IMPLEMENTASI K-MEANS UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDEKS JUMLAH PENGANGGURAN TERBUKA

Muhammad Nurfathullah, Intan Purnamasari
Teknik Informatika, Universitas Singaperbangsa Karawang
Jalan HS. Ronggo Waluyo, Karawang, 41361; (0267) 641177
2010631170100@student.unsika.ac.id

ABSTRAK

Indeks pengangguran terbuka merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja. Berdasarkan rilisan tahunan yang dikeluarkan oleh Dinas Tenaga Kerja dan Transmigrasi pada tahun 2018-2023. Pengelompokkan provinsi dilakukan untuk memetakan daerah dengan indeks pengangguran terbuka sebagai salah satu upaya pemerintah dalam mengetahui indeks pengangguran terbuka setiap provinsi. Penelitian ini menggunakan algoritma k-means. Algoritma k-means merupakan metode dalam data mining yang digunakan untuk mengelompokkan atau mengklasifikasikan data ke dalam satu atau lebih kluster. Penelitian ini memiliki tujuan untuk mengisi kesenjangan tersebut dengan mengimplementasikan algoritma K-Means untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan data Dinas Tenaga Kerja dan Transmigrasi pada tahun 2018-2023. Hasil pengelompokkan membagi provinsi ke dalam 3 *cluster*, yaitu *cluster* 0 (C0) yang mencakup 11 provinsi, *cluster* 1 (C1) yang mencakup 12 provinsi, dan *cluster* 2 (C2) yang mencakup 11 provinsi. Hasill evaluasi pengelompokkan dengan *silhouette* coefficient menghasilkan nilai sebesar 0,54 yang menunjukkan bahwa kriteria pengelompokkan yang dilakukan termasuk ke dalam struktur *cluster* yang standar (*medium structure*).

Kata kunci : *Indeks pengangguran terbuka, K-means clustering, silhouette coefficient*

1. PENDAHULUAN

Di Indonesia, pengangguran terbuka merupakan salah satu masalah sosial dan ekonomi yang signifikan. Tingkat pengangguran yang tinggi dapat menghambat pertumbuhan ekonomi dan mengganggu kesejahteraan masyarakat. Untuk mengatasi masalah ini, penting untuk memahami pola dan karakteristik pengangguran di berbagai wilayah di Indonesia. Pengangguran terbuka sendiri merujuk pada individu yang tidak sedang bekerja, aktif mencari pekerjaan, bekerja kurang dari dua hari dalam seminggu, atau sedang berupaya untuk memperoleh pekerjaan yang sesuai dengan kualifikasi mereka [1].

Salah satu pendekatan yang dapat digunakan untuk menganalisis pola pengangguran di tingkat regional adalah menggunakan metode *clustering* atau pengelompokkan data. Metode *clustering* adalah salah satu metode yang dapat diandalkan dalam data mining dan merupakan alat yang valid untuk menyelesaikan masalah-masalah kompleks dalam bidang ilmu komputer dan statistik. Proses *clustering* melibatkan pengelompokkan titik-titik data ke dalam beberapa kelompok, dimana titik-titik data dalam kelompok yang sama memiliki kemiripan yang lebih besar satu sama lain dibandingkan dengan kelompok data lainnya [2]. Untuk memastikan keberhasilan analisis, diperlukan penerapan suatu algoritma. Dalam penelitian ini, algoritma yang digunakan adalah K-Means.

K-Means merupakan salah satu algoritma *clustering* yang populer dan efektif untuk mengelompokkan data berdasarkan fitur yang serupa, serta memisahkan data dengan karakteristik yang berbeda ke dalam kelompok yang berbeda [3].

Pada penelitian sebelumnya menjelaskan tentang klasifikasi pengangguran terbuka di Indonesia dengan algoritma *classification and regression tree* (cart) dan c4.5. Penelitian tersebut bertujuan untuk klasifikasi dengan jumlah data yang cukup besar dengan banyak faktor serta dapat melakukan analisis klasifikasi pada peubah respon baik nominal, ordinal, maupun kontinu [4].

