapidminer dan sangat mungkin beralasan bahwa konsekuensi estimasi manual dari teknik K-Means menyiratkan dan dari Microsoft excel 2019 aplikasi pemrograman setara dengan informasi ke dalam aplikasi Rapidminer. Penelitian menggunakan data dari 34 provinsi menunjukkan bahwa metode k-means clustering berhasil diterapkan dalam data mining. Data pengangguran provinsi dari tahun 2019 hingga 2022 digunakan sebagai sumber data. Hasil analisis menunjukkan terdapat 2 cluster, yaitu 23 provinsi dengan nilai cluster pengangguran tinggi dan 11 provinsi dengan nilai cluster pengangguran rendah. Dengan begitu pemerintah dapat memperluas lapangan pekerjaan dan mengatasi masalah pengangguran di wilayah tersebut.

Kata kunci: Data Mining, Rapidminer, Clustering, K-Means.

Abstrak

This research aims to provide valid and accurate information to the government to make appropriate decisions and policies in overcoming open unemployment in provinces in Indonesia with high unemployment rates. With an effective and efficient allocation of resources, it is hoped that economic development and social welfare will increase. Information mining is an iterative and intelligent cycle of finding new examples or models, which really helps to make sense of huge data sets. Information mining that contains the search for patterns or ideal examples in very large data sets to help make future choices. To test the information that has been broken down. The following is an explanation of the test using the rapidminer 9.9 programming. The linkages obtained between manual calculation calculations and the results to be shown by the rapidminer instrument and it is quite possible to reason that the consequences of manual estimation from the K-Means technique imply and from the Microsoft excel 2019 programming application are equivalent to information into the Rapidminer application. Research using data from 34 provinces shows that the k-means clustering method is successfully applied in data mining. Provincial unemployment data from 2019 to 2022 is used as the data source. The results

of the analysis show that there are 2 clusters, namely 23 provinces with high unemployment cluster scores and 11 provinces with low unemployment cluster scores. That way the government can expand employment and overcome the problem of unemployment in the region.

Keywords: Data Mining, Rapidminer, Clustering, K-Means.

1. PENDAHULUAN

Pengangguran terbuka adalah tenaga kerja yang tidak memiliki pekerjaan sama sekali. Meski sudah berusaha semaksimal mungkin, angkatan kerja sering mengalami pengangguran akibat malas mencari kerja atau malas bekerja.^[1] Sebagai negara berkembang Indonesia tidak akan lepas dari masalah pengangguran, salah satunya masalah pengangguran terbuka^[1]. Dengan mengatasi masalah pengangguran terbuka, peran pemerintah juga sangat penting. Masyarakat perlu memiliki semangat untuk mencari pekerjaan dan meningkatkan keterampilan mereka agar lebih mudah mendapatkan pekerjaan. Selain itu, masyarakat juga perlu mendukung program-program pemerintah untuk mengatasi masalah pengangguran. Oleh karena itu, pengurangan tingkat pengangguran terbuka menjadi prioritas utama bagi pemerintah dalam memajukan perekonomian dan meningkatkan kesejahteraan masyarakat.

Adanya pengangguran terbuka akan mengakibatkan beberapa hal, antara lain bertambahnya jumlah penduduk miskin karena tidak mampu memperoleh penghasilan dan memenuhi kebutuhan pokoknya. Selain itu, pengangguran terbuka berpotensi menurunkan daya beli masyarakat, serta menimbulkan ketimpangan sosial dan berbagai akibat lainnya.^[2]

Adapun faktor yang menjadi penyebab pengangguran terbuka diantaranya, lajunya pertumbuhan penduduk yang tinggi, ketersediaan lapangan pekerjaan yang tidak berimbang, ketidaksiapan penduduk usia produktif dalam menghadapi pesatnya kemajuan teknologi, tingkat Pendidikan, perubahan struktur perekonomian, dan besaran upah^[3] (Priastiwi and Handayani 2019).

