

Pengelompokan Data Pengangguran Terbuka Menggunakan Algoritma *K-Means* Berdasarkan Provinsi Jawa Barat

Nok Imas Pastia¹, Fatiha Nursari Dikananda²

^{1,2} Program Studi Manajemen Informatika, STMIK IKMI Cirebon

e-mail: ¹nokimaspastia22@gmail.com, ²fatiha.dikananda@gmail.com

Intisari

Pengangguran adalah salah satu permasalahan yang selalu muncul setiap tahunnya. Pengangguran terbuka dikategorikan sebagai angkatan kerja yang masih berusaha dalam mencari pekerjaan dan mempersiapkan usaha. Efek jangka panjang dari pengangguran menimbulkan psikologis yang buruk bagi para penganggur dan keluarganya. Pengangguran merupakan masalah yang berdampak sangat buruk terhadap perekonomian dan masyarakat, sebab itu perlu dilakukan upaya terus menerus untuk mengatasinya. Pada penelitian ini data yang digunakan bersumber dari Open Data Jabar tahun 2013-2021. Penelitian ini menggunakan teknik *data mining* yaitu algoritma *K-Means*. Metode *K-Means* merupakan metode *clustering* yang digunakan untuk membagi suatu kumpulan data menjadi beberapa kelompok. Tujuan penelitian ini untuk mendapatkan hasil pengelompokan Kabupaten/Kota di Provinsi Jawa Barat berdasarkan indikator tingkat pengangguran. Dibagi menjadi 2 bagian yakni kelompok dengan tingkat pengangguran tinggi dan kelompok dengan tingkat pengangguran rendah. Dari hasil pengelompokan diperoleh sebanyak 110 *cluster* rendah dan 104 *cluster* tinggi. Hasil dari penelitian ini adalah sebagai salah satu cara pemerintah dapat menciptakan dan memperluas lapangan pekerjaan untuk meningkatkan perekonomian. Terutama Kabupaten/Kota yang minim lapangan kerja yang menyebabkan pengangguran. Diharapkan penelitian ini dapat memberikan informasi serta masukan bagi pemerintah agar dapat memberikan perhatian lebih kepada Kabupaten/Kota Jawa Barat yang termasuk kategori tinggi pengangguran.

Kata kunci— *Data Mining*, Algoritma *K-Means*, Pengangguran, *Clustering*.

Abstract

Unemployment is a problem that always arises every year. Open unemployment is categorized as a workforce that is still trying to find work and prepare for business. The long-term impact of unemployment is psychologically bad for the unemployed and their families. Unemployment is a problem that has a very bad impact on the economy and society, therefore it is necessary to make continuous efforts to overcome it. In this study the data used comes from West Java Open Data for 2013-2021. This study uses a data mining technique, namely the K-Means Algorithm. The K-Means method is a clustering method used to divide a data set into several groups. The purpose of this study was to obtain the results of the grouping of districts/cities in West Java Province based on the unemployment rate into 2 parts, namely the group with the high unemployment rate and the group with the low unemployment rate. From the grouping results obtained 110 the low cluster and 104 the high cluster. The results of this study are one way the government can create and expand employment opportunities to improve the economy. Especially districts/cities that have minimal employment which causes unemployment. It is hoped that this research can provide information and input for the government so that it can pay more attention to districts/cities in West Java that are included in the category of high unemployment.

Keywords— Data Mining, K-means Algorithm, Unemployment, Clustering.