Ada juga penelitian lain tentang pengangguran terbuka tentang klasifikasi dan prediksi tingkat pengangguran terbuka di Indonesia menggunakan metode *classification and regression tree* (CART) yang mendapatkan hasil yaitu akurasi sebesar 91,17% berdasarkan fungsi *recall, precision, accuracy and error radio*, dengan nilai masing-masing yaitu 96,87%, 93,93%, dan 8,83% [5].

Penelitian lain menjelaskan tentang clustering pengangguran umur 25 tahun keatas di Sumatera Utara menggunakan algoritma K-Medoids. Pada penelitian ini menghasilkan tiga *cluster*, yaitu 14 kabupaten di *cluster* 0, 2 kabupaten di *cluster* 1, dan 17 kabupaten di *cluster* 2 [6].

Pada penelitian sebelumnya menjelaskan tentang perbandingan tingkat pengangguran terbuka provinsi di Indonesia berbasis K-Means Clustering menggunakan data dari tahun 2016-2021 [2]. Selanjutnya, ada juga penelitian terkait penerapan algoritma K-Means untuk mengelompokkan pengangguran di Indonesia sudah pernah dilakukan menggunakan data dari tahun 2014-2019 [7]. Berbeda dengan penelitian penulis, dimana penulis menggunakan data pengangguran terbuka dari tahun 2018-2023.

Dengan menggunakan pendekatan ini, diharapkan dapat ditemukan pola-pola yang bermanfaat dalam memahami distribusi dan karakteristik pengangguran di tingkat provinsi. Informasi ini dapat menjadi dasar bagi pemerintah dan pemangku kepentingan lainnya dalam merancang kebijakan yang lebih efektif dalam mengatasi masalah pengangguran di berbagai wilayah di Indonesia.

Oleh sebab itu, penelitian ini memiliki tujuan untuk mengisi kesenjangan tersebut dengan mengimplementasikan algoritma K-Means untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan jumlah pengangguran terbuka. Dengan demikian, diharapkan hasil dari penelitian ini dapat memberikan wawasan yang berharga bagi pemangku kepentingan dalam upaya mengurangi tingkat pengangguran di Indonesia.

2. TINJAUAN PUSTAKA

2.1. Pengangguran Terbuka

Pengangguran terbuka merupakan individu yang tidak memiliki pekerjaan secara aktif. Ada yang mengalami pengangguran ini karena mereka telah berupaya dengan maksimal namun belum berhasil mendapatkan pekerjaan. Sementara ada pula yang mengalami pengangguran karena kurangnya motivasi dalam mencari pekerjaan atau bekerja [8].

2.2. *Clustering*

Clustering adalah teknik dalam data mining yang bertujuan untuk mengelompokkan atau mengklasifikasikan data ke dalam kelompok-kelompok yang memiliki karakteristik yang serupa, dengan tujuan untuk melakukan klasifikasi atau memprediksi nilai dari variabel target. Setiap kelompok dalam clustering berusaha untuk membuat pembagian data secara homogen, di mana objek yang berada dalam kelompok yang sama memiliki kesamaan karakteristik dan berbeda dari objek dalam kelompok lainnya [7].

2.3. Google Colab

Google Colab atau Google Colaboratory adalah platform pengembangan terpadu (IDE) untuk bahasa pemrograman Python. Di sini, komputasi dieksekusi oleh infrastruktur server Google yang memiliki spesifikasi perangkat keras tinggi [9].

Google Colab adalah program yang dikembangkan oleh Google Research. Google Colab memfasilitasi penulisan dan eksekusi kode Python melalui peramban web tanpa batasan, memberikan manfaat yang besar dalam konteks machine learning, analisis data, dan pendidikan [10].

2.4. Algoritma K-Means

Algoritma K-Means merupakan metode dalam data mining yang digunakan untuk mengelompokkan atau mengklasifikasikan data ke dalam satu atau lebih kluster, di mana data dengan karakteristik yang serupa ditempatkan dalam satu kluster, sedangkan data dengan karakteristik yang berbeda ditempatkan dalam kluster yang berbeda [11].

