IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK MENGELOMPOKAN TINGKAT PENGANGGURAN

Nurul Nurjanah [1], Nana Suarna [2], Willy Prihartono [3]
[1,2] Teknik Informatika, STMIK IKMI Cirebon
[3] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
Jalan Perjuangan No. 10B Majasem Kec. Kesambi Kota Cirebon
nurulnurjanah613@gmail.com

ABSTRAK

Di dalam era digital ini, pertumbuhan data ekonomi dan ketersediaan teknologi informatika memberikan peluang untuk mendekati isu kompleks seperti tingkat pengangguran dengan pendekatan yang lebih canggih, dan provinsi Jawa Barat sebagai pusat ekonomi menjadi fokus penelitian ini. Permasalahan dalam penelitian ini adalah bagaimana cara mengelompokkan tingkat jumlah pengangguran di Jawa Barat menggunakan algoritma *clustering K-Means*, berapa *cluster* yang dihasilkan dan berapa nilai titik pusat (*centroid*) nya. Adapun tujuan dari penelitian ini adalah mengelompokkan lebih detail lagi dalam menganalisis data pengangguran di Provinsi Jawa Barat melalui penerapan algoritma *clustering K-Means* untuk mengidentifikasi tiap *cluster* berdasarkan tingkat kabupaten/kota, pendidikan, dan tahun serta menentukan jarak atau titik pusat (*centroid*). Cara kerja analisis data tingkat pengangguran ini menggunakan *tools RapidMiner Studio* dengan menggunakan metode penelitian *kuantitatif* menggunakan algoritma *clustering K-Means*. Hasil yang diperoleh dari penelitian ini yaitu mengelompokan data tingkat pengangguran di Provinsi Jawa Barat menjadi 3 cluster yaitu cluster 0 dengan kategori pengangguran rendah meliputi kota Banjar, cluster 1 dengan kategori pengangguran sedang meliputi kota/kabupaten Kota/kabupaten Sukabumi, Cianjur, Bandung, Garut Tasikmalaya, Ciamis, Kuningan, Cirebon, Sumedang, Indramayu, Subang, Purwakarta, Karawang, dan Bandung Barat sedangkan pada cluster 2 dengan kategori pengangguran tinggi meliputi kota/kabupaten Bogor, Bekasi, Bandung, dan Depok.

Kata kunci : *Implementasi K-means, Clustering*, Pengangguran

1. PENDAHULUAN

Salah satu ukuran utama kesejahteraan ekonomi adalah tingkat pengangguran. Jawa Barat berkontribusi besar pada perekonomian nasional karena menjadi salah satu pusat pertumbuhan ekonomi Indonesia. Meskipun demikian, karena perubahan ekonomi dan dinamika pasar kerja yang terus berubah, masalah mengelola tingkat pengangguran masih menjadi perhatian utama.[1]

Strategi pengelompokan, atau pengelompokan, dapat menjadi suatu metode yang efektif untuk menemukan pola karakteristik dalam data yang berkaitan dengan masalah tingkat pengangguran. algoritma *k-means* adalah salah satu algoritma *clustering* yang telah terbukti efektif. Dengan menggunakan algoritma ini, dapat membagi provinsi Jawa Barat menjadi beberapa wilayah berdasarkan tingkat pengangguran. Ini akan memungkinkan pemangku kepentingan untuk membuat kebijakan yang lebih tepat sasaran.[2]

Adapun permasalahan yang muncul dari penelitian ini yaitu bagaimana mengelompokkan data tingkat pengangguran menggunakan teknik clustering menggunakan algoritma k-means, berapa cluster yang dihasilkan serta berapa jarak titik pusat (centroid) yang dihasilkan dari penelitian ini. Adapun tujuan dari penelitian ini adalah mengelompokkan lebih detail lagi dalam menganalisis data pengangguran di Provinsi Jawa Barat melalui penerapan algoritma *clustering K-Means* untuk mengidentifikasi tiap *cluster* berdasarkan tingkat kabupaten/kota, pendidikan, dan tahun serta menentukan jarak atau titik pusat (*centroid*). Cara kerja analisis data tingkat pengangguran ini menggunakan *tools RapidMiner Studio* dengan menggunakan metode penelitian *kuantitatif* menggunakan algoritma *clustering K-Means*. Hasil yang diperoleh berupa pengelompokan data tingkat pengangguran berdasarkan tingkat kota/kabupaten, tingkat pendidikan dan tahun.