PENDAHULUAN

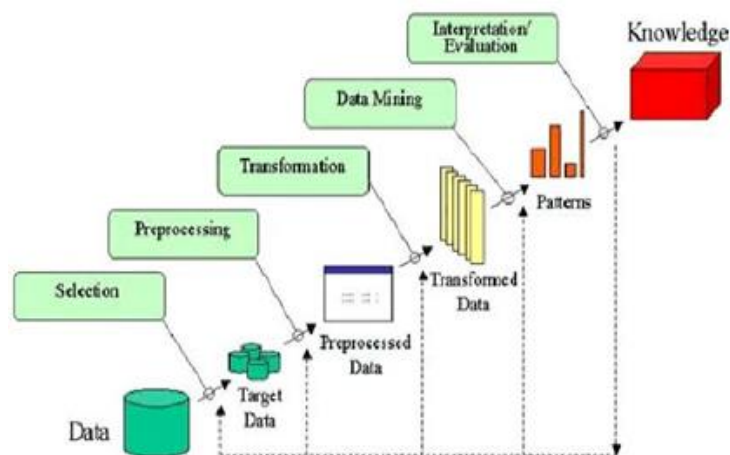
Pengangguran dapat digolongkan menjadi mereka yang tidak bekerja, mereka yang sedang mencari pekerjaan, dan mereka yang sudah bekerja tetapi belum memperoleh hasil [1]. Pengangguran juga dianggap sebagai seseorang atau kelompok orang yang sedang tidak memiliki pekerjaan atau tidak memiliki penghasilan tetap. Sedangkan pengangguran terbuka bisa dikategorikan sebagai angkatan kerja yang sedang berusaha mencari pekerjaan dan mempersiapkan suatu usaha [2]. Pengangguran terjadi karena banyak dari total angkatan kerja tidak selaras dengan pertumbuhan lapangan kerja [3] akibat dari minimnya lapangan pekerjaan di suatu perusahaan atau tempat kerja di suatu daerah. Pengangguran dapat disebabkan oleh beberapa faktor, antara lain tingginya laju pertumbuhan penduduk, tersedianya lapangan pekerjaan yang tidak seimbang, ketidaksiapan penduduk usia produktif menghadapi pesatnya perkembangan teknologi, tingkat pendidikan, perubahan struktur ekonomi, dan besarnya upah sehingga berdampak pada tingkat pengangguran [4].

Pengangguran terbuka biasanya terjadi di kalangan generasi muda yang baru saja menyelesaikan pendidikan menengah dan tinggi. Ada kecenderungan bagi mereka yang baru saja menyelesaikan pendidikannya untuk mencari pekerjaan sesuai dengan keinginannya. Keinginan mereka biasanya bekerja disektor atau kantor modern. Untuk mendapatkan pekerjaan, mereka biasanya rela menunggu sedikit lebih lama. Mereka mencoba berusaha mencari pekerjaan di kota, provinsi atau daerah di mana aktivitas industri telah berkembang lebih pesat [3]. Hal inilah yang membuat tingkat pengangguran terbuka cenderung lebih tinggi di kota atau daerah tersebut.

Penelitian terdahulu menggunakan *k-means clustering*; klusterisasi tingkat pengangguran berdasarkan data pengangguran tahun 2014-2019. Penelitian yang dilakukan [1] bertujuan menemukan kelompok provinsi berada di *cluster* tinggi dan kelompok provinsi yang berada dalam *cluster* rendah dalam rentang waktu 6 tahun. Penelitian ini menemukan 2 *cluster* provinsi dengan tingkat pengangguran tinggi menghasilkan *cluster* tinggi sebanyak 13 provinsi dan *cluster* rendah sebanyak 21 provinsi. Pada penelitian yang berjudul “Penerapan *Data Mining* Menggunakan Metode *Clustering* Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan” Penelitian yang dilakukan oleh [5] bahwa metode *k-means clustering* memiliki kemampuan untuk mengelompokkan sejumlah data yang cukup besar. Metode tersebut diterapkan pada data kepatuhan wajib pajak bumi dan bangunan. Tujuan dari penelitian ini adalah untuk mengetahui kelompok kepatuhan Wajib Pajak dan mengetahui Blok mana yang tingkat kepatuhan dalam membayar pajak PBB. Dari penelitian tersebut, peneliti menemukan pengetahuan baru dengan menggunakan metode *clustering*. Penelitian ini bertujuan untuk menggali pengetahuan lebih dalam data pengangguran dengan menganalisis data tingkat pengangguran terbuka berdasarkan Kabupaten/Kota di Provinsi Jawa Barat pada tahun 2013-2021 serta untuk mendapatkan hasil pengelompokan Kabupaten/Kota di Provinsi Jawa Barat berdasarkan indikator tingkat pengangguran. Selain itu, sebagai dasar pengambilan keputusan dan penentuan kebijakan oleh pemerintah dalam menyelaraskan jumlah angkatan kerja dengan

pertumbuhan penciptaan lapangan kerja terutama Kabupaten/Kota yang minim lapangan kerja yang menyebabkan pengangguran, agar mengantisipasi munculnya masalah baru akibat pengangguran di suatu daerah.