Algoritma K-Means mengelompokkan data menjadi beberapa kluster di mana data yang memiliki karakteristik serupa ditempatkan dalam kluster yang sama, sementara itu, data yang memiliki atribut yang tidak sama akan dikelompokkan ke dalam kluster yang berbeda [3].

3. METODE PENELITIAN

Dalam bagian ini, peneliti menerapkan suatu pendekatan penelitian yang dikenal dengan sebutan Knowledge Discovery In Database (KDD). Proses yang terdapat dalam metodologi KDD meliputi data *selection*, data *preprocessing*, *transformation*, data *mining*, *evaluation* [12].

3.1. *Data Selection*

Tahap pertama adalah proses *data selection* (pemilihan data) yang akan digunakan dalam proses pengelompokkan data. *Dataset* yang digunakan berasal dari data publik melalui situs terbuka data.jabarprov.go.id yang dirilis Dinas Tenaga Kerja Dan Transmigrasi. Data yang digunakan adalah data indeks tingkat pengangguran terbuka berdasarkan provinsi di Indonesia tahun 2018-2023. *Dataset* terdiri dari 203 record dengan 6 atribut. Atribut tersebut diantaranya id, kode_provinsi, nama_provinsi, indeks_tpt, satuan, dan tahun. Adapun isi dataset awal ditunjukkan pada Gambar 1.

	id	kode_provinsi	nama_provinsi	indeks_tpt	satuan	tahun
0	1	1100	ACEH	8.34	POIN	2018
1	2	1200	SUMATERA UTARA	5.55	POIN	2018
2	3	1300	SUMATERA BARAT	5.66	POIN	2018
3	4	1400	RIAU	5.98	POIN	2018
4	5	1500	JAMBI	3.73	POIN	2018
...	...	...	...	...	...	...
198	199	7500	GORONTALO	3.06	POIN	2023
199	200	7600	SULAWESI BARAT	2.27	POIN	2023
200	201	8100	MALUKU	6.31	POIN	2023
201	202	8200	MALUKU UTARA	4.31	POIN	2023
202	203	9100	PAPUA BARAT	5.38	POIN	2023

Gambar 1. Dataset Tingkat Pengangguran Terbuka Provinsi 2018-2023

3.2. *Data Pre-processing*

Dalam tahap pra-pemrosesan, atribut yang tidak diperlukan dihapus untuk mempersiapkan data agar siap digunakan pada tahapan berikutnya. Dalam penelitian ini, hanya atribut nama_provinsi, indeks tpt, dan tahun yang digunakan. Adapun hasil *data pre-processing* setelah dilakukan penghapusan atribut yang tidak diperlukan dapat dilihat pada Gambar 2.

	nama_provinsi	indeks_tpt	tahun
0	ACEH	6.34	2018
1	SUMATERA UTARA	5.55	2018
2	SUMATERA BARAT	5.66	2018
3	RIAU	5.98	2018
4	JAMBI	3.73	2018
...	...	...	...
198	GORONTALO	3.06	2023
199	SULAWESI BARAT	2.27	2023
200	MALUKU	6.31	2023
201	MALUKU UTARA	4.31	2023
202	PAPUA BARAT	5.38	2023

Gambar 2. *Data Pre-processing*

3.3. *Transformation*

Transformasi data dilakukan dengan mengubah struktur dataset awal yang memiliki dua kolom yang merepresentasikan indeks tpt dan tahun menjadi enam kolom terpisah, yaitu dari setiap indeks tpt dan tahun yang berbeda. Hasil dari transformasi data tersebut kemudian ditampilkan dalam Gambar 3.

