

Pengelompokan Kabupaten/Kota Berdasarkan Indikator Tingkat Pengangguran Menggunakan Algoritma K-Means Clustering (Studi Kasus: Provinsi Jawa Barat)

Anggia Arfiani Putrie¹, Rangga Sanjaya²

^{1,2}Program Studi Sistem Informasi, Universitas Adhirajasa Reswara Sanjaya
e-mail: ¹arfianianggia6@gmail.com, ²rangga@ars.ac.id

Abstrak

Permasalahan yang selalu muncul setiap tahunnya adalah masalah pengangguran. Pada tahun 2018 tingkat pengangguran terbuka di Provinsi Jawa Barat sebesar 8,23%, pada tahun 2019 persentase pengangguran terbuka di Jawa Barat berkurang menjadi 8.04% dan pada tahun 2020 bertambah menjadi 10.46%. Oleh karena dalam penelitian ini membahas tentang persoalan pengangguran berdasarkan indikator tingkat pengangguran khususnya jumlah penduduk dan upah minimum kabupaten/kota, kemudian dikelompokkan menggunakan metode K-Means Clustering berdasarkan wilayah/kota. Tujuan penelitian ini untuk mendapatkan hasil pengelompokan Kabupaten/Kota di Provinsi Jawa Barat berdasarkan indikator tingkat pengangguran. Berdasarkan hasil penelitian diperoleh 4 cluster, yaitu cluster 0 menunjukkan daerah dengan tingkat pengangguran paling rendah terdapat 18 kabupaten/kota Cluster 1 menunjukkan daerah dengan tingkat pengangguran rendah terdapat 2 kabupaten/kota. Cluster 2 menunjukkan daerah dengan pengangguran tinggi terdapat 1 kabupaten, dan cluster 3 menunjukkan daerah dengan tingkat pengangguran sedang terdapat 2 Kota. Pada tahap evaluasi menggunakan metode Davies Bouldin Index pada rapid miner menghasilkan Davies Bouldin sebesar 0.328 artinya hasilnya baik karena hasilnya 0 (non-negatif).

Kata kunci— Pengelompokan, Pengangguran, K-Means, Rapidminer, Davies Bouldin Index.

Abstract

The problem that arises every year is the problem of unemployment. In 2018 the open unemployment rate in West Java Province was 8.23%, in 2019 the percentage of open unemployment in West Java decreased to 8.04% and in 2020 increased to 10.46%. Therefore, in this study discussed the problem of unemployment based on indicators of the unemployment rate, especially the number of residents and minimum wage of districts/cities, then grouped using the K-Means Clustering method by region/city. The purpose of this study is to get the results of regency/city grouping in West Java Province based on indicators of unemployment rate. Based on the results of the study obtained by 4 clusters, namely cluster 0 shows the area with the lowest unemployment rate there are 18 districts/cities Cluster 1 shows areas with low unemployment rates there are 2 districts/cities. Cluster 2 shows areas with high unemployment.

Keywords— Grouping, Unemployment, K-Means, Rapidminer, Davies Bouldin Index

Corresponding Author:

Rangga Sanjaya,

Email: rangga@ars.ac.id

1. PENDAHULUAN

Indonesia dianggap sebagai negara. Indonesia adalah salah satu negara dengan potensi pasar paling besar dalam hal ketersediaan tenaga kerja, dan itu harus menjadi kekuatan pendorong dalam perekonomian. Namun, kesulitan di Indonesia adalah bahwa populasi dan

tenaga kerja di negara berkembang menjadi beban bagi pembangunan negara daripada aset potensial yang dapat digunakan untuk mendorong ekonomi. Besarnya tingkat pengangguran berdampak pada laju keberhasilan pembangunan ekonomi[1]. Pengangguran merupakan masalah utama di negara-negara non-industri seperti Indonesia, dengan pertumbuhan penduduk yang tinggi, akses ke pekerjaan yang tidak merata, penduduk usia kerja yang tidak dapat mengikuti kemajuan teknologi yang pesat, tingkat pendidikan, perubahan ekonomi, dll. Struktur dan ukuran upah yang disebabkan oleh berbagai faktor[2, pp. 2010–2014].

