

IMPLEMENTASI KLASTERISASI *K-MEANS* UNTUK MEMETAKAN JUMLAH PENGANGGURAN DI PROVINSI JAWA BARAT BERDASARKAN TINGKAT PENDIDIKAN

Teguh Firmansyah^{1*}, Yuliazmi²

^{1,2}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta,

Email: ^{1*}teguh.firmansyah178@gmail.com, ²yuliazmi@budiluhur.ac.id
(* : corresponding author)

Abstrak-Pengangguran adalah keadaan dimana seseorang yang tidak memiliki pekerjaan, maupun orang yang sedang mencoba untuk mencari pekerjaan. Pengangguran juga disebabkan beberapa faktor, seperti pertumbuhan penduduk yang tinggi, masyarakat yang kurang bisa adaptasi dengan kemajuan teknologi, kurangnya lapangan kerja, persaingan yang ketat, tingkat pendidikan masyarakat, serta banyaknya tenaga asing yang ada di Indonesia. Provinsi Jawa Barat menjadi salah satu Provinsi yang memiliki tingkat pengangguran yang tertinggi di Indonesia dengan jumlah penduduk terbanyak sehingga dapat dilakukan penelitian lebih lanjut. Hasil yang didapat diharapkan dapat dijadikan evaluasi oleh pemerintah jika tingkat pengangguran yang ada di daerah tersebut tinggi dan dapat dibuat suatu solusi untuk mengurangi masalah tersebut. Data yang diperoleh dari Open Data Jabar dengan kurun waktu 2017-2022. Pada penelitian ini menggunakan teknik dari data mining untuk klasterisasi jumlah pengangguran pada provinsi Jawa Barat. Data diolah dengan menggunakan algoritma *K-Means*. Dataset terdiri dari 27 Kabupaten / Kota yang terdiri dari Nama Provinsi, Kode Kabupaten, Nama Kabupaten / Kota, Pendidikan, Satuan, Jumlah Pengangguran, Tahun. Pada tingkat pendidikan Tidak / Belum Sekolah, SD, SMP, SMA, SMK, Diploma. Data yang telah dikumpulkan akan diolah dengan melakukan *Clustering* pada Rapid Miner dalam 3 *Cluster* yaitu *Cluster* 0 (jumlah pengangguran rendah) terdapat 6 tingkat pendidikan dengan jumlah 27 Kabupaten / Kota di Provinsi Jawa Barat. *Cluster* 1 (jumlah pengangguran sedang) terdapat 4 tingkat pendidikan dengan jumlah 6 Kabupaten/Kota. *Cluster* 2 (jumlah pengangguran tinggi) terdapat 5 tingkat pendidikan dengan jumlah 19 Kabupaten/Kota. Hasil akhir berupa pengelompokan data pengangguran rendah, sedang dan tinggi pada setiap Kabupaten/Kota di Provinsi Jawa Barat sehingga menjadi masukan bagi Pemerintah di Provinsi Jawa Barat dalam menganalisis jumlah pengangguran berdasarkan tingkat pendidikan.

Kata Kunci: *Data Mining, Clustering, Pengangguran, K-Means*

IMPLEMENTATION OF *K-MEANS* CLUSTERIZATION TO REDUCE THE NUMBER OF UNEMPLOYMENT IN WEST JAVA PROVINCE BASED ON EDUCATION LEVEL