2. TINJAUAN PUSTAKA

2.1. Pengangguran

Menurut Sukirno (1994), pengangguran tidak berarti seseorang tidak bekerja tetapi tidak secara aktif mencari pekerjaan. Sebaliknya, pengangguran adalah keadaan di mana seseorang yang termasuk dalam angkatan kerja ingin mendapatkan pekerjaan tetapi tidak berhasil. Pengeluaran agregat yang rendah adalah faktor utama yang menyebabkan pengangguran. Pengusaha memproduksi barang dan jasa dengan tujuan untuk memperoleh keuntungan, tetapi keuntungan tersebut hanya akan diperoleh apabila mereka dapat menjual barang dan jasa yang mereka buat. Semakin banyak permintaan untuk barang dan jasa yang mereka buat, lebih banyak yang mereka produksi. Penggunaan tenaga kerja akan meningkat sebagai hasil dari peningkatan produksi.[3]

2.2. *Clustering*

Mencari dan mengelompokkan data yang memiliki persamaan, atau kemiripan, satu sama lain, dikenal sebagai clustering. *Clustering* adalah salah satu metode *data mining* yang tidak diawasi *(unsupervised).* Ini berarti bahwa metode ini dapat digunakan tanpa instruksi atau instruktur, serta tanpa target output. *Clustering hierarchical* dan *non-hierarchical* adalah dua jenis *clustering* yang digunakan dalam *data mining*.[4]

2.3. *K-Means*

Salah satu metode pengelompokan data non-hirarki adalah algoritma k-means clustering, yang mengelompokkan objek dengan mengidentifikasi data mana yang akan dikelompokkan terlebih dahulu. Algoritma k-means adalah algoritma yang sederhana dan dapat digunakan untuk data dalam jumlah besar atau kecil. Titik utama setiap cluster ditetapkan secara bebas pada iterasi pertama. Kemudian dihitung jarak data antara tiap titik utama cluster[5]. Seperti yang dijelaskan dalam penelitian, algoritma K-means beroperasi sebagai berikut:

a. Menghitung jumlah cluster (k), yang ditetapkan pada pusat cluster apa pun;
b. Menggunakan persamaan ukuran jarak geometris, hitung jarak antara setiap data ke dalam cluster dengan jarak terpendek:

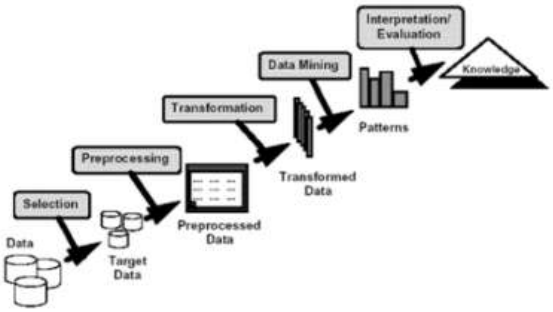
$$(x1,2)= \|x2 - x1\| = \sqrt{\sum} \{x2j = x1 \} p\ 2\ j=1 \ldots\ldots(1)$$

D1(x1, x2) adalah jarak antara data i dan data j, x2j adalah kordinat data x2 pada dimensi j, dan x1j adalah kordinat data x1 pada dimensi j dan P dimensi data.

- Gunakan rumus persamaan
  Vij = ∑ xkj Ni k=1 Ni…… (2)
  untuk menggabungkan data ke dalam kelompok dengan jarak yang paling pendek. Data cluster Vij terdiri dari -i kolom j. Data dari kolom k dan banyaknya anggota cluster i.
- Menggunakan persamaan (1), hitung pusat cluster baru. Ulangi langkah a sampai d hingga tidak ada lagi perpindahan data di cluster lain.

3. METODE PENELITIAN

Penelitian ini memanfaatkan algoritma *K-Means*, yang merupakan algoritma yang efektif untuk mengelompokkan data menjadi berbagai cluster berdasarkan tingkat kemiripan antar data. Dalam penelitian ini, pendekatan penelitian kuantitatif digunakan. Penelitian kuantitatif adalah jenis penelitian yang menggunakan data angka sebagai alat untuk menganalisis informasi yang ingin diketahui. Proses pengolahan data dapat dibagi menjadi beberapa tahap, masing-masing dengan konsep yang berbeda tetapi terkait satu sama lain. Salah satu tahap dalam proses KDD adalah tahap daur ulang data.[6]



Gambar 1. Tahapan *knowledge discovery in databases (kdd)*

Berikut ini adalah tahapan data mining *Knowledge Discovery in Databases (KDD)*, yaitu:

1. *Selection*
   Pilih set data yang di inginkan, atau fokus pada subset variabel atau sampel data. Hasil seleksi disimpan dalam file database yang berbeda. Nama_kabupaten_kota, jumlah_tingkat_pengangguran, pendidikan, satuan (orang), dan tahun akan digunakan untuk proses data mining.[6]
2. *Preprocessing*
   Persiapan awal yang dilakukan adalah membersihkan data (*data cleaning*). Data *cleaning* dilakukan untuk menghilangkan variabel yang tidak digunakan (menghilangkan data *missing value*).[6]
3. *Transformasi* Data
   Mengkonversi data ke format yang dapat diproses kedalam data mining. Operator yang digunakan Nominal *to Numerical*. Langkah ini penting karena *K-Means* merupakan algoritma yang hanya dapat menggunakan data variabel numerik.[6]
4. Data Mining
   Data mining adalah alat yang mengubah data menjadi informasi. Data mining memuat pencarian pola yang diinginkan dalam database besar untuk membantu membuat keputusan di masa depan. Tahap ini merupakan proses utama ketika menerapkan metode untuk menemukan pengetahuan yang berguna dari data. Jenis data dalam penelitian ini adalah clustering yang digunakan untuk mengelompokan data pengangguran berdasarkan tingat pendidian di Kabupaten/Kota Jawa Barat dalam rentang 12 tahun.[6]
5. Interpretasi dan Evaluasi
   Menafsirkan pola yang dihasilkan oleh data mining. Format informasi yang diperoleh dari proses data mining harus disajikan dalam format yang mudah dipahami. Pada tahap evaluasi menggunakan pendekatan *Davies Bouldin Index (DBI)* untuk menentukan cluster (k) yang paling optimum dan dapat mempresentasikan hasil yang ada apakah sudah tercapai atau belum. [6]

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Hasil penelitian yang dilakukan dalam pembahasan ini yaitu akan menguraikan proses bagaimana *clusteriasi* dataset tingkat pengangguran di Jawa Barat menggunakan Algoritma *K-Means.* Pengelompokkan ini dilakukan dengan proses pengujian menggunakan *machine learning* yaitu *Software RapidMiner* Studio[7]. Tahapan analisisnya adalah sebagai berikut :

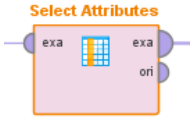
4.1.1. Data

Data yang digunakan dalam penelitian ini adalah data sebanyak 432 *record* dan terdiri 9 atribut pada tahun 2020 sampai dengan 2022 dataset tersebut bersumber dari Open Data Jabar yang bisa dilihat pada gambar 2 berikut ini :

Gambar 2. Data tingkat pengangguran

4.1.2. Data Selection

Langkah kedua dalam analisis data di software *RapidMiner* yaitu menggunakan operator *Select Attributes* untuk memilih attribute mana saja yang akan di pilih untuk tahap analisis data selanjutnya[8]. Operator *Select Attributes* dapat dilihat pada gambar 3 berikut ini :



Gambar 3. Operator *select attributes*

Gambar 4. Hasil *select attribute*

Dari gambar 4 dapat dilihat *attribute* yang di pilih, hasil *selection* dari data yang tidak digunakan hanya digunakan 4 *attribute* karena *attribute* tersebut memiliki keterkaitan yang sama antara satu dengan yang lainnya karena memiliki tujuan yang berkaitan langsung dengan analisis tersebut.

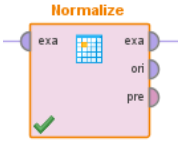
Tabel 1. *Data selection*

Variabel	Tipe Data
nama_kabupaten_kota	*Polynominal*
Pendidikan	*Polynominal*
jumlah_pengangguran	*Integer*
Tahun	*Integer*

Dari tabel 1 Data *Selection* yaitu hasil seleksi *attribute* yang di pilih 4 *attribute* berupa beberapa variabel *attribute* dan tipe data. Variabel nama_kabupaten_kota dan pendidikan memiliki tipe data *Polynominal* karena data yang ada di dalam kolom tersebut memiliki data berupa teks sedangkan variabel kolom jumlah_pengangguran dan tahun memiliki tipe data *Integer* karena data yang berada di dalam kolom tersebut memiliki data berupa angka.