tahun / nama_provinsi	2018	2019	2020	2021	2022	2023
ACEH	6.34	6.17	6.59	6.38	6.17	6.03
BALI	1.40	1.57	5.63	5.37	4.80	2.69
BANTEN	8.47	8.11	10.64	8.98	8.09	7.52
BENGKULU	3.35	3.26	4.07	3.65	3.59	3.42
DI YOGYAKARTA	3.37	3.18	4.57	4.56	4.06	3.69
DKI JAKARTA	6.65	6.54	10.95	8.50	7.18	6.53
GORONTALO	3.79	3.76	4.28	3.01	2.58	3.06
JAMBI	3.73	4.06	5.13	5.09	4.59	4.53
JAWA BARAT	8.23	8.04	10.46	9.02	8.31	7.44
JAWA TENGAH	4.47	4.44	6.48	5.95	5.57	5.13
JAWA TIMUR	3.91	3.82	5.84	5.74	5.49	4.88
KALIMANTAN BARAT	4.18	4.35	5.81	5.82	5.11	5.05
KALIMANTAN SELATAN	4.35	4.18	4.74	4.95	4.74	4.31
KALIMANTAN TENGAH	3.91	4.04	4.58	4.53	4.26	4.10
KALIMANTAN TIMUR	6.41	5.94	6.87	6.83	5.71	5.31
KALIMANTAN UTARA	5.11	4.49	4.97	4.58	4.33	4.01
KEP. BANGKA BELITUNG	3.61	3.58	5.25	5.03	4.77	4.56
KEP. RIAU	8.04	7.50	10.34	9.91	8.23	6.80
LAMPUNG	4.04	4.03	4.67	4.69	4.52	4.23
MALUKU	6.95	6.69	7.57	6.93	6.88	6.31
MALUKU UTARA	4.63	4.81	5.16	4.71	3.98	4.31
NUSA TENGGARA BARAT	3.58	3.28	4.22	3.01	2.89	2.80
NUSA TENGGARA TIMUR	2.85	3.14	4.28	3.77	3.54	3.14
PAPUA	3.00	3.51	4.28	3.33	2.83	NaN
PAPUA BARAT	6.45	6.43	6.88	5.84	5.37	5.38
RIAU	5.98	5.76	6.32	4.42	4.37	4.23
SULAWESI BARAT	3.01	2.98	3.32	3.13	2.34	2.27
SULAWESI SELATAN	4.94	4.62	6.31	5.72	4.51	4.33
SULAWESI TENGAH	3.37	3.11	3.77	3.75	3.08	2.95
SULAWESI TENGGARA	3.19	3.52	4.58	3.92	3.36	3.15

Gambar 3. Hasil Transformasi Data

3.4. *Data Mining*

Pada tahapan ini, dilakukan proses pemodelan dengan menerapkan algoritma k-means. Dengan hasil seperti Gambar 4.

```
k_means_optimum = KMeans(n_clusters = 3, init = 'k-means++', random_state=6)
y = k_means_optimum.fit_predict(X)
print(y)
```

```
[2 1 1 1 2 1 2 0 0 0 2 2 2 2 2 2 2 0 1 0 0 0 0 1 1 1 1 2 1 0 0 1 1 2 1 1
 1 2 1 2 0 0 0 2 2 2 2 2 2 2 0 1 0 0 0 0 1 1 1 1 2 1 0 0 1 1 2 1 1 1 1 2 1
 2 0 0 0 2 2 2 2 2 2 2 2 0 1 0 0 0 0 0 1 1 1 1 2 1 0 0 1 1 2 1 1 1 1 2 1 2 0 0
 0 2 2 2 2 2 2 2 2 0 1 0 0 0 0 0 1 1 1 1 2 1 0 0 1 1 2 1 1 1 1 2 1 2 0 0 0 2 2
 2 2 2 2 2 0 1 0 0 0 0 0 1 1 1 1 2 1 0 0 1 1 2 1 1 1 1 2 1 2 0 0 0 2 2 2 2 2
 2 2 0 1 0 0 0 0 0 1 1 1 1 2 1 0 0 1]
```

Gambar 4. Hasil Data Mining

Gambar 4 memberikan gambaran tentang hasil yang muncul setelah melalui proses pemodelan algoritma k-means. Dari gambar tersebut dapat terlihat representasi visual dari kelompok-kelompok yang terbentuk setelah menerapkan algoritma k-means pada data. Gambar tersebut memberikan gambaran tentang bagaimana provinsi di Indonesia dikelompokkan berdasarkan jumlah pengangguran terbuka.