Jawa Barat, salah satu provinsi di Indonesia, masuk dalam kategori pengangguran yang jumlahnya sering menurun, namun masih cukup tinggi. Jumlah pengangguran dapat meningkat setiap tahunnya. Tingkat pengangguran terbuka di Jawa Barat pada tahun 2018 menurun menjadi 8,23%, tingkat pengangguran terbuka di Jawa Barat pada tahun 2019 menurun menjadi 8,04% dan meningkat menjadi 10,46% pada tahun 2020[3, pp. 2015–2020]. Pengangguran yang tinggi mempengaruhi tingkat kemiskinan di beberapa daerah[4]. Akibatnya, karena pengangguran adalah salah satu penentu utama tingkat kemiskinan, penelitian yang lebih mendalam tentang pengangguran dan indikator yang mempengaruhi tingkat pengangguran perlu dilakukan. Memanfaatkan metode pengelompokan salah satunya.

Clustering adalah salah satu teknik analisa *data mining*[5], Menurut Widodo (2013:9)[6] *Clustering* adalah teknik untuk membagi dataset menjadi kelompok-kelompok berdasarkan kesamaan yang telah ditentukan. Metode KMeans merupakan metode pengelompokan yang mengelompokkan data berdasarkan pusat cluster (centroid) yang paling dekat dengan data tersebut. K Means bertujuan untuk mengorganisasikan data dengan meningkatkan kesamaan data dalam kelompok dan mengurangi kesamaan data antar cluster[7], [8].

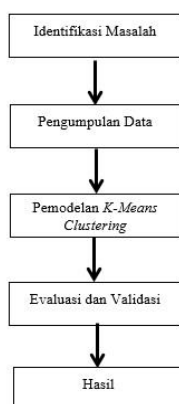
Pada penelitian yang telah dilakukan sebelumnya, yaitu pengelompokan tingkat pengangguran di Sulawesi Selatan dengan indikator TPT, UMK, dan IPM, penelitian ini kemudian akan memanfaatkan indikator jumlah penduduk dan upah minimum kabupaten/kota disamping jumlah pengangguran sebagai acuan utama. Data yang digunakan adalah data Badan Pusat Statistik (BPS) Jawa Barat tahun 2018-2020.

Hasil pengelompokan yang diperoleh dapat digunakan untuk menjelaskan bagaimana pemerintah daerah menangani masalah pengangguran dengan mengkaji faktor-faktor yang mempengaruhi pengangguran, seperti jumlah penduduk dan upah.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Tahap ini, akan menjelaskan gambaran metode penelitian yang digunakan. Gambar 1 menggambarkan langkah-langkah yang diambil pada penelitian ini[9].



Gambar 1. Desain Penelitian

1. Identifikasi Masalah

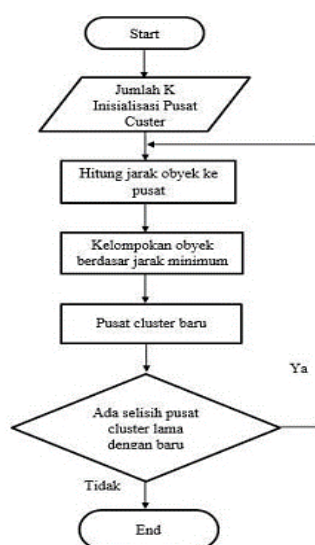
Bagaimana menerapkan algoritma *K-Means Clustering* untuk mengidentifikasi pengelompokan daerah pengangguran sesuai dengan indikator pengangguran yang digunakan.

2. Pengumpulan Data

Penulis menggunakan website resmi Badan Pusat Statistik (BPS) Jawa Barat <https://jabar.bps.go.id> untuk melakukan pengamatan langsung terhadap apa yang sedang diteliti. Data yang diperoleh adalah pengangguran, jumlah penduduk dan upah minimum periode 2018-2020.