METODE PENELITIAN



Gambar 1 Tahap Penelitian

Sumber : [6]

Penelitian ini menggunakan metode *Knowledge Discovery In Database*. Berikut proses tahapan KDD :

1. *Selection*

Pilih set data atau fokus pada *subset variabel* atau sampel data. Hasil seleksi disimpan dalam file yang terpisah dari database. Atribut yang akan digunakan kedalam proses *data mining* yakni nama_kabupaten_kota, presentase_tingkat_pengangguran_terbuka, satuan (persen) dan rentang tahun.

2. *Preprocessing*

Persiapan awal yang dilakukan adalah membersihkan data (*data cleaning*). *Data Cleaning* dilakukan untuk menghilangkan *variabel* yang tidak digunakan (menghilangkan *data missing value*).

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 [7].

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 [8]. 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 mengelompokkan data pengangguran berdasarkan tingkat pengangguran Kabupaten/Kota Jawa Barat dalam rentang 8 tahun.

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.

HASIL DAN PEMBAHASAN

A. Hasil Pengelompokan

Data utama yang digunakan dalam penelitian ini adalah data persentase tingkat pengangguran terbuka menurut Kabupaten/Kota di Jawa Barat tahun 2013-2021 yang diperoleh dari Open Data Jabar. Data diterima dalam bentuk file excel.

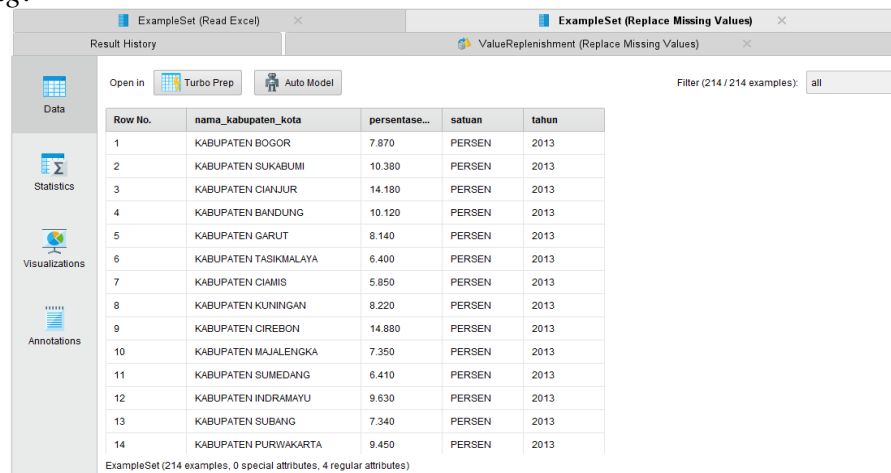
Tabel 1 *Dataset* Tingkat Pengangguran Terbuka

no.	nama_kabupaten_kota	presentase_tingkat_pengangguran_terbuka	satuan	tahun
1.	Kabupaten Bogor	7,87	Persen	2013
2.	Kabupaten Sukabumi	10,38	Persen	2013
3.	Kabupaten Cianjur	14,18	Persen	2013
4.	Kabupaten Bandung	10,12	Persen	2013
5.	Kabupaten Garut	8,14	Persen	2013
6.	Kabupaten Tasikmalaya	6,4	Persen	2013
7.	Kabupaten Ciamis	5,85	Persen	2013
8.	Kabupaten Kuningan	8,22	Persen	2013
9.	Kabupaten Cirebon	14,88	Persen	2013
10.	Kabupaten Majalengka	7,35	Persen	2013
...

213.	Kota Tasikmalaya	7,66	Persen	2021
214.	Kota Banjar	6,09	Persen	2021

1. Data Cleaning

Data *Cleaning* dilakukan untuk menghilangkan data yang kosong atau *missing value* dengan menggunakan operator *Replace missing values* pada saat proses menginputkan data pengangguran terbuka kedalam *rapidminer*. Hal ini dihasilkan bahwa kenyataannya data pengangguran terbuka menurut Kabupaten/Kota Provinsi Jawa Barat tidak memiliki data yang kosong dan *missing*.