3.5. *Evaluation*

Pada tahap evaluasi, dilakukan pengujian untuk menilai kualitas kluster yang terbentuk setelah proses pengelompokan. Metode evaluasi yang diterapkan dalam penelitian ini adalah dengan menggunakan nilai *Silhouette Coefficient*. *Silhouette Coefficient* adalah gabungan dari metode validasi klaster, termasuk metode kohesi yang mengevaluasi kedekatan objek dalam sebuah klaster, dan metode separasi yang mengevaluasi jarak antar klaster. Ketika nilai koefisien silhouette mendekati atau positif, maka struktur klaster dianggap baik. Sebaliknya, jika nilai koefisien silhouette adalah 0, maka struktur klaster dianggap tidak jelas. Tabel 1 menggambarkan kategori nilai koefisien silhouette berdasarkan Rousseeuw (1987).

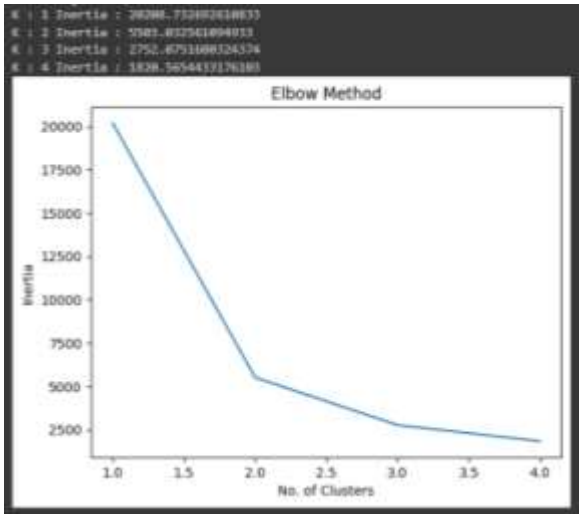
Tabel 1. Kategori Nilai *Silhouette Coefficient*

Nilai SC	Kriteria
0,71 - 1,00	Struktur *cluster* kuat (*Strong Structure)*
0,51 - 0,70	Struktur *cluster* standar (Medium *Structure)*
0,26 - 0,50	Struktur *cluster* lemah (Weak *Structure)*
≤ - 0,25	Tidak memiliki struktur (*No Structure)*

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini, *clustering* dilakukan menggunakan algoritma k-means yang diimplementasikan dengan menggunakan bahasa pemrograman Python. Pada tahap awal proses *clustering*, jumlah cluster optimal ditentukan menggunakan metode elbow. Metode elbow digunakan untuk menentukan jumlah *cluster* dengan memperhatikan perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik tertentu dengan nilai *Sum of Square Error* dari masing-masing

cluster. Grafik dari metode elbow yang digunakan dalam penelitian ini ditampilkan dalam Gambar 5.



Gambar 5. Grafik Metode Elbow

Berdasarkan grafik tersebut, terdapat tiga *cluster* yang terbentuk, yaitu *cluster* 0 (C0), *cluster* 1 (C1), dan *cluster* 2 (C2). Setelah itu, dilakukan pengelompokan data dengan memilih jarak terdekat dari setiap data. Hasil dari pengelompokan tersebut tersaji dalam Tabel 2.