Abstract-Unemployment is a situation where someone does not have a job or someone who is trying to find a job. Unemployment is also caused by several factors, such as high population growth, people who are less able to adapt to technological advances, lack of jobs, intense competition, the level of public education, and the large number of foreign workers in Indonesia. West Java Province is one of the provinces that has the highest unemployment rate in Indonesia with the largest population so further research can be carried out. The results obtained are expected to be used as an evaluation by the government if the unemployment rate in the area is high and a solution can be made to reduce the problem. Data obtained from West Java Open Data for the period 2017-2022. This study uses techniques from data mining to cluster the number of unemployed in the province of West Java. Data is processed using the *K-Means* algorithm. The dataset consists of 27 Regencies/Cities consisting of Province Name, Regency Code, Regency/City Name, Education, Unit, Number of Unemployed, and Year. At the level of education / Not yet Schooled, Elementary, Middle School, High School, Vocational School, Diploma. The data that has been collected will be processed by clustering the Rapid Miner in 3 clusters namely *Cluster* 0 (low unemployment) there are 6 levels of education with a total of 27 districts/cities in West Java Province. *Cluster* 1 (moderate number of unemployed) There are 4 levels of education with a total of 6 Regencies/Cities. *Cluster* 2 (high unemployment) has 5 levels of education with a total of 19 districts/cities. The final result is in the form of grouping data on low, medium, and high unemployment in each Regency/City in West Java Province so that it becomes the input for the Government in West Java Province in analyzing the number of unemployed based on education level.

Keywords: *Data Mining, Clustering, Unemployment, K-Means*

1. PENDAHULUAN

Pengangguran adalah keadaan dimana seseorang yang tidak memiliki pekerjaan, maupun orang yang sedang mencoba untuk mencari pekerjaan. Pengangguran tidak hanya merugikan diri sendiri tetapi juga kepada negara salah satunya Indonesia yang disebut sebagai negara berkembang. Pengangguran di Indonesia memiliki angka yang cukup tinggi sehingga banyak masyarakat yang ada di Indonesia hidup dengan kekurangan yang disebabkan tidak adanya pemasukkan yang tetap dan mengakibatkan orang tersebut menggantungkan hidupnya kepada orang lain. Tidak hanya itu, pengangguran juga disebabkan beberapa faktor, seperti pertumbuhan penduduk yang melesat tinggi, masyarakat yang kurang bisa adaptasi dengan kemajuan dari teknologi yang ada sekarang, kurangnya lapangan kerja, persaingan yang ketat, tingkat pendidikan masyarakat, serta banyaknya tenaga asing yang ada di Indonesia.

Jawa Barat merupakan salah satu provinsi dengan tingkat pengangguran tertinggi, namun tingkat pengangguran di Jawa Barat semakin menurun setiap tahunnya, namun jumlahnya masih tinggi. Tingkat pengangguran terbuka di Jawa Barat turun menjadi 8,23% pada tahun 2018, 8,04% pada tahun 2019, dan naik menjadi 10,46% pada tahun 2020 [1]. Jawa Barat berada di urutan pertama sebagai provinsi dengan jumlah penduduk terbanyak di Indonesia. Jumlah penduduk Jabar sebagaimana tercatat di BPS sebanyak 48.274.162 jiwa [2]. Jawa Barat memiliki 27 kabupaten dan kota, kabupaten merupakan wilayah yang lebih luas dibanding pemerintahan kota.

Terdapat penelitian dahulu yang menggunakan metode klusterisasi, Pengelompokan Kabupaten/Kota Berdasarkan Indikator Tingkat Pengangguran Menggunakan Algoritma *K-Means Clustering* (Studi Kasus: Provinsi Jawa Barat) [1]. Penerapan *Data Mining* Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma *K-Means* [3]. Analisis Kluster Tingkat Pengangguran di Provinsi Kalimantan Timur Menggunakan Algoritma *K-Means* [4]. *K-Means Clustering Analysis* pada Persebaran Tingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan [5]. Penerapan Algoritma *K-Means* Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru [6].