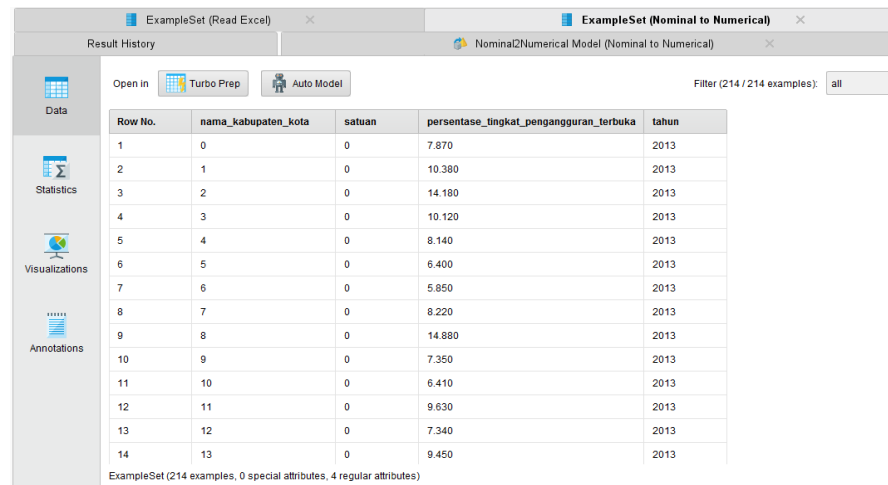


Row No.	nama_kabupaten_kota	persentase...	satuan	tahun
1	KABUPATEN BOGOR	7.870	PERSEN	2013
2	KABUPATEN SUKABUMI	10.380	PERSEN	2013
3	KABUPATEN CIANJUR	14.180	PERSEN	2013
4	KABUPATEN BANDUNG	10.120	PERSEN	2013
5	KABUPATEN GARUT	8.140	PERSEN	2013
6	KABUPATEN TASIKMALAYA	6.400	PERSEN	2013
7	KABUPATEN CIAMIS	5.850	PERSEN	2013
8	KABUPATEN KUNINGAN	8.220	PERSEN	2013
9	KABUPATEN CIREBON	14.880	PERSEN	2013
10	KABUPATEN MAJALENGKA	7.350	PERSEN	2013
11	KABUPATEN SUMEDANG	6.410	PERSEN	2013
12	KABUPATEN INDRAMAYU	9.630	PERSEN	2013
13	KABUPATEN SUBANG	7.340	PERSEN	2013
14	KABUPATEN PURWAKARTA	9.450	PERSEN	2013

Gambar 2 Hasil Proses Data *Cleaning*

2. Transformasi Data

Data akan ditransformasikan sesuai dengan metode algoritma yang digunakan agar dapat diproses kedalam *data mining*. *K-Means* merupakan algoritma yang hanya dapat menggunakan data *variabel* numerik. Dengan menggunakan operator *Nominal To Numerical* yang berguna untuk mengubah *variabel* menjadi numerik. Tahap ini mengubah atribut *nama_kabupaten_kota* dan *satuan* (persen) menjadi atribut numerik.



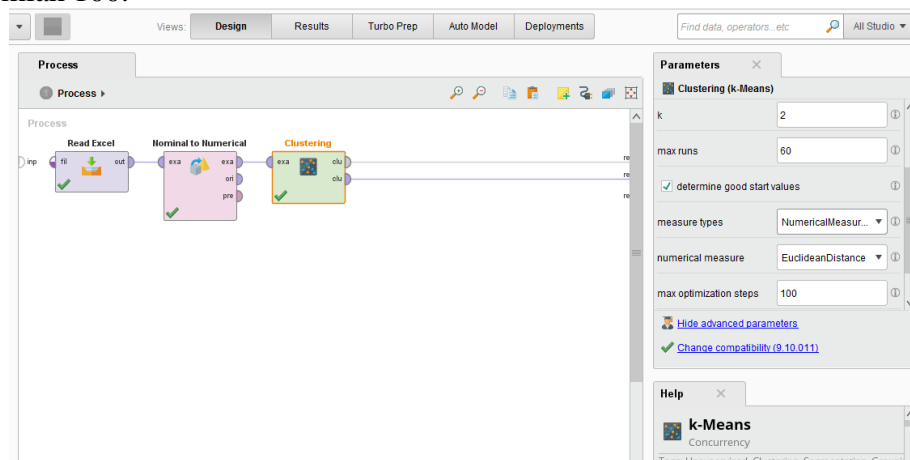
Row No.	nama_kabupaten_kota	satuan	persentase_tingkat_pengangguran_terbuka	tahun
1	0	0	7.870	2013
2	1	0	10.380	2013
3	2	0	14.180	2013
4	3	0	10.120	2013
5	4	0	8.140	2013
6	5	0	6.400	2013
7	6	0	5.850	2013
8	7	0	8.220	2013
9	8	0	14.880	2013
10	9	0	7.350	2013
11	10	0	6.410	2013
12	11	0	9.630	2013
13	12	0	7.340	2013
14	13	0	9.450	2013