3. Pemodelan *K-Means Clustering*

Pemodelan yang digunakan terhadap objek penelitian yaitu menentukan jumlah *cluster* dalam data, kemudian menentukan nilai pusat atau *centroid*, dan terakhir menggunakan jarak *Euclidean Distance* untuk menghitung jarak terdekat ke *centroid* untuk mendapatkan kelompok terbaik.



Gambar 2. Flowchart *K-Means*

4. Evaluasi dan Validasi

Langkah selanjutnya adalah evaluasi *cluster* yang dihasilkan berdasarkan jumlah dan kedekatan antar data hasil *cluster*. Proses evaluasi disini menggunakan *Davies-Bouldin Index (DBI)*.

5. Hasil

Hasil analis data yang diproses sebelumnya akan diuji pada tahap ini. Adapun mekanisme pengujian hasil adalah sebagai berikut:

- Dataset pertama akan dikumpulkan berupa statistik pengangguran, jumlah penduduk dan upah minimum di wilayah/kota Jawa Barat dari tahun 2018 hingga 2020. Data berikut dievaluasi menggunakan algoritma *K-Means*.
- Data yang telah dievaluasi, dimasukan menggunakan kriteria pada metode *K-Means* dan dihitung secara manual.
- Data yang di analisis kemudian akan di proses dan diperiksa menggunakan aplikasi *Rapidminer* (alat pengukuran pengujian, tingkat kesalahan/akurasi)
- Hasil yang didapat dari perhitungan data secara manual menggunakan kriteria metode *K-Means*, kemudian dibandingkan dengan perhitungan data yang dilakukan menggunakan program *Rapidminer*.

2.2. Instrumen Penelitian

Berdasarkan permasalahan yang dijelaskan sebelumnya, maka bahan dan peralatan yang dibutuhkan untuk penelitian ini adalah:

a. Dataset

Terdapat 23 data dalam dataset yang digunakan dalam penelitian ini, antara lain pengangguran, jumlah penduduk dan Upah Minimum Kabupaten/Kota tahun 2018-2020 sebanyak 243 data. Data tersebut di dapatkan dari hasil riset *website* resmi Badan Pusat Statistik (BPS) Jawa Barat.

b. Peralatan

Dibawah ini merupakan kebutuhan dari penelitian diantaranya:

- 1) *Microsoft office excel professional 2016*, aplikasi ini digunakan penulis sebagai media penulisan dan hitung manual dari *dataset* yang diperoleh.
- 2) *Rapidminer versi 9.9* merupakan aplikasi yang digunakan untuk menguji keakuratan dalam pengelompokkan pengangguran pada provinsi Jawa Barat memanfaatkan teknik *data mining* yaitu metode *K-Means clustering* terhadap *dataset*.

2.3 Metode Pengumpulan Data

Data penelitian ini dikumpulkan menggunakan teknik pengamatan (observasi) di situs resmi Badan Pusat Statistik (BPS) Jawa Barat.

2.4 Variabel Data

Pengangguran, jumlah penduduk, dan upah minimum di 27 Kabupaten/Kota tahun 2018-2020 di Provinsi Jawa Barat merupakan variabel data yang diperoleh untuk penelitian ini.

2.5 Metode Analisis Data

Pendekatan KDD, yang terdiri dari Sembilan Langkah Maimon dan Rokatch dalam (Muningsih & Kiswati, 2018), digunakan untuk menganalisis data dalam penelitian ini yaitu:

1. Tahap *Domain Understanding and KDD Goals*, menurut situs resmi Badan Pusat Statistik, Provinsi Jawa Barat memiliki populasi terbesar di Indonesia, dengan 48.274.162 juta orang, dan juga menjadi salah satu provinsi dengan tingkat pengangguran tertinggi kedua. Mengenai pengangguran, yang masih mengalami naik turun. Tingkat pengangguran Jawa Barat sebesar 8,23 persen pada 2018, 8,04 persen pada 2019, dan 10,46 persen pada 2020, dengan peningkatan yang cukup besar sebesar 10,46 persen pada 2020. Pemerintah menghadapi tugas berat dalam mengurangi pengangguran di setiap kabupaten/kota di Provinsi Jawa Barat.
2. Tahap *selection and addtion*, yaitu mengumpulkan data pengangguran, jumlah penduduk, dan upah minimum kab/kota Provinsi Jawa Barat. Nilai rata-rata setiap atribut yang digunakan menjadi fokus pada penelitian ini.

