

## Penerapan Algoritme K-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat

Tasyifa Nafsiah Muthmainnah<sup>1\*</sup>, Siti Indriyana<sup>2</sup>, Ultach Enri<sup>3</sup>

<sup>1,2,3</sup>Jurusan SISTEM INFORMASI, Fakultas ILMU KOMPUTER, Universitas Singaperbangsa Karawang

\*Email: 2010631250023@student.unsika.ac.id

### Abstrak

*Pengangguran merupakan masalah sosial utama di banyak daerah, termasuk provinsi Jawa Barat di Indonesia. Provinsi Jawa Barat merupakan salah satu daerah terpadat dengan tingkat urbanisasi yang tinggi. Dengan pertumbuhan penduduk dan urbanisasi, tantangan untuk menciptakan pekerjaan yang cukup menjadi lebih sulit. Oleh karena itu, tujuan dari penelitian ini untuk mengklasterisasi data pengangguran terbuka pada masyarakat Jawa Barat yang digolongkan berdasarkan jumlah pengangguran berdasarkan kabupaten atau kota. Pada penelitian ini menggunakan metode CRISP-DM dengan Algoritma K-Means clustering. Hasil dari penelitian ini adalah 10 kabupaten/kota yang memiliki jumlah pengangguran tingkat rendah, lalu ada 15 kabupaten/kota yang memiliki jumlah pengangguran tingkat sedang dan terdapat 2 kabupaten/kota yang memiliki jumlah pengangguran dengan tingkat tinggi. Hasil dari pengujian dengan menggunakan Davies Bouldin Index cluster = 3 mempunyai kualitas cluster terbaik, karena nilai dari hasil pengujian Davies Bouldin Index dengan  $c=3$  adalah dengan nilai terkecil sebesar 0.28, yang dimana semakin rendah, semakin baik clusternya.*

**Kata kunci:** Clustering, Data Mining, Pengangguran

### Abstract

*Unemployment is a major social problem in many regions, including West Java province in Indonesia. West Java province is one of the most populous regions with a high level of urbanization. With population growth and urbanization, the challenge of creating enough jobs becomes more difficult. Therefore, the purpose of this study is to cluster open unemployment data in West Java communities classified by the number of unemployed people by district or city. This research uses CRISP-DM method with K-Means clustering algorithm. The result of this research is 10 regencies/cities that have low level of unemployment, then there are 15 regencies/cities that have medium level of unemployment and there are 2 regencies/cities that have high level of unemployment. The result of the test using Davies Bouldin Index cluster = 3 has the best cluster quality, because the value of the Davies Bouldin Index test result with  $c = 3$  is the smallest value of 0.28, which is the lower, the better the cluster.*

**Keywords:** Clustering, Data Mining, Unemployment

## PENDAHULUAN

Pengangguran Terbuka adalah situasi kondisi dimana seseorang yang termasuk dalam golongan kelas pekerja juga dapat disebut dengan seorang yang mempunyai keinginan untuk mendapatkan sebuah pekerjaan akan tetapi sama sekali belum bisa mencapainya. Pengangguran merupakan keadaan seseorang yang dimana seseorang ingin bekerja, tetapi tidak bisa mendapatkan pekerjaan yang sesuai dengan bidangnya (Tanjung, Windarto, and Fauzan 2021). Hal ini disebabkan oleh kurangnya kesempatan kerja meskipun mereka

telah berusaha untuk mendapatkannya, atau karena kurangnya motivasi untuk mencari atau bekerja. Pengangguran menjadi masalah serius di negara kita (Wahyudi and Pujiastuti 2020).

Meningkatkan kualitas dan produktivitas sumber daya manusia (SDM) sangat penting dalam upaya mengurangi tingkat pengangguran. Pendidikan dianggap sebagai sarana yang efektif untuk meningkatkan kualitas SDM karena mampu menghasilkan tenaga kerja yang berkualitas, memiliki pola pikir yang baik, dan memiliki sikap yang positif. Oleh karena itu, pendidikan sebagai salah satu metode yang

efektif dalam meningkatkan kualitas SDM (Allo et al. 2022).