ExampleSet (214 examples, 0 special attributes, 4 regular attributes)

Gambar 3 Hasil Proses *Transformasi Data*

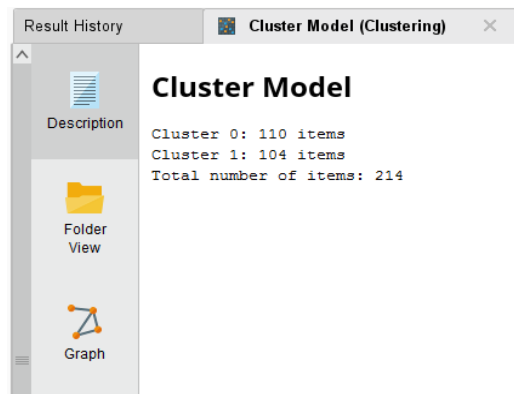
3. Proses *Data Mining*

Untuk menggali informasi yang diperoleh dari *dataset* tersebut, algoritma yang digunakan pada tahap ini adalah *k-means* dengan parameter $k = 2$ dan maximal run 60 dengan *Measure Type Numerical Measures*, *Numerical Measure* nya adalah *Euclidean Distance* dan *max optimization steps* nya berjumlah 100.



Gambar 4 Model Proses *K-Means*

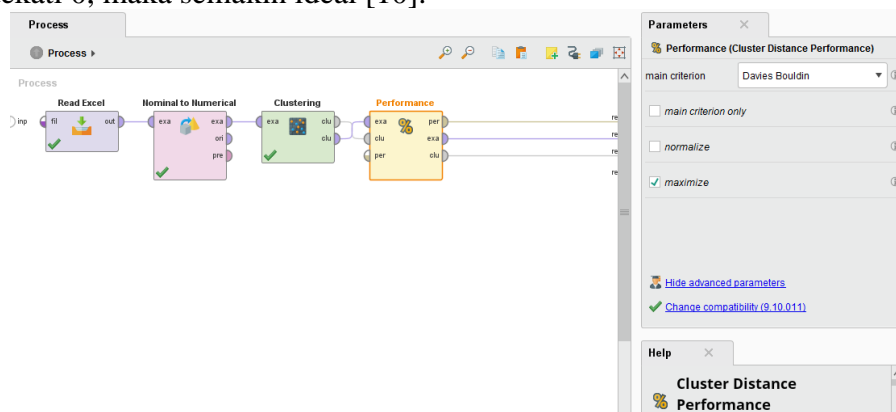
Diperoleh hasil untuk $k = 2$ yaitu *Cluster 0* sebanyak 110 anggota dan *Cluster 1* sebanyak 104 anggota.



Gambar 5 Hasil *Cluster Model*

B. Hasil *Cluster Optimum*

Setelah melakukan pemodelan *clustering*, selanjutnya adalah memvalidasi *clustering* menggunakan operator *cluster performance distance* untuk mengetahui nilai *Davies Bouldin Index* (DBI). Pada operator ini menggunakan *parameter main criterion Davies Bouldin Index*. DBI yang berupaya untuk menambah jarak antar *cluster* dan mengurangi jarak antar titik dalam sebuah *cluster* [9]. Semakin rendah nilai DBI atau semakin kecil nilai positifnya mendekati 0, maka semakin ideal [10].