Tabel 1. Rata-Rata Pengangguran, Jumlah Penduduk, dan Upah Minimum Kab/Kota

Kabupaten/Kota	Pengangguran	Jumlah Penduduk	Upah Minimum Kab/Kota
Kabupaten Bogor	147322	2348760	377691
Kabupaten Sukabumi	42779	2465728	265933
Kabupaten Cianjur	50759	2262673	234439
Kabupaten Bandung	55194	3774692	290345
Kabupaten Garut	40042	2621820	181377
Kabupaten Tasikmalaya	26269	1753711	195930
Kabupaten Ciamis	14857	1195163	173938
Kabupaten Kuningan	22609	1080802	174122
Kabupaten Cirebon	47980	2192916	203909

Kabupaten Majalengka	14886	1205014	179812
Kabupaten Sumedang	23929	1152245	290346
Kabupaten Indramayu	33831	1728427	212532
Kabupaten Subang	30714	1595806	274271
Kabupaten Purwakarta	18870	9627320	373566
Kabupaten Karawang	51724	2353471	425221
Kabupaten Bekasi	81417	3764603	416101
Kabupaten Bandung Barat	34750	1699530	290915
Kabupaten Pangandaran	4808	3993210	171135
Kota Bogor	25044	1111945	385658
Kota Sukabumi	6521	3285510	234012
Kota Bandung	55951	2507233	335157
Kota Cirebon	6675	3193040	205276
Kota Bekasi	62332	3003837	424494
Kota Depok	44181	2407115	388645
Kota Cimahi	14601	6141690	290346
Kota Tasikmalaya	10626	6634090	209024
Kota Banjar	2587	1830760	169428

3. Tahap *preprocessing and data cleansing*, pembersihan data akan dilakukan pada tahap ini, dan data yang hilang akan diisi menggunakan operator yang diberikan oleh rapidminer versi 9.9, yaitu mengganti nilai yang hilang. Namun proses ini tidak dilakukan karena ada data yang hilang.
4. Tahap *transformasi* data, dengan mengkonversi data ke format yang dapat diproses kedalam *data mining*, karena data tersebut sesuai untuk di proses maka tidak dilakukan tahap *transformasi*.
5. Tahap *data mining*, jenis data yang digunakan dalam penelitian ini adalah *clustering*, yang digunakan untuk mengelompokkan pengangguran berdasarkan jumlah penduduk dan upah minimum di setiap Kabupaten/Kota di provinsi Jawa Barat selama 3 tahun.
6. Tahap *data mining*, Penelitian ini menggunakan teknik clustering *K-Means* untuk mengklasifikasikan pengangguran pada berbagai tingkatan berdasarkan jumlah penduduk dan upah minimum Jawa Barat.
7. Tahap *data mining*, penggunaan algoritma *data mining* yaitu dengan menerapkan algoritma yang telah ditentukan sebelumnya.
 - a. Menentukan jumlah cluster (k) pada *dataset*, tahap ini akan dibuat sebanyak 4 *cluster* yaitu tingkat tinggi, sedang, rendah, dan sangat rendah.
 - b. Menentukan titik pusat (*centroid*) awal secara acak
8. Tahap evaluasi, pada langkah ini, kelompok atau cluster yang ada di evaluasi dengan memanfaatkan pendekatan *davies bouldin index* (DBI) untuk menentukan jarak terdekat antara data hasil cluster
9. Tahap *Discovered Knowledge (Visualization dan Integration)*, Pada fase ini, teknik clustering KMeans digunakan untuk mengelompokkan data pengangguran terhadap kelompok kabupaten/kota Jawa Barat berdasarkan jumlah penduduk kabupaten/kota

tahun 2018-2020 dan upah minimum. Mengelompokkan kategori tinggi, sedang, rendah, dan sangat rendah