Tingginya tingkat pengangguran di provinsi Jawa Barat, sehingga perlu adanya upaya untuk mengatasi masalah pengangguran di daerah tersebut. Data Tingkat Pengangguran Terbuka di Jawa Barat tidak hanya berisi informasi angka-angka yang menunjukkan tingkat pengangguran berdasarkan pendidikan. Banyak pengetahuan yang dapat ditambang (mining) dari data tingkat pengangguran tersebut. Metode dalam pengelompokan data ini menggunakan metode *data mining* dengan algoritma K-Means untuk melakukan clustering data (Dan 2022).

Beberapa penelitian sebelumnya yang menggunakan algoritma K-Means untuk analisis *clustering* yaitu (Nasution, Windarto, and Fauzan 2020), penelitian tersebut menggunakan algoritma K-Means dengan pengujian data dengan *tools rapidminer* untuk mengelompokkan data penduduk miskin menurut provinsi, diperoleh 2 *cluster* berdasarkan pengelompokan data penduduk miskin yaitu jumlah penduduk miskin tinggi dan rendah. Penelitian selanjutnya dilakukan oleh (Febianto and Palasara 2019), dengan menggunakan algoritma K-Means. Hasil dari penelitian tersebut diperoleh 5 *cluster* yang dihasilkan dengan pemetaan karakteristik dari setiap kelompok yang terbentuk. Penelitian yang dilakukan oleh (Riani, Voutama, and Ridwan 2023), menggunakan algoritma K-Means dalam penelitiannya dalam pengelompokan hasil belajar siswa mendapatkan hasil mengelompokkan peserta didik menjadi 3 *cluster* yaitu *cluster 0* dengan nilai rata-rata cenderung rendah yang memiliki jumlah anggota *cluster* yaitu 57 peserta didik. Kemudian *cluster 2* yaitu kelompok rata-rata nilai sedang dengan jumlah anggota *cluster* yaitu 42 peserta didik. Lalu *cluster 1* merupakan kelompok nilai dengan rata-rata tinggi yang hanya berjumlah 1 peserta didik.

Penelitian selanjutnya yang menggunakan Algoritma K-Means adalah (Ali 2023) yang berjudul Klasterisasi Pendidikan Masyarakat untuk mengetahui Daerah dengan Pendidikan Terendah menggunakan Algoritma K-Means, dengan tujuan penelitiannya yaitu untuk mengetahui desa mana saja yang memiliki tingkat pendidikannya rendah. Hasil

dari penelitiannya yaitu dari pengelompokan yang terbentuk pada data tingkat pendidikan sebanyak 229 record adalah 8 *cluster* yang dinyatakan desa pendidikannya terendah terdapat pada (C1) dengan jumlah 33 desa. Penelitian sebelumnya tentang algoritma K-Means juga dilakukan oleh (Nur Afidah 2023) yang berjudul Penerapan Metode Clustering dengan Algoritma K-Means untuk Pengelompokan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang. Hasil dari penelitian tersebut menunjukkan bahwa terdapat sedikit perbedaan antara perhitungan manual dengan aplikasi *RapidMiner*, pada perhitungan manual *cluster 1* (C1) sebanyak 7 kecamatan, *cluster 2* (C2) sebanyak 6 kecamatan, dan *cluster 3* (C3) sebanyak 1 kecamatan. Sedangkan pada aplikasi *RapidMiner cluster 1* (C1) sebanyak 10 kecamatan, *cluster 2* (C2) sebanyak 3 kecamatan, dan *cluster 3* (C3) sebanyak 1 kecamatan.

Mengacu pada penelitian sebelumnya yang membahas terkait pengangguran terbuka dengan variable yaitu upah minimum kabupaten/kota dan pertumbuhan indeks dan pembangunan manusia (Akramunnisa and Fajriani 2020). Maka tujuan dari penelitian ini yaitu untuk mengklasterisasi data pengangguran terbuka pada masyarakat Jawa Barat yang digolongkan berdasarkan jumlah pengangguran berdasarkan kabupaten atau kota. Diharapkan hasil analisis penelitian ini dapat membantu pemerintah untuk membuat kebijakan yang dapat memberikan solusi untuk memfokuskan pada wilayah dengan angka pengangguran tertinggi serta membuat kebijakan yang dapat meminimalisir angka pengangguran tiap masyarakat di kota maupun kabupaten di Jawa Barat.