Gambar 6 Proses Pengujian

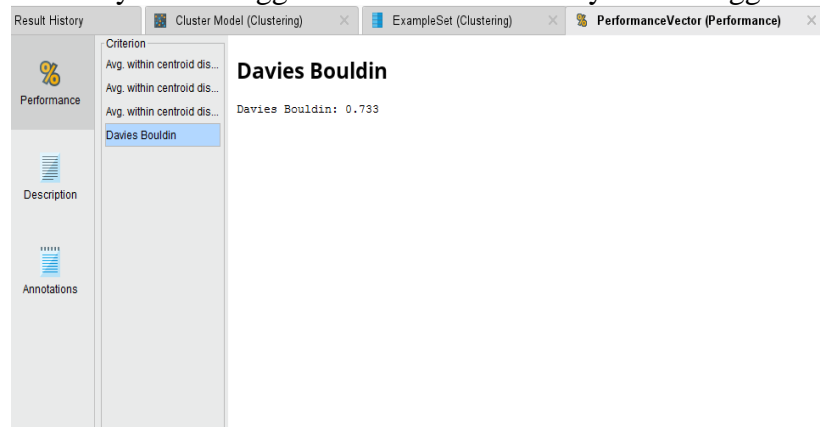
Dari hasil percobaan pengujian nilai $k = 2$ sampai nilai $k = 10$, diperoleh nilai DBI seperti tampak pada Tabel 2.

Tabel 2 Nilai DBI dalam menentukan nilai k

K	DBI
2	0.733
3	0.891
4	1.063
5	1.144
6	1.089

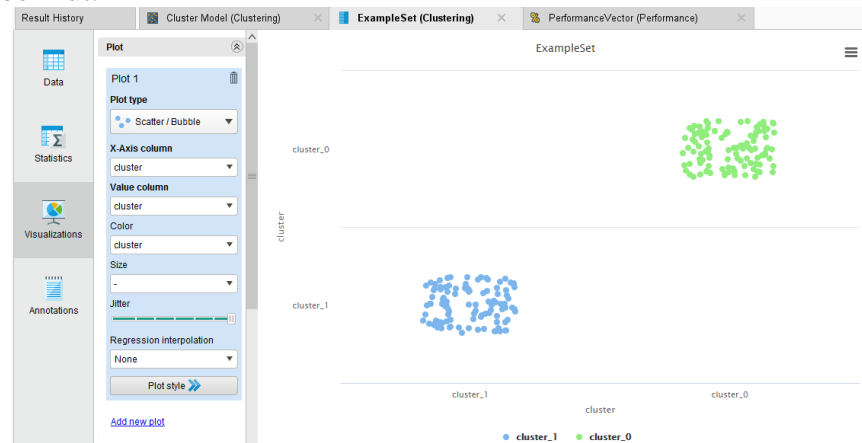
7	1.050
8	0.976
9	0.944
10	0.926

Perhitungan *K-Means* menunjukkan bahwa kelompok terbaik ada pada uji coba terhadap $k = 2$ dengan nilai DBI 0.733 yang membagi data menjadi 2 *cluster* yaitu *Cluster 0* sebanyak 110 anggota dan *Cluster 1* sebanyak 104 anggota.



Gambar 7 Nilai DBI Optimum pada $k = 2$

Adapun Diagram *Scatter/Bubble* untuk nilai $k = 2$ ditunjukan pada Gambar 8 berikut.



Gambar 8 Diagram Hasil pengelompokan $k = 2$

KESIMPULAN

Berdasarkan hasil pengelompokan data pengangguran terbuka tahun 2013-2021 menurut Provinsi Jawa Barat menggunakan algoritma *clustering* yaitu *K-Means*. Algoritma *K-Means* dengan nilai $k = 2$ memberikan *cluster* terbaik. Hasil *clustering* divalidasi menggunakan metode *Davies Bouldin Index* (DBI) dengan nilai DBI 0.733. Pada percobaan *K-Means* dengan nilai $k = 2$ data dibagi menjadi dua *cluster*, yaitu *cluster 0* berisi 110 anggota dan *cluster 1* berisi 104 anggota.