Pemerintah Indonesia harus memiliki akses terhadap informasi yang dapat dipercaya untuk mengatasi masalah pengangguran terbuka. Tidak ada data yang dapat diandalkan mengenai pengangguran terbuka spesifik provinsi^[4]. Agar pemerintah dapat menentukan provinsi mana yang termasuk dalam klaster dengan tingkat pengangguran tinggi dan provinsi mana yang termasuk dalam klaster dengan tingkat pengangguran rendah, diperlukan suatu metode untuk mengklasifikasikan data pengangguran terbuka menurut provinsi dengan mempertimbangkan permasalahan tersebut di atas. Oleh karena itu, kami dapat mengelompokkan informasi pengangguran terbuka di Indonesia berdasarkan wilayah, jadi peneliti sangat membutuhkan strategi penambangan informasi menggunakan teknik pengelompokan *K-Means*^[2].

Adapun peneliti terdahulu yang sebelumnya dilakukan diantaranya ialah penelitian Riski Muliono, Zulfikar Sembiring pada *data mining clustering* menggunakan *algoritma k-means* untuk klasterisasi tingkat tridarma pengajaran dosen tujuan dari penelitian ini diharapkan dapat membantu proses klasterisasi dengan nilai yang akan mendekati karakteristik menjadi lebih efektif. Ketepatan prediksi yang akan dilakukan oleh algoritma *k-means* terhadap 15 data mengalami perbedaan ketepatan, hanya sebanyak 53,33% akurasi prediksi bernilai besar[5]. Penelitian Zulfa Nabila (dkk) analisis data mining untuk clustering kasus *covid-19* di provinsi lampung dengan *algoritma k-means* dari penelitian ini adalah untuk menganalisa data kasus *covid-19* agar dapat mengetahui pengelompokan pada masalah *covid-19* di provinsi lampung. Pengelompokan data kasus *covid-19* di provinsi lampung dilakukan dengan menggunakan

metode *clustering* dengan *algoritma k-means* dengan atribut kabupaten/kota, suspek, probable, konfirmasi positif, selesai isolasi, dan kematian yang digunakan dalam proses perhitungan dan membagi data ke dalam 4 *cluster* yang dikategorikan sebagai zona merah, zona orange, zona kuning dan zona hijau. Serta validasi menggunakan perhitungan *Davies-Bouldin Index (DBI)*[6]. Akramunnisa, Fajriani Penelitian sebelumnya antara lain *k-means clustering analysis* pada persebaran tingkat pengangguran kabupaten/kota di Sulawesi dan penelitian ini bertujuan untuk mengetahui hasil persebaran kabupaten/kota di Sulawesi selatan berdasarkan indikator tingkat pengangguran dengan menggunakan *clustering analysis*. Indikator yang digunakan adalah tingkat pengangguran terbuka (TPT), upah minimum kabupaten/kota (UMK) dan laju pertumbuhan indeks pembangunan manusia (IPM).

Tujuan dari teknik analisis data yang di kenal dengan istilah “*Clustering*” adalah untuk mengumpulkan dan mengatur sekumpulan data informasi yang memiliki kesamaan[7]. *K-Means* adalah sejenis algoritma yang digunakan untuk mengelompokkan item serupa menjadi satu. Pendekatan *K-Means* adalah contoh teknik pemodelan dan analisis data tanpa pengawasan (dipandu sendiri) yang menggunakan skema partisi untuk mengklasifikasikan data. Untuk temuan penelitian ini dimaksudkan untuk menginformasikan keputusan kebijakan tentang alokasi sumber data ke provinsi-provinsi di Indonesia dengan tingkat pengangguran yang tinggi secara tidak proposional, yang menjadi penghambat pembangunan ekonomi negara.

2. METODE PENELITIAN

2.1. Data Mining

Menemukan pola atau model baru yang sangat berguna untuk dipahami dalam *database* yang sangat besar adalah tujuan dari penambangan data, yang merupakan proses berulang dan interaktif. Penambangan informasi yang berisi pencarian pola atau contoh ideal dalam kumpulan data yang sangat besar untuk membantu dinamika di kemudian hari[8]

2.2. K-Means

Perhitungan *K-Means* merupakan perhitungan pengelompokan yang dapat mengelompokkan informasi berdasarkan titik tempat kelompok (*centroid*) terdekat dengan informasi tersebut. Pengelompokan data dengan *K-Means* bertujuan untuk meminimalkan [9]kesamaan data antar *cluster* dan memaksimalkan kesamaan data dalam satu cluster. untuk menentukan kesamaan data menggunakan jarak terpendek dari titik pusat ke data.