4.1.3. Data Pre-Processing

Langkah kedua dari analisis data di *software RapidMiner* adalah tahap Data *Pre-Processing* yaitu dengan menggunakan operator *Normalize*[9]. Data *pre-processing* di *RapidMiner* menggunakan operator *normalize* bertujuan untuk mengubah nilai-nilai variabel dalam dataset agar memiliki skala seragam. Normalisasi membantu algoritma pembelajaran mesin bekerja lebih efisien dan menghasilkan model yang baik, membantu mengatasi masalah ketidakseimbangan skala antar variabel dan meningkatkan stabilitas dan interpretasi model. Operator *Normalize* dapat dilihat pada gambar 5 berikut ini :



Gambar 5. Operator *normalize*

Adapun parameter yang digunakan dalam operator Normalize pada analisis data di software RapidMiner ini yaitu :

Tabel 2. Parameter operator *normalize*

No	Parameter	Digunakan
1	*Attribute filter type*	*Subset*
2	*Attribute*	jumlah_pengangguran
3		Tahun
4	*method*	*range transformation*
5	min	0.0
6	max	0.1

Setelah melakukan normalisasi, variabel-variabel dalam dataset yang sebelumnya memiliki rentang nilai yang berbeda akan diubah menjadi skala numeric yang seragam. Normalisasi memiliki dampak pada distribusi nilai-nilai variabel, misalnya jika menggunakan metode *min-max scalling*, nilai-nilai akan dibawa ke dalam rentang 0 hingga 1, variabel yang awalnya memiliki perbedaan skala besar akan di sesuaikan sehingga perbedaannya menjadi lebih kecil. Hasil dari proses normalisasi pada attribute

jumlah_pengangguran dan tahun dapat dilihat pada gambar 6 berikut ini :

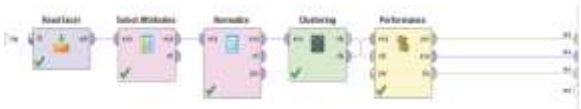
Gambar 6. Hasil normalisasi data

4.1.4. Data Transformasi

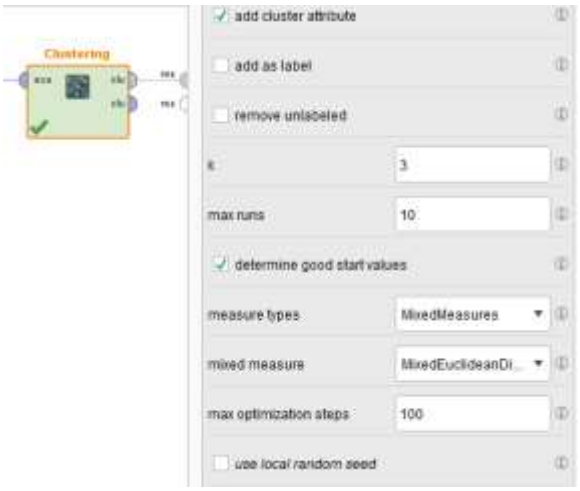
Pada tahap analisis data ini tidak memerlukan tahap transformasi dengan operator *"Nominal to Numerical"* karena tipe data yang sudah sesuai dan dapat diterima oleh model atau analisis yang akan dilakukan maka tidak perlu mengubahnya menjadi *numeric* dan pada tahap sebelumnya sudah melakukan normalisasi atau standarisasi data, variabel *numeric* yang dihasilkan mungkin sudah sesuai dengan kebutuhan tanpa perlu mengubah variabel kategori.[10]

4.1.5. Data Mining

Langkah selanjutnya merupakan lagkah yang paling penting dalam analisis data di *software RapidMiner* meggunakan algoritma *Clustering K-Means* adalah menambahkan operator *K-Means*, yang berfungsi untuk menentukan jumlah cluster pada data tersebut[11]. Operator untuk menganalisis algoritma *K-Means* dapat dilihat pada gambar 7 berikut ini :



Gambar 7. Data mining algoritma *k-means*



Gambar 8. Proses *k-means*

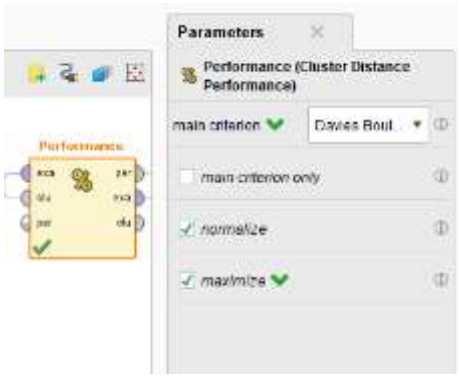
Untuk menentukan berapa *cluster* yang di tentukan pada dataset ini, maka di perlukan operator algoritma k-means dengan parameter k = 3, max run 10, dengan measure types *Mixed Measures*, dengan Mixed Measur *MixedEuclideanDistance*, dengan max optimization steps nya berjumlah 100. Seperti yang ditampilkan pada gambar 8 berikut ini