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan *K-Means*

1. Perhitungan pengelompokan *K-Means Clustering* adalah menentukan *cluster* (*k*) dalam dataset, tahap terdiri dari 4 *cluster* yaitu *cluster C0, C1, C2, C3*.
2. Menentukan titik pusat (*centroid*) awal secara acak kemudian menghitung *centroid cluster k-1*:

Tabel 2. Centroid Awal

Attribute	C0	C1	C2	C3
Pengangguran	147322	55194	18870	10626
Jumlah Penduduk	2348760	3774692	9627320	6634090
UMK	377691	290345	373566	209024

Setelah menentukan centroid awal, *cluster* ditentukan dengan menggunakan rumus jarak *Eulidean Distance* untuk menghitung jarak setiap data yang ada ke *centroid* terdekat. Berikut contoh cara menghitung setiap *centroid*.

- Jarak objek data ke-1 dari *cluster (C0)*

$$= \sqrt{(147322 - 147322)^2 + (2348760 - 2348760)^2 + (377691 - 377691)^2} = \sqrt{0}$$

$$= 0$$
- Jarak objek data ke-1 dari *cluster (C1)*

$$= \sqrt{(147322 - 55194)^2 + (2348760 - 3774692)^2 + (377691 - 290345)^2} = \sqrt{2,049,398} = 1,431,572$$
- Jarak objek data ke-1 dari *cluster (C2)*

$$= \sqrt{(147322 - 11870)^2 + (2348760 - 9627320)^2 + (377691 - 373566)^2} = \sqrt{52,977,460}$$

$$= 7,279,561$$
- Jarak objek data ke-1 dari *cluster (C3)*

$$= \sqrt{(147322 - 10626)^2 + (2348760 - 6634090)^2 + (377691 - 209024)^2} = \sqrt{18,411,187}$$

$$= 4,290,825$$

Kemudian melakukan perhitungan yang sama dari data selanjutnya sampai data ke 27 dan mencari jarak terpendeknya. Berikut hasil iterasi pertama:

Tabel 3. Hasil Pengelompokan Iterasi 1

Kabupaten/Kota	C0	C1	C2	C3
Kabupaten Bogor	1			
Kabupaten Sukabumi	1			
Kabupaten Cianjur	1			
Kabupaten Bandung		1		
Kabupaten Garut	1			

Kabupaten/Kota	C0	C1	C2	C3
Kabupaten Tasikmalaya	1			
Kabupaten Ciamis	1			
Kabupaten Kuningan	1			
Kabupaten Cirebon	1			
Kabupaten Majalengka	1			
Kabupaten Sumedang	1			
Kabupaten Indramayu	1			
Kabupaten Subang	1			
Kabupaten Purwakarta			1	
Kabupaten Karawang	1			
Kabupaten Bekasi		1		
Kabupaten Bandung Barat	1			
Kabupaten Pangandaran		1		
Kota Bogor	1			
Kota Sukabumi		1		
Kota Bandung	1			
Kota Cirebon		1		
Kota Bekasi	1			
Kota Depok	1			
Kota Cimahi				1
Kota Tasikmalaya				1
Kota Banjar	1			

Setelah melakukan pengelompokan hasil dari iterasi 1 didapat, *cluster* 0 sebanyak 19 kab/kota, *cluster* 1 sebanyak 5 kab/kota, *cluster* 2 sebanyak 1 kab, dan *cluster* 3 sebanyak 2 kota dan didapatkan hasil *centroid* baru untuk melakukan iterasi berikutnya.