Tabel 2. Hasil Pengelompokkan 2018-2019

Provinsi	2018	cluster	2019	cluster
Aceh	6,34	2	6,17	2
Bali	1,40	2	1,57	2
Banten	8,47	2	8,11	2
Bengkulu	3,35	2	3,26	2
DIY Yogyakarta	3,37	2	3,18	2
DKI Jakarta	6,65	2	6,54	2
Gorontalo	3,70	2	3,76	2
Jambi	3,73	2	4,06	2
Jawa Barat	8,23	2	8,04	2
Jawa Tengah	4,47	2	4,44	2
Jawa Timur	3,91	2	3,82	2
Kalimantan Barat	4,18	0	4,35	0
Kalimantan Selatan	4,35	0	4,18	0
Kalimantan Tengah	3,91	0	4,04	0
Kalimantan Timur	6,41	0	5,94	0
Kalimantan Utara	5,11	0	4,49	0
Kep. Bangka Belitung	3,61	0	3,58	0
Kep. Riau	8,04	0	7,50	0
Lampung	4,04	0	4,03	0
Maluku	6,95	0	6,69	0
Maluku Utara	4,63	0	4,81	0
Nusa Tenggara Barat	3,58	0	3,28	0
Nusa Tenggara Timur	2,85	1	3,14	1
Papua	3,00	1	3,51	1
Papua Barat	6,45	1	6,43	1
Riau	5,98	1	5,76	1
Sulawesi Barat	3,01	1	2,98	1
Sulawesi Selatan	4,94	1	4,62	1
Sulawesi Tengah	3,37	1	3,11	1
Sulawesi Tenggara	3,19	1	3,52	1
Sulawesi Utara	6,61	1	6,01	1
Sumatera Barat	5,66	1	5,38	1
Sumatera Selatan	4,27	1	4,53	1
Sumatera Utara	5,55	1	5,39	1

Tabel 3. Hasil Pengelompokkan 2020-2021

Provinsi	2020	cluster	2021	cluster
Aceh	6,59	2	6,30	2
Bali	5,63	2	5,37	2
Banten	10,64	2	8,98	2
Bengkulu	4,07	2	3,65	2
DIY Yogyakarta	4,57	2	4,56	2
DKI Jakarta	10,95	2	8,50	2
Gorontalo	4,28	2	3,01	2
Jambi	5,13	2	5,09	2
Jawa Barat	10,46	2	9,82	2
Jawa Tengah	6,48	2	5,95	2
Jawa Timur	5,84	2	5,74	2
Kalimantan Barat	5,81	0	5,82	0
Kalimantan Selatan	4,74	0	4,95	0
Kalimantan Tengah	4,58	0	4,53	0
Kalimantan Timur	6,87	0	6,83	0
Kalimantan Utara	4,97	0	4,58	0
Kep. Bangka Belitung	5,25	0	5,03	0
Kep. Riau	10,34	0	9,91	0
Lampung	4,67	0	4,69	0
Maluku	7,57	0	6,93	0
Maluku Utara	5,15	0	4,71	0
Nusa Tenggara Barat	4,22	0	3,01	0
Nusa Tenggara Timur	4,28	1	3,77	1
Papua	4,28	1	3,33	1
Papua Barat	6,80	1	5,84	1
Riau	6,32	1	4,42	1
Sulawesi Barat	3,32	1	3,13	1
Sulawesi Selatan	6,31	1	5,72	1
Sulawesi Tengah	3,77	1	3,75	1
Sulawesi Tenggara	4,58	1	3,92	1
Sulawesi Utara	7,37	1	7,06	1
Sumatera Barat	6,88	1	6,52	1

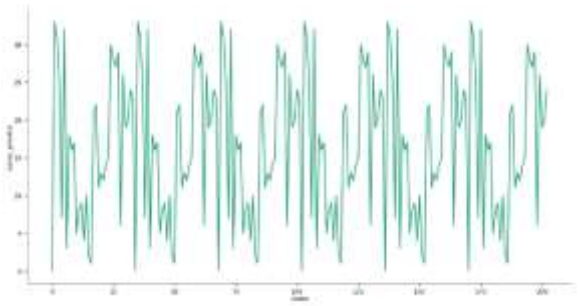
Provinsi	2020	cluster	2021	cluster
Sumatera Selatan	5,51	1	4,98	1
Sumatera Utara	6,91	1	6,33	1