Dengan operator algoritma *K-Means* dan untuk mengukur besarnya *Davies Bouldin* yang diperoleh dari algoritma *K-Means* menggunakan operator *Cluster Distance Performance*, maka menghasilkan parameter seperti pada tabel 3 sebagai berikut :

Tabel 3. Parameter *k-means*

Parameter	Keterangan
k	3
max runs	10
measure type	*Mixed Measures*
Mixed Measure	*MixedEuclideanDistance*
clustering algorithm	*K-Means*

Sebelum menentukan berapa jumlah *cluster* yang diambil dalam penelitian ini, terlebih dahulu menguji dari K=2 sampai K=10 untuk menentukan nilai K optimal dengan menggunakan operator *Cluster Distance Performance* untuk menentukan nilai *Davies Bouldin Index* terbaik yang mendekati nilai 0. Berikut ini adalah operator *Cluster Distance Performance* dengan *main criterion Davies Bouldin* dapat dilihat pada gambar 9. berikut ini :



Gambar 9. Operator *cluster distance performance*

Parameter yang digunakan pada operator *Cluster Distance Performance* dapat dilihat pada tabel 4 berikut ini :

Tabel 4. Parameter *cluster distance performance*

No.	Parameter	Isi
1	*Main criterion*	*Davies Bouldin*
2	*Normalize*	Digunakan
3	*Maximize*	Digunakan

4.1.6. Interpretasi Data/Evaluation

Teori evaluasi yang digunakan dalam penelitian ini adalah *Davies Bouldin Index* (DBI). Hasil pengujian nilai K=2 sampai K=10 untuk menentukan nilai *Davies Bouldin Index* yang optimal

yang memiliki nilai K mendekati 0 adalah pada K=3 dengan nilai *Davies Bouldin* 0.214. Berikut adalah hasil pengujian K=2 sampai K=10 dapat dilihat pada tabel 5 berikut ini :

Tabel 5. Hasil perhitungan nilai k optimal

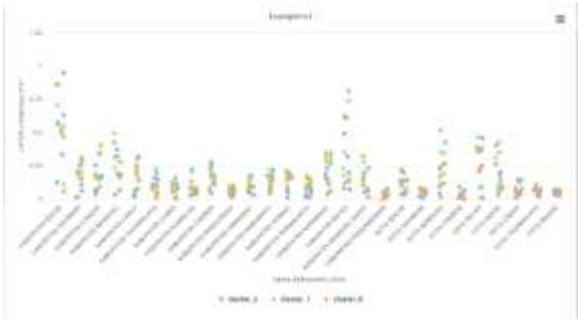
Cluster	*Davies Bouldin Index*
K=2	0.259
K=3	0.214
K=4	0.319
K=5	0.309
K=6	0.282
K=7	0.267
K=8	0.284
K=9	0.254
K=10	0.262

Berdasarkan hasil analisis dengan menggunakan *software RapidMiner* dari data Jumlah Pengangguran di Jawa Barat dari tahun 2020-2022 diperoleh 3 cluster. 3 cluster diperoleh setelah dilakukannya percobaan dari *cluster* 2 sampai *cluster* 10 untuk mencari nilai *Davies Bouldin Indeks*. Setelah melakukan percobaan menghasilkan nilai k optimum K=3 dengan nilai DBi = **0.214**.

4.2. Pembahasan

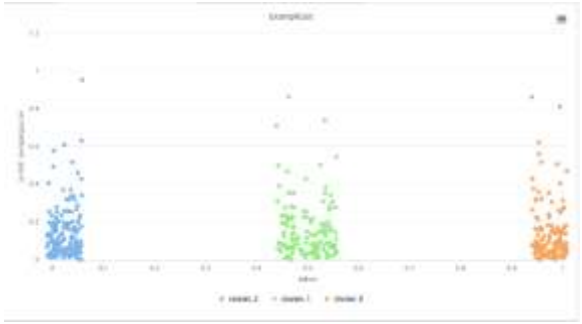
4.2.1. Pengelompokkan Menggunakan Teknik *Clustering*

Penelitian ini berhasil mengimplementasikan teknik *clustering*, dari proses tersebut menghasilkan 3 visualisasi data dalam bentuk Diagram Scatter/Buuble pada gambar 10, gambar 11, dan gambar 12 berikut ini :