2.3. Clustering

Tujuan *clustering*, teknik analisis data yang sering dimasukkan ke dalam data mining, adalah untuk mengelompokkan data dengan karakteristik serupa ke dalam satu wilayah dan data dengan karakteristik berbeda ke dalam wilayah lain[7]. Menurut Baskoro menyatakan bahwa “*Clustering atau klusterisasi adalah salah satu alat batu pada data mining yang bertujuan mengelompokkan objek-objek ke dalam cluster-cluster*”. Karena tidak memiliki variabel target, cluster berbeda dari klasifikasi. Nilai variabel target tidak dikategorikan, diestimasi, atau diprediksi oleh klaster. Algoritma berdasarkan rata-rata terdekat dapat digunakan untuk menempatkan objek ke dalam cluster tertentu dengan menggunakan metode *K-Means*. [10] Ada beberapa langkah yang dapat dilakukan dalam proses pembuatan *K-Means Cluster*, yaitu sebagai berikut:

1. Tentukan jumlah *cluster* (*k*) yang dibutuhkan.
2. Menghasilkan *k centroid* awal (rata-rata dari setiap tandan).

3. Masukkan objek ke dalam *cluster* yang sesuai berdasarkan jarak terpendek antara objek tersebut dengan centroid setelah dihitung jaraknya.
4. Tentukan centroid cluster baru.
5. Terus lakukan langkah 3 dan 4 hingga tidak ada lagi objek yang ditransfer antar *cluster*.

2.4. Rapid Miner

Rapid Miner adalah perangkat lunak yang (*open source*) dan dapat diakses oleh siapa saja. *Rapid Miner*, yang dapat dimanfaatkan sebagai solusi analisis data.

3. HASIL & PEMBAHASAN

Berikut adalah beberapa gambaran perhitungan manual yang digunakan dalam *algorithm k-means clustering* pada data pengangguran terbuka per provinsi dengan menggunakan konsep *data mining* untuk mendapatkan temuan penelitian.

1. menentukan jumlah data yang akan diklaster dan lokasi yang akan digunakan sampel data pengangguran terbuka dalam proses *clustering*, khususnya data pengangguran terbuka yang mencakup 34 provinsi dari tahun 2019 hingga 2022. Dapat dilihat dari tabel 1 Data Pengangguran Terbuka Menurut Provinsi sebagai berikut.

Tabel 1. Data Pengangguran Terbuka Menurut Provinsi

Provinsi	Tingkat Pengangguran Terbuka Menurut Provinsi (Persen)				
	Total 2019	Total 2020	Total 2021	Total 2022	Rata-rata
Aceh	5,83	6,00	6,30	6,07	6,05
Sumatera Utara	5,48	5,81	6,17	5,82	5,82
Sumatera Barat	5,38	6,07	6,60	6,23	6,07
Riau	5,56	5,62	4,69	4,39	5,06
Jambi	3,79	4,70	4,93	4,65	4,51
Sumatera Selatan	4,28	4,71	5,08	4,69	4,69
Bengkulu	2,84	3,58	3,69	3,49	3,40
Lampung	3,99	4,47	4,62	4,42	4,37
Kep. Bangka Belitung	3,45	4,30	5,04	4,48	4,32
Kep. Riau	7,26	8,16	10,02	8,13	8,39
Dki Jakarta	6,02	8,05	8,51	7,59	7,54
Jawa Barat	7,91	9,09	9,37	8,33	8,67
Jawa Tengah	4,32	5,34	5,96	5,66	5,32
Di Yogyakarta	3,04	3,98	4,42	3,90	3,83
Jawa Timur	3,80	4,72	5,46	5,15	4,78
Banten	7,83	9,32	9,00	8,31	8,61
Bali	1,40	3,44	5,40	4,82	3,76
Nusa Tenggara Barat	3,22	3,63	3,49	3,41	3,44
Nusa Tenggara Timur	3,06	3,46	3,58	3,42	3,38
Kalimantan Barat	4,21	5,14	5,78	4,99	5,03
Kalimantan Tengah	3,63	3,96	4,39	4,23	4,05
Kalimantan Selatan	3,80	4,21	4,64	4,47	4,28
Kalimantan Timur	6,30	6,80	6,82	6,24	6,54
Kalimantan Utara	5,17	5,34	4,63	4,48	4,90