Tabel 4. Centroid Baru

Attribute	C0	C1	C2	C3
Pengangguran	40660	30923	18870	12614
Jumlah Penduduk	1921945	3602211	9627320	6387890
UMK	277833	263374	373566	249685

Lakukan perhitungan *Euclidean Distance* dengan *centroid* baru sampai menghasilkan pola yang sama dari iterasi sebelumnya dengan yang terakhir. Pada proses ini berhenti pada iterasi ke tiga dan menghasilkan sebagai berikut:

Tabel 5. Jumlah pada centroid

Iterasi	C0	C1	C2	C3	Total
1	19	5	1	2	27
2	18	6	1	2	27
3	18	6	1	2	27

Pada penelitian ini proses iterasi terhenti pada iterasi 3, seperti pada penjelasan proses *k-means* bahwa iterasi berhenti dan dianggap stabil jika hasil iterasi baru sama dengan hasil iterasi sebelumnya. Maka proses iterasi sudah dianggap stabil karena hasil iterasi 2 dan 3 sudah sama.

3.2 Karakteristik Cluster

- Cluster 0

Rata-rata setiap variabel (*centroid*) cluster 0

Pengangguran	39456.333
Jumlah Penduduk	1861840
Upah Minimum Kab/Kota	264406.889
Rata-rata seluruh variabel cluster 0	721901.056

Berdasarkan rata-rata pada cluster 0, nilai *centroid* untuk pengangguran yaitu sebesar 39456.333, sedangkan untuk jumlah penduduk yaitu sebesar 1861840, dan untuk upah minimum Kab/Kota yaitu sebesar 264406.889. Maka rata-rata cluster yang diperoleh dari seluruh variabel di cluster 0 yaitu sebesar 721901.056. Anggota cluster 0 antara lain Kab. Bogor, Kab. Sukabumi, Kab. Cianjur, Kab. Garut, Kab. Tasikmalaya, Kab. Ciamis, Kab. Kuningan, Kab. Cirebon, Kab. Majalengka, Kab. Sumedang, Kab. Indramayu, Kab. Subang, Kab. Karawang, Kab. Bandung Barat, Kota Bogor, Kota Bandung, Kota Depok, dan Kota Banjar.

- Cluster 1

Rata-rata setiap variabel (*centroid*) cluster 1

Pengangguran	36157.833
Jumlah Penduduk	3502482
Upah Minimum Kab/Kota	290227.167
Rata-rata seluruh variabel cluster 1	1276289.000

Berdasarkan rata-rata pada cluster 1, nilai *centroid* untuk pengangguran yaitu sebesar 36157.833, sedangkan untuk jumlah penduduk yaitu sebesar 3502482, untuk upah minimum Kab/Kota yaitu sebesar 290227.167. Maka rata-rata cluster yang diperoleh dari seluruh variabel di cluster 1 yaitu sebesar 1276289.000. Anggota cluster 1 antara lain Kab. Bandung, Kab. Bekasi, Kab. Pangandaran, Kota Sukabumi, Kota Cirebon, dan Kota Bekasi.

- Cluster 2

Rata-rata setiap variabel (*centroid*) cluster 2

Pengangguran	11870
Jumlah Penduduk	9627320
Upah Minimum Kab/Kota	373566
Rata-rata seluruh variabel cluster 2	3339918.667

Berdasarkan rata-rata pada cluster 2, nilai *centroid* untuk pengangguran yaitu sebesar 11870, sedangkan untuk jumlah penduduk yaitu sebesar 9627320, dan untuk upah minimum Kab/Kota yaitu sebesar 373566. Maka rata-rata cluster yang diperoleh dari seluruh variabel di cluster 2 yaitu sebesar 3339918.667. Anggota cluster 2 yaitu Kab. Purwakarta.