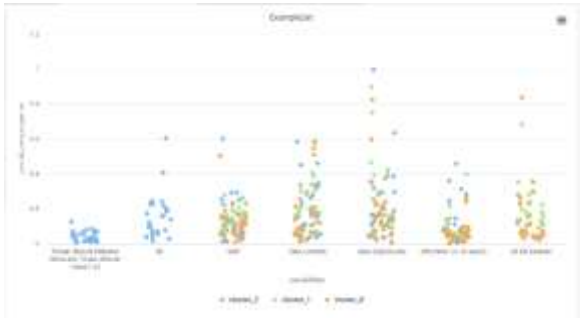
Gambar 10. Hasil visualisasi jumlah pengangguran terhadap kabupaten/kota

Pada gambar 10 menampilkan data pengangguran yang dianalisis berdasarkan masing-masing kabupaten di Jawa Barat. Ini penting untuk mengetahui daerah mana yang memiliki tingkat pengangguran tertinggi dan terendah, serta untuk memahami distribusi pengangguran di tingkat provinsi. Seperti yang ditunjukkan oleh visualisasi data, Kabupaten Bogor menunjukkan tingkat pengangguran tertinggi yang terletak pada cluster 2, dan Kota Banjar menunjukkan tingkat pengangguran paling rendah yang terletak pada cluster 0.



Gambar 11. Hasil visualisasi jumlah pengangguran terhadap tahun

Kemudian pada gambar 11 diatas dengan mengumpulkan data pengangguran berdasarkan tahun tertentu, kita dapat melihat trend pengangguran dan mengetahui apakah ada peningkatan atau penurunan signifikan dalam jangka waktu tertentu. Sebagai contoh, dapat melihat bahwa tingkat pengangguran pada tahun 2020 terletak pada cluster 2 yang berwarna biru meningkat sebagai akibat dari dampak pandemi COVID-19 di Indonesia pada perekonomian dan lapangan pekerjaan, dan pada tahun 2021 terletak pada cluster 1 berwarna hijau dan tahun 2022 terletak pada clster 0 berwarna orange tingkat pengangguran menurun dan dampak terhadap perekonomian mulai stabil.



Gambar 12. Hasil visualisasi jumlah pengangguran terhadap pendidikan

Sedangkan pada gambar 12 diatas menampilkan pengelompokan data berdasarkan tingkat pendidikan yang tersedia untuk pengangguran saat ini. Ini dapat memberikan informasi tentang tingkat pengangguran di setiap tingkat pendidikan di Jawa Barat. Seperti yang ditunjukkan oleh visualisasi data, tingkat pendidikan SMA menunjukkan tingkat pengangguran tertinggi yang terletak pada cluster 2 berwarna biru, sedangkan tingkat pendidikan Tidak/Belum Pernah Sekolah/Tidak/Belum Tamat SD menunjukkan tingkat pengangguran paling rendah yang terletak pada cluster 0 yang berwarna orange.

Berdasarkan hasil analisis Scatter/Bubble yang terdapat pada Gambar 10, 11 dan 12 diatas dapat di simpulkan bahwa analisis tingkat pengangguran di Jawa Barat berdasarkan tingkat kabupaten, tahun dan pendidikan distribusi jumlah pengangguran pada setiap *cluster* menunjukan hampir serupa dengan ketiga kategori tersebut. Hasil analisis jumlah

pengangguran berdasarkan pendidikan diatas dapat ditampilkan dalam bentuk gambar 13 berikut ini :

Gambar 13. Berdasarkan tiap *cluster*

Berdasarkan hasil analisis dataset jumlah pengangguran di Provinsi Jawa Barat menggunakan algoritma *K-Means* yang di visualisasikan dalam bentuk *scatter/buuble* diatas pada gambar 13 diatas dapat di simpulkan bahwa terdapat 3 *cluster* yang berbeda dan terdapat 3 kategori cluster yaitu cluster 0 yang berwarna orange dengan kategori rendah, cluster 1 yang berwarna hijau dengan kategori sedang dan cluster 2 berwarna biru dengan kategori tinggi yang dapat di lihat pada tabel 6 berikut ini :

Tabel 6. Hasil rangkuman tiap *cluster*

Cluster	**Tingkat Pendidikan**	**Kategori Berdasarkan Jumlah Pengangguran**
0	SMP, SMA UMUM dan KEJURUAN, DIPLOMA I/II/III/AKADEMI/UNIVERSITAS dan SD KEBAWAH	RENDAH
1	SMP, SMA UMUM dan KEJURUAN, DIPLOMA I/II/III/AKADEMI/UNIVERSITAS dan SD KEBAWAH	SEDANG
2	TIDAK/BELUM PERNAH SEKOLAH/TIDAK/BELUM TAMAT SD, SD, SMP, SMA UMUM dan KEJURUAN dan DIPLOMA I/II/III/AKADEMI/UNIVERSITAS	TINGGI