Dari beberapa penelitian sebelumnya diatas ada perbedaan dengan penelitian yang telah dilakukan sekarang. Pada penelitian ini melakukan klasterisasi pada data pengangguran terbuka di provinsi Jawa Barat dengan atribut yang digunakan yaitu kode kabupaten kota, kabupaten kota, dan jumlah pengangguran. Perbedaan lain dari penelitian ini adalah *tools* yang digunakan untuk melakukan klasterisasi ini dengan *python*. Dan penelitian ini melakukan evaluasi dengan menggunakan

metode *Davies Bouldin Index* agar dapat menentukan jumlah cluster yang optimal.

**TINJAUAN PUSTAKA**

**2.1 Data Mining**

Data mining atau dikenal sebagai *knowledge discovery in database* (KDD) merupakan sebuah proses yang melibatkan pengumpulan dan penggunaan data historis untuk menemukan pola atau hubungan yang ada dalam kumpulan data yang berukuran besar (Handoko, Fauziah, and Handayani 2020) . Hasil dari proses data mining ini dapat digunakan untuk mengambil keputusan di masa depan (Sunia, Kurniabudi, and Alam Jusia 2019). Di dalam data mining juga terdapat berbagai metode yang dapat digunakan, seperti Classification (klasifikasi), Clustering (pengelompokan), Estimation (perkiraan), Prediction (prediksi), dan Association (asosiasi) (Sugianto, Rahayu, and Gusman 2020).

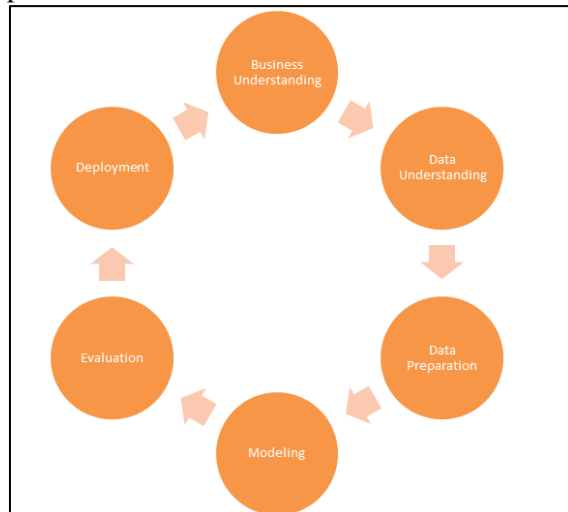
**2.2 K-Means**

K-Means *clustering* adalah metode pengelompokan yang menggunakan pendekatan deskriptif untuk menentukan objek mana yang akan dikelompokkan ke dalam *cluster* tertentu berdasarkan jarak terdekat ke rata-rata *cluster* tersebut (Kamila, Khairunnisa, and Mustakim 2019). K-Means *clustering* memiliki beberapa keunggulan, seperti tidak membutuhkan matriks jarak seperti yang diperlukan dalam clustering hierarkis, dan juga memiliki waktu komputasi yang cepat (Auliasari and Mariza Kertaningtyas 2022). Tools yang digunakan adalah dengan bahasa pemrograman *python*. Algoritma K-Means *Clustering* merupakan metode yang sederhana namun efektif untuk mengidentifikasi *cluster* dalam data dengan algoritma sebagai berikut (Suriani 2020) :

- Langkah 1: Menentukan Jumlah *cluster*.
- Langkah 2: Menentukan nilai yang akan menjadi lokasi pusat *cluster* awal.
- Langkah 3: Menghitung pusat *cluster* yang paling terdekat untuk setiap rekaman.
- Langkah 4: Pada setiap *cluster* k, menghitung centroid *cluster* dan memperbarui posisi setiap pusat *cluster* .
- Langkah 5: Ulangi langkah 3-5 sampai terjadi penghentian.