- Cluster 3

Rata-rata setiap variabel (*centroid*) cluster 3

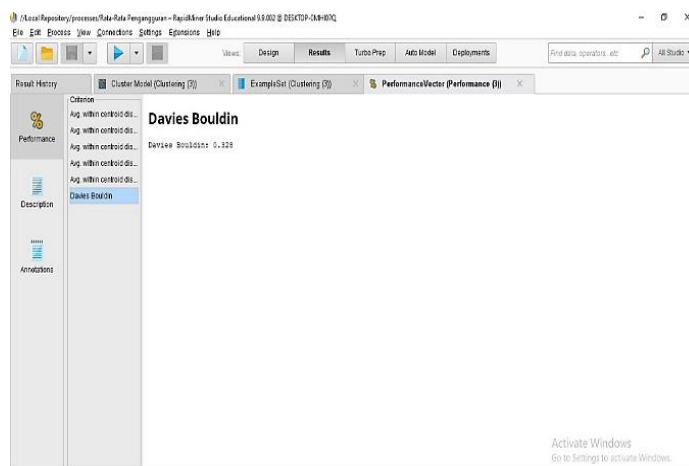
Pengangguran	12613.5
Jumlah Penduduk	6387890
Upah Minimum Kab/Kota	294685
Rata-rata seluruh variabel cluster 3	2216729.500

Berdasarkan rata-rata pada cluster 3 nilai *centroid* untuk pengangguran yaitu sebesar 12613.5, sedangkan untuk jumlah penduduk yaitu sebesar 6387890, dan untuk Upah Minimum

Kota yaitu sebesar 294685. Maka rata-rata *cluster* yang diperoleh dari seluruh variabel di *cluster* 3 yaitu sebesar 2216729.500. Anggota *cluster* 3 antara lain Kota Cimahi dan Kota Tasikmalaya.

3.3 Validasi Hasil

Berdasarkan validasi hasil pada aplikasi *Rapidminer* menggunakan *Davies Bouldin Index*, terbentuk 4 *cluster* yang dianggap mampu mengelompokan data dengan baik. Gambar 3 menggambarkan *Performance vector*, yang merupakan hasil evaluasi dari *cluster* yang dibuat sebanyak 4 *cluster* dan memiliki nilai 0,328.



Gambar 3. Performance Vector Davies Bouldin Index

Sebelumnya telah dilakukan pengujian $k=2$ sampai $k=5$. Nilai *Davies Bouldin Index* $k=4$ didapat sebesar 0,328. *Davies Bouldin Index* bisa dikatakan baik jika hasil pengelompokan semakin kecil atau minimal [10].

Tabel 6. Hasil pengujian *DBI* $k=2$ sampai $k=5$ menggunakan aplikasi *Rapidminer*

Cluster	Nilai Davies Bouldin
K=2	0.416
K=3	0.539
K=4	0.328
K=5	0.331

3.3 Pembahasan

Menurut hasil pengelompokan *K-Means* menggunakan aplikasi *Rapidminer*, *cluster* 0 adalah cluster dengan nilai rata-rata terendah dari semua variabel dibandingkan dengan cluster lainnya. Maka *cluster* 0 termasuk dalam kategori kelompok pengangguran paling rendah, yang di dalamnya terdapat 18 Kabupaten/Kota dengan nilai rata-rata sebesar 721901.056. *Cluster* 1 termasuk dalam kategori kelompok pengangguran rendah yang di dalamnya terdapat 6 Kabupaten/Kota dengan rata-rata sebesar 1276289.000. *Cluster* 2 termasuk dalam kelompok pengangguran tinggi yaitu terdapat 1 Kabupaten dengan rata-rata sebesar 3339918.667. *Cluster* 3 termasuk dalam kategori kelompok rendah yang di dalamnya terdapat 2 Kota dengan rata-rata sebesar 2216729.500.

Dari 4 *cluster* yang terbentuk *cluster* 2 yang harus diberi perhatian lebih oleh pemerintah karena *cluster* tersebut memiliki rata-rata tertinggi. Metode untuk menurunkan

tingkat pengangguran pada *cluster* ini diantaranya pemerintah melakukan evaluasi upah minimum kab/kota pekerja sehingga dapat menetapkan kebijakan yang berdampak pada peningkatan kesejahteraan dan taraf hidup layak pekerja, pemerintah juga dapat menciptakan lapangan kerja yang dibutuhkan dan menyerap sejumlah besar pekerja untuk mengurangi tingkat pengangguran saat ini.