4.2.2. *Cluster* yang dihasilkan oleh Algoritma *K-Means*

Hasil pengelompokan jumlah tingkat pengangguran menggunakan k-means diperoleh sebanyak 3 *cluster*, yaitu *cluster* 0 sebanyak 135 items, *cluster* 1 sebanyak 135 items dan cluster 2 sebanyak 162 items dari total sebanyak 432 items. Hasil pengelompokan jumlah *cluster* dapat dilihat pada gambar 14 berikut ini :

Cluster Model

```
Cluster 0: 135 items
Cluster 1: 135 items
Cluster 2: 162 items
Total number of items: 432
```

Gambar 14. Jumlah *cluster*

4.2.3. Jarak Antar *Cluster* Terhadap Titik Pusat *Cluster*

Berdasarkan hasil analisis jumlah tingkat pengangguran di Provinsi jawa Barat jarak antar *cluster* diukur dati titik pusat *cluster* terdekat dan terjauh untuk memahami distribusi atau perbedaan antar *cluster* berdasarkan atribut-atribut yang dianalisis. Jumlah rata-rata dalam jarak *centroid* setiap *cluster* pada tabel 7 berikut ini :

Tabel 7. Jumlah rata-rata dalam jarak *centroid* setiap *cluster*

Cluster	**Jarak Rata-Rata *Centroid***
K Pusat	0.011
K0	0.011
K1	0.011
K2	0.011

Berdasarkan tabel 4.7 diatas, dengan menghitung jarak antara rata-rata *cluster* diperoleh bahwa jarak rata-rata antar *cluster* pada yaitu 0.011 sedangkan jarak rata-rata *cluster* ke titik pusat yaitu *cluster* 0 = 0.011, jarak *cluster* 1 = 0.011 dan jarak *cluster* 2 = 0.011.

5. KESIMPULAN DAN SARAN

Berikut adalah simpulan dari penelitian yang dilakukan terhadapat data pengangguran di Provinsi Jawa Barat dari tahun 2020 hingga 2022 menggunakan metode *k-means clustering* melalui perangkat lunak RapidMiner yaitu sebagai berikut : Algoritma *K-Means* dapat diterapkan pada data pengangguran di Jawa Barat dengan mengkategorikan data berdasarkan tingkat pendidikan. Melalui proses iteratif, algoritma ini mengidentifikasi pusat *cluster* yang optimal yang kemudian digunakan untuk mengelompokkan data pengangguran ke dalam 3 *cluster* yang berbeda. Hasil *clustering* menunjukkan adanya korelasi yang konsisten antara tingkat pendidikan dengan jumlah pengangguran, tahun, dan kabupaten. *Cluster 0* dengan rata-rata tingat pendidikan rendah di dominasi oleh SMP, SMA UMUM dan KEJURUAN, DIPLOMA I/II/III/AKADEMI/UNIVERSITAS dan SD KEBAWAH, *cluster 1* dengan rata-rata tingat pendidikan sedang di dominasi oleh SMP, SMA UMUM dan KEJURUAN, DIPLOMA I/II/III/AKADEMI/UNIVERSITAS dan SD KEBAWAH dan *cluster 2* dengan rata-rata tingat pendidikan tinggi di dominasi oleh TIDAK/BELUM PERNAH SEKOLAH/TIDAK/BELUM TAMAT SD, SD, SMP, SMA UMUM dan KEJURUAN dan