**METODE PENELITIAN**

Pada penelitian ini, dilakukan beberapa proses untuk memahami hubungan antar atribut pengklasifikasi pada dataset pengangguran terbuka di provinsi Jawa Barat. Metode yang digunakan dalam metode ini adalah CRISP-DM (Cross Industry Standard Process for Data Mining). Dalam metode CRISP-DM ada 6 tahap yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluasi, dan Deployment. persiapan data, pemilihan atribut pengklasifikasi, penggunaan K-Means Clustering untuk melakukan klasterisasi, dan menganalisis hasil klasterisasi. Proses dalam metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metode CRISP-DM

Penjelasan dari setiap tahapan pada metode CRISP-DM pada penelitian ini yaitu:

**3.1 Business Understanding**

Pada tahap pertama ini, dilakukan sebuah analisis masalah atau pemahaman tentang masalah apa yang dapat diangkat dari penelitian ini. Setelah mendapatkan tujuan dan permasalahan yang terjadi tahap berikutnya yaitu menentukan solusi untuk mengatasi permasalahan tersebut.

**3.2 Data Understanding**

Pada tahap kedua, dilakukan pengumpulan data, mendeskripsikan data, mencari data mana saja yang bermanfaat. Tahapan ini tujuannya untuk mendapatkan gambaran awal mengenai data. Data dalam

penelitian ini didapat dari BPS tentang data pengangguran terbuka di Provinsi Jawa Barat berdasarkan wilayah pada tahun 2022.

### 3.3 Data Preparation

Pada tahap ini, dilakukan beberapa tahapan yaitu pemilihan data dengan memilih kolom yang akan digunakan untuk proses modeling, pembersihan data, kemudian mengelompokkan atribut ke dalam data yang baru. Pada proses ini sangat penting untuk menyesuaikan dengan kebutuhan penelitian.

### 3.4 Modeling

Pada tahap modeling, dilakukan proses penentuan model dan penerapan teknik *data mining* yang digunakan. Algoritma yang digunakan dalam penelitian ini adalah *K-Means clustering* terhadap data pengangguran terbuka Provinsi Jawa Barat dengan menggunakan *python*.

### 3.5 Evaluation

Pada tahap ini, dilakukan analisis dari hasil yang telah diperoleh. Langkah ini merupakan hasil dari proses interpretasi model *data mining* yang digunakan. Evaluasi dilakukan dengan menggunakan metode *Davies Bouldin Index* untuk menentukan jumlah *cluster* yang optimal. *Davies Bouldin Index* adalah salah satu metode pengukuran validitas atau jumlah *cluster* optimal dalam metode *clustering*, dimana kohesi diartikan sebagai penjumlahan kedekatan data dengan pusat *cluster* dari *cluster* yang dipantau (Muningsih, Maryani, and Handayani 2021).

### 3.6 Deployment

Pada tahap terakhir, memberikan hasil dari pemodelan dan evaluasi pada proses *data mining* yang telah dilakukan. Hasil tersebut diberikan kepada Pemerintah Provinsi Jawa Barat dan dapat digunakan untuk mengambil keputusan yang tepat dalam mengatasi pengangguran dari hasil di wilayah yang banyak pengangguran di Provinsi Jawa Barat.

## HASIL DAN PEMBAHASAN

### 4.1 Business Understanding

Objek penelitian ini adalah Pengangguran di Provinsi Jawa Barat. Permasalahan bisnis Pengangguran di Jawa

Barat adalah meningkatnya angka pengangguran di provinsi Jawa Barat berdasarkan kabupaten atau kota. Dari masalah tersebut maka tujuan dari penelitian ini adalah untuk mengklasterisasi data pengangguran terbuka pada masyarakat Jawa Barat yang digolongkan berdasarkan jumlah pengangguran berdasarkan kabupaten atau kota dengan menggunakan algoritma *K-Means*.

### 4.2 Data Understanding

Pada tahap pemahaman data, memahami dan menganalisis data yang dimiliki dan apa yang dapat dilakukan pada data tersebut. Pada proses pengumpulan data, data yang diperoleh dari portal open data Jabar berupa data pengangguran di Provinsi Jawa Barat. Dimana ada 27 kabupaten atau kota dalam 4 tahun terakhir. Dengan memiliki atribut yaitu id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, pendidikan, jumlah pengangguran, satuan, tahun.