4. KESIMPULAN

Maka, dengan menggunakan Algoritma *K-Means* Clustering, kesimpulan yang dapat ditarik mengenai pengelompokan pengangguran berdasarkan jumlah penduduk dan Upah Minimum di Kabupaten/Kota Provinsi Jawa Barat tahun 2018-2020 yaitu:

1. Hasil pengolahan pengangguran menurut indikator pengangguran, yaitu jumlah penduduk dan Upah Minimum Kabupaten/Kota dengan menggunakan algoritma *K-Means Clustering* dapat menghasilkan informasi, kelompok daerah mana saja yang memerlukan perhatian khusus dari pemerintah dalam menciptakan dan meratakan lapangan kerja, mengevaluasi Upah Minimum Kabupaten/Kota dan memperlambat laju pertumbuhan penduduk di Kabupaten/Kota dengan tingkat pengangguran yang tinggi.
2. Hasil *clustering* pada *rapid miner* diperoleh *cluster* 0 (pertama) terdapat 18 Kabupaten/Kota dengan rata-rata pengangguran paling rendah, *cluster* 1 (kedua) terdapat 6 Kabupaten/Kota dengan rata-rata pengangguran rendah, *cluster* 2 (ketiga) terdapat 1 Kabupaten dengan rata-rata pengangguran tinggi dan *cluster* 3 (keempat) terdiri dari 2 Kota dengan rata-rata pengangguran sedang.
3. Berdasarkan evaluasi *Davies Bouldin Index* (DBI) pembentukan 4 *cluster* telah dinilai baik, karena dengan uji validasi pembentukan jumlah k (*cluster*) 2-5 k (*cluster*), *cluster* 4 memiliki nilai yang paling dekat dengan 0, yaitu 0.328.

DAFTAR PUSTAKA

- [1] S. Prawira, "Pengaruh pertumbuhan ekonomi, upah minimum provinsi, dan tingkat pendidikan terhadap pengangguran terbuka di Indonesia," *J. Ecogen*, vol. 1, no. 1, pp. 162–168, 2018.
- [2] T. B. Hartanto, "Analisis Pengaruh Jumlah Penduduk, Pendidikan, Upah Minimum Dan Produk Domestik Regional Bruto (Pdrb) Terhadap Jumlah Pengangguran Di Kabupaten Dan Kotaprovisi Jawa Timur Tahun 2010-2014," *JJET J. Ilmu Ekon. Terap.*, vol. 2, no. 1, 2017.
- [3] K. Berliani, "Pengaruh Tingkat Pengangguran, Tingkat Pendidikan dan Laju Pertumbuhan Penduduk Terhadap Tingkat Kemiskinan Penduduk Provinsi Jawa Barat Tahun 2015-2020," *Syntax Lit. J. Ilm. Indones.*, vol. 6, no. 2, pp. 872–882, 2021.
- [4] I. Itang, "Faktor Faktor Penyebab Kemiskinan," *Tazkiya*, vol. 16, no. 01, pp. 1–30, 2017.
- [5] R. Sanjaya and F. Fitriyani, "Prediksi Bedah Toraks Menggunakan Seleksi Fitur Forward Selection dan K-Nearest Neighbor," *JEPIN J. Edukasi Dan Penelit. Inform.*, vol. 5, no. 3, pp. 316–320, 2019.
- [6] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, 2015.
- [7] M. Wahyudi, L. Pujiastuti, and S. Solikhun, "Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma K-Means," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2020, vol. 2, pp. 432–440.
- [8] F. Fitriyani and R. Sanjaya, "KOMPARASI ALGORITMA LR, K-NN DAN SVM UNTUK ESTIMASI AREA KEBAKARAN HUTAN," *Infotronik J. Teknol. Inf. Dan Elektron.*, vol. 3, no. 2, pp. 103–110, 2018.

- [9] R. Sanjaya, D. Puspitasari, and others, “Noise Reduction through Bagging on Neural Network Algorithm for Forest Fire Estimates,” in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 2018, pp. 1–5.
- [10] V. Kotu and B. Deshpande, *Predictive analytics and data mining: A framework for optimizing decisions with R tool*. 2015. doi: 10.4018/978-1-4666-4940-8.ch004.