Tabel 4. Hasil Pengelompokkan 2022-2023

Provinsi	2022	cluster	2023	cluster
Aceh	6,17	2	6,03	2
Bali	4,80	2	2,69	2
Banten	8,09	2	7,52	2
Bengkulu	3,59	2	3,42	2
DIY Yogyakarta	4,06	2	3,69	2
DKI Jakarta	7,18	2	6,53	2
Gorontalo	2,58	2	3,06	2
Jambi	4,59	2	4,53	2
Jawa Barat	8,31	2	7,44	2
Jawa Tengah	5,57	2	5,13	2
Jawa Timur	5,49	2	4,88	2
Kalimantan Barat	5,11	0	5,05	0
Kalimantan Selatan	4,74	0	4,31	0
Kalimantan Tengah	4,26	0	4,10	0
Kalimantan Timur	5,71	0	5,31	0
Kalimantan Utara	4,33	0	4,01	0
Kep. Bangka Belitung	4,77	0	4,56	0
Kep. Riau	8,23	0	6,80	0
Lampung	4,52	0	4,23	0
Maluku	6,88	0	6,31	0
Maluku Utara	3,98	0	4,31	0
Nusa Tenggara Barat	2,89	0	2,80	0
Nusa Tenggara Timur	3,54	1	3,14	1
Papua	2,83	1	6,45	1
Papua Barat	5,37	1	5,38	1
Riau	4,37	1	4,23	1
Sulawesi Barat	2,34	1	2,27	1
Sulawesi Selatan	4,51	1	4,33	1
Sulawesi Tengah	3,00	1	2,95	1
Sulawesi Tenggara	3,36	1	3,15	1
Sulawesi Utara	6,61	1	6,10	1
Sumatera Barat	6,28	1	5,94	1
Sumatera Selatan	4,63	1	4,11	1
Sumatera Utara	6,16	1	5,89	1

Tabel 5. Hasil Pengelompokkan Jumlah

Provinsi	Jumlah	cluster
Aceh	37,60	2
Bali	21,46	2
Banten	51,81	2
Bengkulu	21,34	2
DIY Yogyakarta	23,43	2
DKI Jakarta	46,35	2
Gorontalo	20,39	2
Jambi	27,13	2
Jawa Barat	52,30	2
Jawa Tengah	32,04	2
Jawa Timur	29,68	2
Kalimantan Barat	30,32	0
Kalimantan Selatan	27,27	0
Kalimantan Tengah	25,42	0
Kalimantan Timur	37,07	0
Kalimantan Utara	27,49	0
Kep. Bangka Belitung	26,80	0
Kep. Riau	50,82	0
Lampung	26,18	0
Maluku	41,33	0
Maluku Utara	27,59	0
Nusa Tenggara Barat	19,78	0
Nusa Tenggara Timur	20,72	1
Papua	16,95	1
Papua Barat	36,27	1
Riau	31,08	1
Sulawesi Barat	17,05	1
Sulawesi Selatan	30,43	1
Sulawesi Tengah	19,95	1
Sulawesi Tenggara	21,71	1
Sulawesi Utara	39,76	1
Sumatera Barat	36,66	1
Sumatera Selatan	28,03	1
Sumatera Utara	36,23	1



Gambar 6. Data sesudah pengelompokkan

Tabel 5 menunjukkan bahwa *cluster* 0 (C0) mencakup 11 provinsi diantaranya Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kep. Bangka Belitung, Kep. Riau, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat. *Cluster* 1 (C1) mencakup 12 provinsi diantaranya Nusa Tenggara Timur, Papua, Papua Barat, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Barat, Sumatera Selatan, dan Sumatera Utara. *Cluster* 2 (C2) mencakup 11 provinsi diantaranya Aceh, Bali, Banten, Bengkulu, DIY Yogyakarta, DKI Jakarta, Gorontalo, Jambi, Jawa Barat, Jawa Tengah, dan Jawa Timur. Hasil evaluasi pengelompokkan dengan *silhouette coefficient* menghasilkan nilai sebesar 0,54 yang menunjukkan bahwa kriteria pengelompokkan yang dilakukan termasuk dalam struktur *cluster* yang standar (*medium structure*).