### 4.3 Data Preparation

Pada tahap ini, melakukan sebuah proses persiapan data yang terdiri dari *data cleaning*, dan *data selection*. Pada tahap pertama yaitu pembersihan data, pada saat proses pembersihan data ini menggunakan *python*. Dalam tahap ini, tidak ada nilai yang kosong atau *null* pada data sehingga dilanjutkan ke proses berikutnya yaitu *data selection*. Pada tahap pemilihan data, dari total 9 atribut hanya 3 atribut yang akan digunakan. Dataset tersebut dapat dilihat pada tabel 1.

Tabel 1. Dataset yang digunakan

No	kode_kabupaten_kota	kabupaten_kota	Jumlah Pengangguran
1	3201	KABUPATEN BOGOR	308165
2	3202	KABUPATEN SUKABUMI	102142
3	3203	KABUPATEN CIANJUR	102803
4	3204	KABUPATEN BANDUNG	126289
5	3205	KABUPATEN GARUT	101135
6	3206	KABUPATEN TASIKMALAYA	39247
7	3207	KABUPATEN CIAMIS	24923
8	3208	KABUPATEN KUNINGAN	52075
9	3209	KABUPATEN CIREBON	90118
10	3210	KABUPATEN MAJALENGKA	26808
11	3211	KABUPATEN SUMEDANG	47055
12	3212	KABUPATEN INDRAMAYU	61849
13	3213	KABUPATEN SUBANG	65605

14	3214	KABUPATEN PURWAKARTA	41287	3	3203	KABUPATEN CIANJUR	102803	1	Sedang
15	3215	KABUPATEN KARAWANG	118008	4	3204	KABUPATEN BANDUNG	126289	1	Sedang
16	3216	KABUPATEN BEKASI	206839	5	3205	KABUPATEN GARUT	101135	1	Sedang
17	3217	KABUPATEN BANDUNG BARAT	78920	6	3206	KABUPATEN TASIKMALAYA	39247	0	Rendah
18	3218	KABUPATEN PANGANDARAN	4077	7	3207	KABUPATEN CIAMIS	24923	0	Rendah
19	3219	KOTA BOGOR	59971	8	3208	KABUPATEN KUNINGAN	52075	0	Rendah
20	3220	KOTA SUKABUMI	14096	9	3209	KABUPATEN CIREBON	90118	1	Sedang
21	3221	KOTA BANDUNG	137098	10	3210	KABUPATEN MAJALENGKA	26808	0	Rendah
22	3222	KOTA CIREBON	13779	11	3211	KABUPATEN SUMEDANG	47055	0	Rendah
23	3223	KOTA BEKASI	140322	12	3212	KABUPATEN INDRAMAYU	61849	0	Rendah
24	3224	KOTA DEPOK	98426	13	3213	KABUPATEN SUBANG	65605	0	Rendah
25	3225	KOTA CIMAHI	34536	14	3214	KABUPATEN PURWAKARTA	41287	0	Rendah
26	3226	KOTA TASIKMALAYA	22964	15	3215	KABUPATEN KARAWANG	118008	1	Sedang
27	3227	KOTA BANJAR	5243	16	3216	KABUPATEN BEKASI	206839	2	Tinggi
				17	3217	KABUPATEN BANDUNG BARAT	78920	1	Sedang
				18	3218	KABUPATEN PANGANDARAN	4077	0	Rendah
				19	3219	KOTA BOGOR	59971	0	Rendah
				20	3220	KOTA SUKABUMI	14096	0	Rendah
				21	3221	KOTA BANDUNG	137098	1	Sedang
				22	3222	KOTA CIREBON	13779	0	Rendah
				23	3223	KOTA BEKASI	140322	1	Sedang
				24	3224	KOTA DEPOK	98426	1	Sedang
				25	3225	KOTA CIMAHI	34536	0	Rendah
				26	3226	KOTA TASIKMALAYA	22964	0	Rendah
				27	3227	KOTA BANJAR	5243	0	Rendah

Data pada tabel 1 berisi atribut yaitu kode kabupaten kota, nama kabupaten kota, dan jumlah pengangguran yang akan digunakan pada proses clustering.