Sulawesi Utara	5,59	6,36	7,17	5,62	6,18
Sulawesi Tengah	3,29	3,35	3,74	3,34	3,43
Sulawesi Selatan	4,86	6,01	5,76	5,13	5,44
Sulawesi Tenggara	3,20	3,84	4,07	3,61	3,68
Gorontalo	3,51	3,79	3,21	2,92	3,35
Sulawesi Barat	2,14	2,86	3,21	2,73	2,73
Maluku	6,65	7,14	6,83	6,66	6,82
Maluku Utara	4,89	4,62	4,89	4,48	4,72
Papua Barat	6,12	6,79	6,01	5,58	6,12
Papua	3,37	3,85	3,55	3,22	3,50

1. Menentukan nilai *Centroid* yang akan ditentukan secara manual atau acak yang akan diambil dari Rata-Rata data.

Tabel 2. Centroid Data Awal

Cluster	Hasil
C1 (Max)	8,67
C2 (Min)	2,73

2. Menggunakan *Jarak Euclidean* untuk menentukan jarak antara setiap pusat massa objek dan titik yang sesuai. Berikut adalah rumus 1 yang menjelaskan cara menghitung jarak dari *centroid* pertama:

$$D_{(i,f)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

$$D(1.1) = \sqrt{(8,67 - 6,05)^2} = 2,63$$

$$D(1.2) = \sqrt{(8,67 - 5,82)^2} = 2,86$$

$$D(1.3) = \sqrt{(8,67 - 6,07)^2} = 2,61$$

Seterusnya perhitungan berlanjut ke perhitungan jarak *Centroid* ke-2 adalah sebagai berikut :

$$D(2.1) = \sqrt{(2,73 - 6,05)^2} = 3,32$$

$$D(2.2) = \sqrt{(2,73 - 5,82)^2} = 3,09$$

$$D(2.3) = \sqrt{(2,73 - 6,07)^2} = 3,34$$

Hingga $D_{2,34}$, dan seterusnya. Sebagai hasilnya, kita dapat menyusun tabel jarak dari pusat massa dan mencari nilai minimal tiga pusat massa. Berikut tabel 3 yang menjelaskan jarak dari *centroid* (Iterasi ke 1).

Tabel 3. Jarak Centroid (Iterasi Ke-1)

C1	C2	Jarak Terpendek
2,63	3,32	2,63
2,86	3,09	2,86
2,61	3,34	2,61
3,61	2,33	2,33
4,16	1,78	1,78
3,99	1,96	1,96

5,28	0,67	0,67
4,30	1,64	1,64
4,36	1,59	1,59
0,28	5,66	0,28
1,13	4,81	1,13
0,00	5,94	0,00
3,36	2,59	2,59
4,84	1,10	1,10
3,89	2,05	2,05
0,06	5,88	0,06
4,91	1,03	1,03
5,24	0,71	0,71
5,30	0,65	0,65
3,65	2,30	2,30
4,62	1,32	1,32
4,40	1,55	1,55
2,14	3,81	2,14
3,77	2,17	2,17
2,49	3,45	2,49
5,25	0,70	0,70
3,24	2,71	2,71
4,99	0,95	0,95
5,32	0,62	0,62
5,94	0,00	0,00
1,85	4,09	1,85
3,96	1,99	1,99
2,55	3,39	2,55
5,18	0,77	0,77

3. Mencirikan Tandan atau Pengelompokan

Dalam menentukan Grup dengan mencari nilai Grup berdasarkan nilai dasar dari nilai Grup dan akan ditetapkan dalam Grup yang dibandingkan dengan nilai dasar pada Penekanan 1. Pada Iterasi 1, tabel 4 *Cluster* terlihat seperti ini:

Tabel 4. Cluster (Iterasi ke-1)

C1	C2
	1
	1
	1
1	
1	
1	
1	
1	
1	
	1
	1

Abdul Aziz, et., al [Klastrisasi Data Tenaga Kerja Terbuka Menurut Provinsi Dengan Penggunaan Algoritma K-Means]

$$C2x = \frac{6,05 + 5,82 + 6,07 + 6,82 + 8,39 + 7,54 + 8,67 + 6,12 + 8,61 + 6,54 + 6,18}{11} = 6,98$$