Pengetahuan diperoleh dengan bantuan *software RapidMiner* yang digunakan dalam penelitian ini :

1. *Cluster 0* adalah dengan tingkat pengangguran rendah yang terdiri dari Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar dalam rentang 8 tahun (2013-2021) dan Kabupaten Pangandaran dalam rentang 6 tahun (2015-2021). *Cluster 0* dengan daerah Kabupaten/Kota tertinggi pengangguran yakni Kota Cimahi tahun 2020 dengan presentase 13,3 persen dan daerah Kabupaten/Kota terendah pengangguran yakni Kabupaten Pangandaran tahun 2021 dengan presentase 3,25 persen. *Cluster 0* memiliki rentang presentase sebesar 3,25 sampai dengan 13,3 persen.
2. *Cluster 1* adalah dengan tingkat pengangguran tinggi yang terdiri dari Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu dan Kabupaten Subang dalam rentang 8 tahun (2013-2021). *Cluster 1* dengan daerah Kabupaten/Kota tertinggi pengangguran yakni Kabupaten Cirebon tahun 2013 dengan presentase 14,88 persen dan daerah Kabupaten/Kota terendah pengangguran yakni Kabupaten Bandung tahun 2017 dengan presentase 3,92 persen. *Cluster 1* memiliki rentang presentase sebesar 3,92 sampai dengan 14,88 persen.

SARAN

Saran untuk penelitian selanjutnya, hasil kelompok yang terbentuk dapat dikembangkan lebih lanjut menggunakan metode lain dan dengan penambahan *variabel* untuk mendapatkan informasi baru yang belum diketahui. Penelitian selanjutnya diharapkan menggunakan data ber-skala besar untuk mengetahui peningkatan atau penurunan nilai K optimal.

DAFTAR PUSTAKA

- [1] F. A. Tanjung, A. P. Windarto, and M. Fauzan, "Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, pp. 61–74, 2021, doi: 10.30645/jurasik.v6i1.271.
- [2] S. Muharni and S. Andriyanto, "Penerapan Metode K-Means Clustering pada Data Tingkat Pengangguran Terbuka Tahun 2016-2018 dan 2019-2021," *J. Inform.*, vol. 22, no. 01, pp. 89–99, 2022, doi: 10.31294/coscience.v2i2.1151.
- [3] W. A. Amrullah, N. Istiyani, F. Muslihatinningsih, J. I. Ekonomi, and S. Pembangunan, "Analisis Determinan Tingkat Pengangguran Terbuka di Pulau Jawa Tahun 2007-2016 (Determinant Analysis of Open

- Unemployment Rate In Java Island Year,” *e-Journal Ekon. Bisnis dan Akunt.*, vol. VI (I), no. 1, pp. 43–49, 2019.
- [4] D. Priastiwi and H. R. Handayani, “Analisis Pengaruh Jumlah Penduduk, Pendidikan, Upah Minimum dan PDRB terhadap Tingkat Pengangguran Terbuka Di Provinsi Jawa Tengah,” *Diponegoro J. Econ.*, vol. 1, no. 1, p. 159, 2019.
- [5] M. A. Putri, N. Rahaningsih, F. M. Basysyar, and O. Nurdiawan, “Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan,” *J. Account. Inf. Syst.*, vol. 5, no. 2, pp. 145–156, 2022.
- [6] I. Sumiati, A. R. F. Adinata, M. H. Al- Musa, R. Herdiana, and C. hERDIANA Rohmat, “Penerapan Metode Two-Step Cluster Untuk Pengelompokan Desa Berdasarkan Kepadatan Penduduk,” *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 5, no. 02, pp. 36–41, 2021, doi: 10.32485/kopertip.v5i2.154.
- [7] D. R. Chaerunisa, N. Rahaningsih, F. M. Basysyar, A. I. Purnamasari, and N. Suarna, “Pengelompokan Penjualan Madu Menggunakan Algoritma K-Means,” *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 5, no. 01, pp. 23–28, 2021, doi: 10.32485/kopertip.v5i1.144.
- [8] S. Al Syahdan and A. Sindar, “Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, pp. 56–63, 2018, doi: 10.32672/jnkti.v1i2.771.
- [9] I. Kamila, U. Khairunnisa, and M. Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119125, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [10] D. Ananda, S. Rohimah, B. Susilo, D. Wulan, and A. Mustofa, “Implementasi K-Means Dalam Pengelompokkan Data Akta Kelahiran di Indonesia: Implementation of K-Means in Grouping Birth Certificate Data in Indonesia,” *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 66–71, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/247%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/247/135>.