#### 4.4 Modeling

Setelah data siap untuk digunakan, langkah berikutnya adalah pemodelan. Pada penelitian ini, algoritma yang digunakan yaitu algoritma K-Means yang kemudian akan dimodelkan berdasarkan data hasil transformasi. Jumlah cluster yang digunakan dalam penelitian ini adalah 3 cluster. Sebelum melakukan proses clustering data dengan menggunakan k-means, yang harus dilakukan yaitu menentukan titik pusat centroid. Gambar 2 menunjukkan hasil centroid yang digunakan dengan menggunakan pemrograman python.

```
[[[3.21250000e+03 5.57511250e+04 1.25000000e-01]
 [3.21600000e+03 2.06839000e+05 2.00000000e+00]
 [3.21177778e+03 1.12926778e+05 1.00000000e+00]
 [3.20100000e+03 3.08165000e+05 2.00000000e+00]
 [3.21937500e+03 1.83032500e+04 0.00000000e+00]]]
```

Gambar 2. Titik Pusat Centroid

Dari hasil centroid diatas menunjukkan bahwa cluster yang akan dibuat yaitu 3 cluster yaitu cluster 0, cluster 1, dan cluster 2. Langkah selanjutnya yaitu proses clustering dengan menggunakan k-means pada python. Tabel 2 merupakan tabel hasil dari proses clustering yang dijalankan dengan python .

Tabel 2. Hasil proses clustering dengan k-means menggunakan python

No	kode_kabupaten_kota	kabupaten_kota	Jumlah_Pengangguran	cluster	ket
1	3201	KABUPATEN BOGOR	308165	2	Tinggi
2	3202	KABUPATEN SUKABUMI	102142	1	Sedang

Dari hasil clustering pada tabel 2, menunjukkan bahwa jumlah pengangguran dengan tingkat rendah atau cluster 0 yaitu kabupaten Tasikmalaya, Ciamis, Kuningan, Majalengka, Sumedang, Indramayu, Purwakarta, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, dan Kota Banjar. Cluster 0 merupakan cluster yang mempunyai jumlah pengangguran rendah, lalu diikuti oleh cluster 1 dan cluster 2 yang mempunyai jumlah pengangguran sedang dan tinggi. Kemudian untuk jumlah pengangguran dengan tingkat sedang atau cluster 1 yaitu Kabupaten Sukabumi, Cianjur, Bandung, Garut, Cirebon, Karawang, Bandung Barat, Kota Bandung, dan Depok. Dan untuk jumlah pengangguran dengan tingkat tinggi atau cluster 2 yaitu kabupaten Bogor dan Kabupaten Bekasi.

#### 4.5 Evaluation

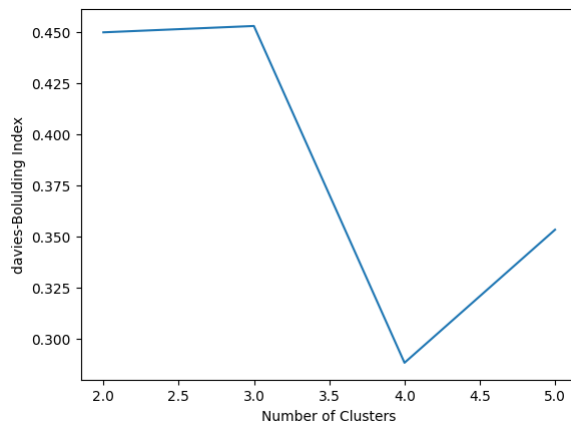
Pada tahap evaluasi dilakukan untuk mengukur kinerja hasil cluster jumlah pengangguran tahun 2022 dengan jumlah cluster = 3 menggunakan Davies Bouldin

Index. Hasil adalah angka, yang jika hasilnya rendah maka kualitas cluster tersebut semakin baik. Tabel 3 merupakan tabel hasil pengujian dengan menggunakan *Davies Bouldin Index* yang dijalankan pada python :

Tabel 3. Hasil pengujian

No	Jumlah Cluster	<i>Davies Bouldin Indeks</i>
1	2	0.4499245887963461
2	3	0.453032730094223
3	4	0.28817443998317
4	5	0.3533091254663149

Gambar 3 merupakan visualisasi terbaik dari hasil pengujian dengan menggunakan *Davies Bouldin Index*.