Selanjutnya akan dilakukan kembali perhitungan. Proses iterasi dihentikan jika nilai *centroid* yang diperoleh dari iterasi sebelumnya dan nilai *centroid* yang diperoleh dari iterasi saat ini identik atau jika nilai *centroid* sudah optimal dan posisi cluster data pengangguran tidak akan berubah lebih lanjut. Namun, jika ada nilai *centroid* yang tidak terlalu mirip atau tidak ideal dan datanya masih berubah, maka siklus penekanan berlanjut pada penekanan berikutnya. Tabel 7 terlampir menunjukkan konsekuensi dari kumpulan iteratif ke iterasi 2:

Tabel 7. Nilai Centroid Baru (Iterasi 2)

Cluster	Hasil
C1 (Max)	4,17
C2 (Min)	6,98

Hitung nilai *centroid* baru dengan menggunakan nilai data rata-rata setelah didapatkan cluster iterasi kedua sebagai berikut:

$$D(1.1) = \sqrt{(4,17 - 6,05)^2} = 1,88$$

$$D(1.2) = \sqrt{(4,17 - 5,82)^2} = 1,65$$

$$D(1.3) = \sqrt{(4,17 - 6,07)^2} = 1,89$$

Ikuti prosedur ini hingga $D_{1\ 34}$. Jarak antara penganggur dan *centroid* cluster kedua kemudian dapat dihitung sebagai berikut:

$$D(2.1) = \sqrt{(6,98 - 6,05)^2} = 0,94$$

$$D(2.2) = \sqrt{(6,98 - 5,82)^2} = 1,16$$

$$D(2.3) = \sqrt{(6,98 - 6,07)^2} = 0,92$$

Hingga $D_{2\ 34}$, dan seterusnya. Sebagai hasilnya, kita dapat menyusun tabel jarak dari pusat massa dan mencari nilai minimal tiga pusat massa. Berikut tabel 9 jarak dari *centroid*:

Tabel 9. Cluster (Iterasi ke-2)

C1	C2	Jarak Terpendek
1,88	0,94	0,94
1,65	1,16	1,16
1,89	0,92	0,92
0,89	1,92	0,89
0,34	2,47	0,34
0,51	2,30	0,51
0,78	3,59	0,77
0,20	2,61	0,20
0,14	2,67	0,15
4,22	1,41	1,41
3,37	0,56	0,56
4,50	1,69	1,69

1,15	1,67	1,15
0,34	3,15	0,34
0,61	2,20	0,61
4,44	1,63	1,63
0,41	3,22	0,41
0,74	3,55	0,73
0,79	3,60	0,79
0,85	1,96	0,86
0,12	2,93	0,12
0,11	2,71	0,11
2,37	0,45	0,44
0,73	2,08	0,73
2,01	0,80	0,80
0,74	3,56	0,74
1,27	1,55	1,27
0,49	3,30	0,49
0,82	3,63	0,82
1,44	4,25	1,44
2,65	0,16	0,16
0,55	2,27	0,55
1,95	0,86	0,86
0,68	3,49	0,68

Anggota dari grup tersebut akan menjadi data dengan jarak *centroid* terkecil. Berikut adalah tabel dengan sepuluh posisi data untuk setiap cluster pada iterasi 2, dengan angka 1 yang menunjukkan bahwa data tersebut adalah anggota *cluster*. Jarak antara data dan setiap *centroid* kemudian dihitung:

Tabel 10. Cluster (Iterasi 2)