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat diambil kesimpulan bahwa pengelompokan provinsi menggunakan algoritma k-means terbagi menjadi 3 *cluster* berdasarkan kemiripan karakteristik indeks pengangguran terbuka di Indonesia. Adapun ketiga *cluster* tersebut antara lain: cluster 0 (C0) yang mencakup 11 provinsi, cluster 1 (C1) yang mencakup 12 provinsi, dan cluster 2 (C2) yang mencakup 11 provinsi. Hasil evaluasi pengelompokan dengan *silhouette coefficient* menghasilkan nilai sebesar 0,54 yang menunjukan bahwa kriteria pengelompokan yang dilakukan termasuk dalam struktur *cluster* yang standar (*medium structure*).

Adapun saran untuk penelitian selanjutnya adalah dengan menggunakan metode selain k-means, seperti k-medoids, Agglomerative Nesting (AGNES), dan lainnya untuk melihat perbandingan hasil sehingga dapat mengetahui evaluasi yang lebih optimal.

DAFTAR PUSTAKA

[1] R. C. Rambe, P. H. Prihanto and H. , "Analisis faktor-faktor yang mempengaruhi pengangguran terbuka di Provinsi Jambi," *e-Jurnal Ekonomi Sumberdaya dan Lingkungan*, vol. Vol. 8. No. 1, pp. 54-67, 2019.

[2] R. Maliki, K. Falgenti, S. Priani, F. Fithri, M. Suherman and D. S. Nugraha, "Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis K-Means Clustering," *Computer Science (CO-SCIENCE)*, vol. Volume 2 No. 2, pp. 109-116, 2022.

[3] T. Amalina, D. B. A. Pramana and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8(15), pp. 574-583, 2022.

[4] I. Fatika, K. Suryowati, N. Pratiwi and M. Sholeh, "Klasifikasi Tingkat Pengangguran Terbuka Di Indonesia Dengan Algoritma Classification and Regression Tree (CART) dan C4.5," *Jurnal Statistika Industri dan Komputasi*, Vols. Volume 07, No. 2, pp. 77-85, 2022.

[5] R. Yulistiani, N. C. Putra, Q. Said and I. Ernawati, "Klasifikasi dan Prediksi Tingkat Pengangguran Terbuka Di Indonesia Menggunakan Classification and Regression Tree (CART)," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, Vols. Vol 1, No 1, pp. 123-130, 2020.

[6] D. S. M. Simanjuntak, I. Gunawan, S. P. and I. P. Sari, "Penerapan Algoritma K-Medoids Untuk Pengelompokkan Pengangguran Umur 25 tahun Keatas Di Sumatera Utara," *Jurnal Krisnadana*, vol. Volume 3 Number 1, pp. 289-309, 2023.

[7] F. A. Tanjung, A. P. Windarto and M. Fauzan, "Penerapan Metode K-Means Pada Pengelompokkan Pengangguran Di Indonesia," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vol. Volume 6 Nomor 1, pp. 61-74, 2021.

[8] J. and A. Junaidi, "Pengaruh produk domestik regional bruto dan pendidikan serta upah terhadap tingkat pengangguran," *Jurnal Ekonomi dan Manajemen*, vol. Volume 20, no. 3, pp. 455-466, 2023.

[9] R. G. Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. Vol. 5 No. 1, pp. 55-60, 2023.

[10] G. I. E. Soen, M. and R. , "Implementasi Cloud Computing dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom Participants," *JITU : Journal Informatic Technology And Communicatione*, Vols. Vol. 6, No. 1, pp. 24-30, 2022.

[11] R. K. Dinata, S. N. Hasdyna and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *Informatics Journal*, vol. Vol. 5 No. 1, pp. 10-17, 2020.

[12] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *Jurnal Ilmiah Informatika (JIF)*, vol. Vol. 09 No. 02, pp. 75-81, 2021.