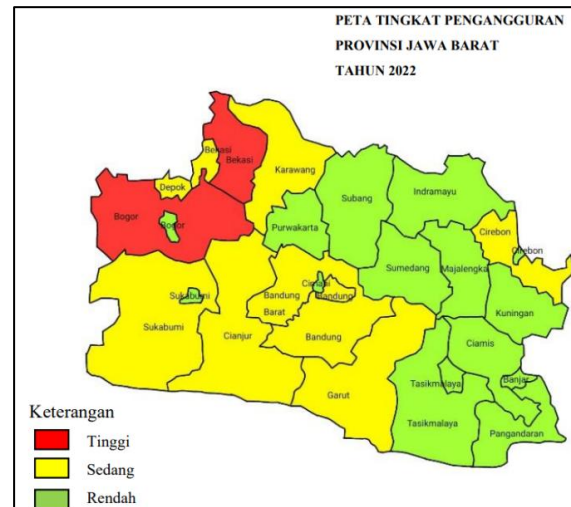


Gambar 3. Visualisasi terbaik dari hasil pengujian

Dari hasil pengujian *Davies Bouldin Index* pada gambar 3 dapat disimpulkan bahwa cluster = 3 mempunyai kualitas cluster terbaik, karena nilai dari hasil pengujian *Davies Bouldin Index* dengan  $c=3$  adalah dengan nilai terkecil sebesar 0.28, yang dimana semakin rendah nilai *Davies Bouldin Index*, semakin baik cluster tersebut.

#### 4.6 Deployment

Setelah selesai semua proses tahapan dari awal sampai pemodelan, menghasilkan sebuah visualisasi yang dapat ditampilkan dari data yang telah diolah dengan menggunakan python. Visualisasi tersebut dapat dilihat pada Gambar 4.



Gambar 4. Hasil Visualisasi

Dari visualisasi pemetaan pada gambar 4 tersebut dapat terlihat Kota/ Kabupaten mana saja yang termasuk dalam tingkat pengangguran tinggi, sedang, dan rendah di provinsi Jawa Barat.

#### SIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan tentang Penerapan Algoritme K-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat pada tahun 2022. Atribut yang digunakan yaitu kode kabupaten kota, kabupaten kota, dan jumlah pengangguran dengan jumlah cluster = 3. Dapat disimpulkan bahwa terdapat 10 kabupaten/kota yang memiliki jumlah pengangguran tingkat rendah atau cluster 0, kemudian ada 15 kabupaten/kota yang memiliki jumlah pengangguran tingkat sedang atau cluster 1 dan terdapat 2 kabupaten/kota yang memiliki jumlah pengangguran dengan tingkat tinggi atau cluster 2. Lalu hasil dari pengujian dengan menggunakan *Davies Bouldin Index* cluster = 3 mempunyai kualitas cluster terbaik, karena nilai dari hasil pengujian *Davies Bouldin Index* dengan  $c=3$  adalah dengan nilai terkecil sebesar 0.28, yang dimana semakin rendah, semakin baik clusternya. Dari hasil yang telah diperoleh pada penelitian ini diharapkan dapat dijadikan sebagai acuan pada penelitian selanjutnya dalam mengkaji tingkat pengangguran baik pada tingkat Jawa Barat maupun daerah lainnya dengan menggunakan pendekatan yang lebih kompleks

dengan menggabungkan algoritma K-means dengan metode clustering lainnya.