C1	C2
	1
	1
	1
1	
1	
1	
1	
1	
1	
	1
	1
	1
1	
1	
1	
	1
1	

1	
1	
1	
1	
1	
	1
1	
	1
1	
1	
1	
1	
1	
	1
1	
	1
1	
23	11

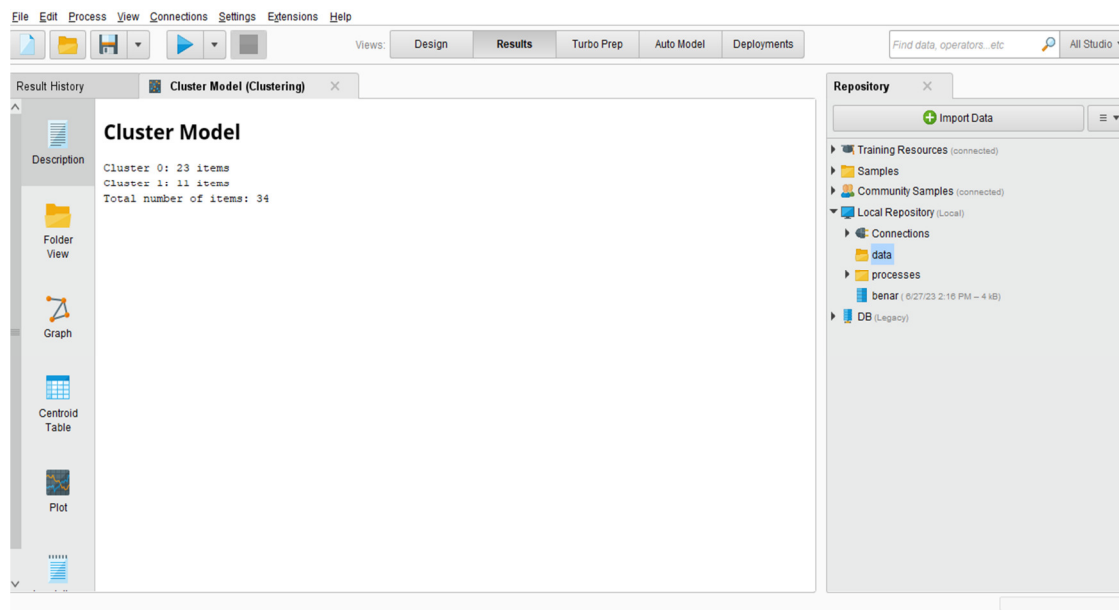
Tabel 11. Hasil Cluster (Iterasi 2)

Cluster	Hasil
C1 (Max)	23
C2 (Min)	11

Berikut produk akhir pada siklus pertama dan kedua, didapatkan hasil yang sama antara kedua penekanan senilai $C1 = 23$, $C2 = 11$ agar posisi kelompok dalam informasi tidak berubah sehingga siklus penekanan berhenti. berdasarkan nilai cluster dari iterasi 2 dan posisi cluster dari setiap kumpulan data pengangguran.

Hasil Analisis

Uji data yang telah dianalisis menggunakan perangkat lunak *Rapidminer 9.9* pada poin penelitian ini. Menghasilkan Penjelasan dari pengujian dengan menggunakan *software RapidMiner 9.9* yang dapat dilihat pada gambar di bawah ini. Hal ini disebabkan oleh koneksi yang dibuat antara hasil yang akan ditampilkan pada alat *rapidminer* dan hasil yang dihitung secara manual menggunakan metode K-means pada aplikasi *Microsoft Excel 2019*. Sehingga dapat disimpulkan bahwa hasil perhitungan manual menggunakan *software Microsoft Excel 2019* sama dengan data yang telah diinput dan diproses pada aplikasi *Rapidminer 9.9*.



Gambar 1. Nilai Cluster Model

KESIMPULAN

Untuk kesimpulan pada hasil penelitian ini, dengan menggunakan data penduduk pengangguran yang ada di Indonesia dari tahun 2019 s/d 2022, dan menggunakan tehnik *K-means Clustering*, Penulis dapat menyimpulkan, bahwa

Untuk menekankan bahwa teknik k-means clustering dapat digunakan untuk mengaplikasikan data mining dalam mengolah data. Terdapat dua kluster data, yang pertama data *cluster* tinggi dengan jumlah sebanyak 23 provinsi yaitu : Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI,Yogyakarta, Jawa Timur, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara dan Papua. *Cluster* rendah dengan jumlah sebanyak 11 provinsi selain dari *cluster* tinggi provinsi yang ada di Indonesia. Kluster tinggi adalah istilah yang digunakan dalam konteks analisis data dan teknik clustering seperti *k-means clustering*. Kluster tinggi merujuk pada kelompok atau kluster data yang memiliki keseragaman yang tinggi di dalamnya, artinya objek-objek dalam kluster tersebut sangat mirip atau serupa satu sama lain berdasarkan atribut atau fitur tertentu yang digunakan untuk mengelompokkan. Kluster rendah adalah istilah yang digunakan dalam konteks analisis data dan teknik clustering seperti k-means clustering. Kluster rendah merujuk pada kelompok atau kluster data yang memiliki tingkat keseragaman yang rendah di dalamnya, artinya objek-objek dalam kluster tersebut memiliki perbedaan yang signifikan dalam atribut atau fitur tertentu yang digunakan untuk mengelompokkan.