DIPLOMA I/II/III/AKADEMI/UNIVERSITAS. Dapat menganalisa dan pengelompokan dataset jumlah tingkat pengangguran di Provinsi Jawa Barat berdasarkan tingkat pendidikan, kota/kabupaten dan tingkat tahun menggunakan algoritma *k-means*. Dapat pengelompokan menggunakan k-means diperoleh terbaik sebanyak 3 *cluster*, yaitu *cluster* 0 sebanyak 135 item, *cluster* 1 sebanyak 135 item dan *cluster* 2 sebanyak 165 item dari total 432 items. Dapat mengetahui informasi pengelompokan dataset jumlah tingkat pengangguran di provinsi jawa barat terbaik berdasarkan tingkat pendidikan, nilai *Davies Bouldin Index* yang dihasilkan dari algoritma *k-means* ini sebesar 0.214. Dengan menghitung jarak antara rata-rata cluster diperoleh dengan jarak terbaik sebesar 0.011. Hasil pengelompokan data tingkat pengangguran berdasarkan tingkat kabupaten/kota diperoleh bahwa kota/kabupaten Bogor termasuk kategori tingkat pengangguran tinggi dan Kota Banjar termasuk kategori tingkat pengangguran sedang, selanjutnya untuk pengelompokan data tingkat pengangguran berdasarkan tahun diperoleh bahwa pada tahun 2020 merupakan tingkat pengangguran paling tinggi akibat dari adanya COVID 19 di Indonesia kemudian dari tahun 2021-2022 jumlah pengangguran menurun dan perekonomian mulai stabil kembali. Sedangkan untuk pengelompokan data tingkat pengangguran berdasarkan pendidikan tingkat SMA merupakan tingkat pengangguran paling tinggi dan untuk pendidikan DIPLOMA I/II/III/AKADEMIK/UNIVERSITAS merupakan tingkat pengangguran rendah.

Peneliti mengharapkan dapat digunakan dalam penelitian-penelitian selanjutnya, Pada penelitian selanjutnya di sarankan untuk menggunakan dataset dengan *attribute* yang lebih banyak diantaranya attribute jenis kelamin, umur, jumlah lapangan pekerjaan, skill/ keterampilan yang di miliki dan lain sebagainya agar hasil yang diperoleh lebih akurat. Di karenakan keterbatasan waktu dan energi, peneliti menyadari bahwa hasil penelitian ini masih jauh dari kata kesempurnaan. Oleh karena itu, dalam penelitian selanjutnya dapat di lakukan perbandingan dengan algoritma clustering yang berbeda *misalnya k-medoids, c-means*, dan lain sebagainya.

DAFTAR PUSTAKA

[1] I. M. Nur, M. Rizky, and S. P. Milasari, "Pengelompokan Tingkat Kemiskinan di Provinsi Jawa Barat dengan Metode K-Means Clustering," vol. 1, no. 2, pp. 51–61, 2023.

[2] N. Fadilah *et al.*, "Penerapan Metode Algoritma K-Means Untuk Clustering," vol. 6, no. 1, pp. 1–5, 2022.

[3] S. Sonang, A. T. Purba, and F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means," *J. Tek. Inf. dan Komput.*, vol. 2, no. 2, p. 166, 2019, doi: 10.37600/tekinkom.v2i2.115.

[4] C. Commons and P. J. Barat, "Penerapan algoritma k-means pada data pengangguran di jawa barat," vol. 3, no. 1, pp. 23–34, 2023.

[5] Y. Nurohmah *et al.*, "Optimalisasi Performa K-Means Clustering Dengan PCA Dalam Analisis Tingkat Kemiskinan Di Jawa Barat," vol. 7, no. 3, pp. 1657–1665, 2023.

[6] N. I. Pastia and F. N. Dikananda, "Pengelompokan Data Pengangguran Terbuka Menggunakan Algoritma K-Means Berdasarkan Provinsi Jawa Barat," *J. Din. Inform.*, vol. 12, no. 1, pp. 59–69, 2023.

[7] A. Aziz *et al.*, "Klastrisasi Data Tenaga Kerja Terbuka Menurut Provinsi Dengan Penggunaan Algoritma K - Means 1,2," vol. 10, no. 3, pp. 444–456, 2023.

[8] D. Harits, A. A. Puji, M. Andivas, D. Dermawan, and E. A. Thoriq, "Analisis Klaster Tingkat Pengangguran di Provinsi Kalimantan Timur Menggunakan Algoritma K-Means," *J. Surya Tek.*, vol. 9, no. 2, pp. 456–460, 2022, doi: 10.37859/jst.v9i2.4430.

[9] A.- Akramunnisa and F. Fajriani, "K-Means Clustering Analysis pada PersebaranTingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan," *J. Varian*, vol. 3, no. 2, pp. 103–112, 2020, doi: 10.30812/varian.v3i2.652.

[10] D. Safira, M. Mustakim, E. D. Lestari, M. Iffa, and S. Annisa, "Pengelompokan Jumlah Penduduk Sumatera Barat Berdasarkan Angkatan Kerja Menggunakan Algoritma K-Means," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 6, no. 1, p. 26, 2020, doi: 10.24014/rmsi.v6i1.8682.

[11] F. A. Tanjung, A. P. Windarto, and M. Fauzan, "Penerapan Metode K-Means Pada Pengelompokkan Pengangguran Di Indonesia," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, p. 61, 2021, doi: 10.30645/jurasik.v6i1.271.