#### DAFTAR PUSTAKA

- Akramunnisa, Akramunnisa -, and Fajriani Fajriani. 2020. "K-Means Clustering Analysis Pada Persebaran Tingkat Pengangguran Kabupaten/Kota Di Sulawesi Selatan." *Jurnal Varian* 3 (2): 103–12. <https://doi.org/10.30812/varian.v3i2.652>.
- Ali, Universitas Darwan. 2023. "Klasterisasi Pendidikan Masyarakat Untuk Mengetahui Daerah Dengan Pendidikan Terendah Menggunakan Algoritma K-Means" 5 (1): 38–44.
- Allo, Yosua Sitoto Tandi, Verra Sofica, Noor Hasan, and Minda Septiani. 2022. "Penggunaan Metode Naïve Bayes Dalam Mengklasifikasi Pengangguran Pada Desa Bojong Kulur." *Bianglala Informatika* 10 (1): 30–35. <https://doi.org/10.31294/bi.v10i1.12333>.
- Auliasari, Karina, and Mariza Kertaningtyas. 2022. "Analisis Cluster Atribut Audio Pada Lagu Terpopuler Aplikasi TikTok." *Jurnal Sains Dan Informatika* 8 (2): 140–49. <https://doi.org/10.34128/jsi.v8i2.497>.
- Dan, Tahun. 2022. "3172-7028-1-Pb" 22 (01): 89–99.
- Febianto, Nugroho Irawan, and Nicodias Palasara. 2019. "Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018." *Jurnal Sisfokom (Sistem Informasi Dan Komputer)* 8 (2): 130–40. <https://doi.org/10.32736/sisfokom.v8i2.653>.
- Handoko, Suhandio, Fauziah Fauziah, and Endah Tri Esti Handayani. 2020. "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering." *Jurnal Ilmiah Teknologi Dan Rekayasa* 25 (1): 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>.
- Kamila, Insanul, Ulya Khairunnisa, and Mustakim Mustakim. 2019. "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pengelompokan Data Transaksi Bongkar Muat Di Provinsi Riau." *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi* 5 (1): 119. <https://doi.org/10.24014/rmsi.v5i1.7381>.
- Masykur, Abdussalam Al, Siska Kurnia Gusti, Suwanto Sanjaya, and Febi Yanto. 2023. "Penerapan Metode K-Means Clustering Untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar" 10 (1).
- Muningsih, Elly, Ina Maryani, and Vembria Rose Handayani. 2021. "Penerapan Metode K-Means Dan Optimasi Jumlah Cluster Dengan Index Davies Bouldin Untuk Clustering Propinsi Berdasarkan Potensi Desa." *Jurnal Sains Dan Manajemen* 9 (1): 95–100. <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>.
- Nasution, Irmanita, Agus Perdana Windarto, and M Fauzan. 2020. "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi." *Building of Informatics, Technology and Science (BITS)* 2 (2): 76–83. <https://doi.org/10.47065/bits.v2i2.492>.
- Nur Afidah, Nayla. 2023. "Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Pengelompokan Data Migrasi Penduduk Tiap Kecamatan Di Kabupaten Rembang." *PRISMA, Prosiding Seminar Nasional Matematika* 6: 729–38. <https://journal.unnes.ac.id/sju/index.php/prisma/>.
- Riani, Adella Putri, Apriade Voutama, and Taufik Ridwan. 2023. "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow." *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)* 6 (1): 164. <https://doi.org/10.53513/jsk.v6i1.7351>.
- Sugianto, Castaka Agus, Ayu Hendrati Rahayu, and Aditia Gusman. 2020. "Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah." *Journal of Information Technology* 2 (2): 39–44. <https://doi.org/10.47292/joint.v2i2.30>.
- Sunia, Dina, Kurniabudi, and Pareza Alam Jusia. 2019. "Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means." *Jurnal Ilmiah Mahasiswa Teknik Informatika* 1 (2): 121–34.
- Suriani, Lilis. 2020. "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola

Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering.” *Jurnal Sistem Komputer Dan Informatika (JSON)* 1 (2): 151.

<https://doi.org/10.30865/json.v1i2.1955>.

Tanjung, Fadhillah Azmi, Agus Perdana Windarto, and M Fauzan. 2021. “Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia.” *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)* 6 (1): 61.

<https://doi.org/10.30645/jurasik.v6i1.271>.

Wahyudi, Mochamad, and Solikhun Pujiastuti. 2020. “Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan.” *Prosiding Seminar Nasional Riset Dan Informatika Science (SENARIS)* 2: 432–40.