Alat Rapidminer 9.9 dapat digunakan untuk melakukan uji k-means pada kasus persentase data pengangguran pada penelitian ini, dan menghasilkan hasil yang sama dengan analisis perhitungan manual dan *alat Rapidminer 9.9*. 23 dari 34 provinsi dengan nilai kluster tinggi akan menjadi fokus upaya pemerintah untuk meningkatkan kesempatan kerja bagi pengangguran di provinsi-provinsi tersebut. Hal ini menjadi fokus pemerintah karena tingkat pengangguran yang tinggi dapat menyebabkan sejumlah masalah sosial dan ekonomi. fokus

pemerintah untuk meningkatkan kesempatan kerja di provinsi dengan kluster tinggi adalah untuk mengatasi masalah pengangguran, meningkatkan stabilitas sosial, dan mendorong pertumbuhan ekonomi di wilayah-wilayah tersebut.

4. SARAN

Saran untuk penelitian lanjutan, agar Dinas Sosial dapat memberikan data pengangguran berdasarkan skill dari pelatihan yang diberikan pada kesempatan tertentu agar, penelitian berikutnya mampu memisahkan data pengangguran berdasarkan nilai skill yang dimiliki setiap individu di setiap daerah yang ada di Indonesia.

5. UCAPAN TERIMA KASIH

Penulis ucapkan terima kasih banyak kepada Dinas Sosial, yang telah memberikan data untuk Penulis gunakan dalam penelitian ini, dan Penulis ucapkan kepada Bapak Abdul Aziz selaku Dosen Pembimbing dan turut serta membantu mengarahkan Penulis agar dapat menyelesaikan Penelitian ini.

DAFTAR PUSTAKA

- [1] R. Franita and A. Fuady, “Analisa Pengangguran di Indonesia”.
- [2] L. Padang and M. Murtala, “Pengaruh Jumlah Penduduk Miskin dan Tingkat Pengangguran Terbuka Terhadap Pertumbuhan Ekonomi di Indonesia,” *j. ekonomika indonesia*, Vol. 9, No. 1, p. 9, Nov. 2020, doi: 10.29103/ekonomika.v9i1.3167.
- [3] H. Aimon and E. Syofyan, “Analisis Faktor-faktor yang Mempengaruhi Produktivitas dan Tingkat Pengangguran di Indonesia”.
- [4] N. Syamsiyah, E. Novianti, and A. Pamungkas, “Penerapan Algoritma Clustering Dalam Mengelompokkan Tingkat Pengangguran pada Kelurahan Ujung Menteng Jakarta Timur,” Vol. 6, No. 2, 2022.
- [5] R. Muliono and Z. Sembiring, “Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen,” Vol. 4, No. 2, 2019.
- [6] Z. Nabila, A. R. Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 di Provinsi Lampung Dengan Algoritma K-Means,” *Jurnal Teknologi dan Sistem Informasi*, Vol. 2, No. 2.
- [7] A.- Akramunnisa and F. Fajriani, “K-Means Clustering Analysis pada Persebaran Tingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan,” *Varian*, Vol. 3, No. 2, pp. 103–112, Apr. 2020, doi: 10.30812/varian.v3i2.652.

- [8] M. Wahyudi and L. Pujiastuti, "*Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma K-Means*," Vol. 2, 2020.
- [9] S. Muharni, "*Penerapan Metode K-Means Clustering Pada Data Tingkat Pengangguran Terbuka Tahun 2016-2018 Dan 2019-202*," *Jurnal Informatika*, Vol. 22, No. 01, 2022.
- [10] A. A. Putrie and R. Sanjaya, "*Pengelompokan Kabupaten/Kota Berdasarkan Indikator Tingkat Pengangguran Menggunakan Algoritma K-Means Clustering (Studi Kasus: Provinsi Jawa Barat)*," Vol. 2, No. 2, 2021.