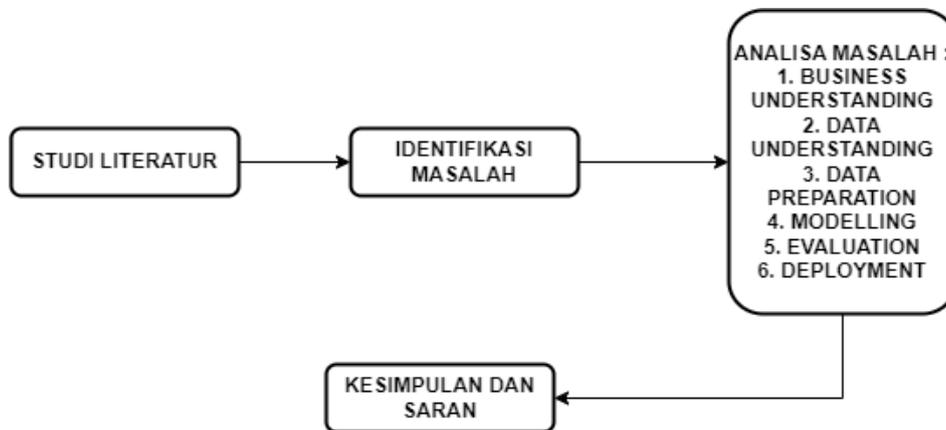
Data mining merupakan pemrosesan sebuah informasi dari suatu database yang dapat dilakukan meski tidak terlihat secara langsung [7]. *Data mining* adalah komponen dari kecerdasan buatan. Metode dalam proses penambangan data meliputi asosiasi, pengelompokan, estimasi, dan prediksi. Dalam penelitian ini, teknik yang digunakan untuk *data mining* adalah pengelompokan (*clustering*). *Clustering* merupakan proses mengelompokkan atau membagi pola data menjadi beberapa kelompok data. Pola-pola yang mirip akan dikelompokkan bersama dalam satu kelompok, sementara pola-pola yang berbeda akan terpisah dan dikelompokkan dalam kumpulan yang berbeda. Pada penelitian ini teknik dari data mining untuk klusterisasi Kabupaten/Kota Provinsi Jawa Barat yaitu dengan menggunakan algoritma *K-Means*. Algoritma *K-Means clustering* memiliki kelebihan yaitu mudah dipahami dan diimplementasikan untuk suatu penelitian. Dalam mempelajari algoritma tersebut membutuhkan waktu yang relatif cepat. *Tools* yang digunakan yaitu *Rapid Miner* untuk membentuk 3 cluster, cluster pengangguran rendah, pengangguran sedang dan pengangguran tinggi. Dengan menerapkan teknik *data mining* menggunakan metode *K-Means Clustering*, diharapkan dapat mempermudah pengelompokan tingkat pengangguran menjadi kategori rendah, sedang, dan tinggi di setiap wilayah Kabupaten/Kota di Provinsi Jawa Barat. Hal ini diharapkan akan membantu Pemerintah dalam mengidentifikasi solusi serta mengambil keputusan yang lebih terinformasi berdasarkan hasil pengelompokan data tersebut.

2. METODE PENELITIAN

Menurut [8] untuk melakukan proses *mining*, diperlukan tahapan dengan standarisasi yang dikenal dengan Gambar 1, yaitu model *Cross Standard Industry for Data Mining* (CRISP-DM) yang terdiri dari 6 fase analisa masalah sebagai berikut:

- Business Understanding*: Pada penelitian ini menggunakan data publik yang diambil dari Open Data Jabar. Langkah-langkah yang dilakukan melibatkan pemahaman terhadap kebutuhan dan tujuan dari perspektif bisnis, kemudian menerjemahkan pengetahuan tersebut menjadi definisi masalah yang sesuai untuk proses data mining. Setelah itu, melakukan perencanaan dan menetapkan strategi guna mencapai tujuan dalam proses data mining.
- Data Understanding*: Tahap ini dimulai dengan pengumpulan data, deskripsi data, dan penilaian terhadap kualitas data. Dalam penelitian ini menggunakan data 27 Kabupaten/Kota di Provinsi Jawa Barat yang meliputi : Kode Kabupaten/Kota, Nama Kabupaten/Kota, Tingkat Pendidikan, Jumlah Pengangguran, dan Tahun.
- Data Preparation*: Pada langkah ini, basis data akan disusun sedemikian rupa untuk mempermudah proses

- penambahan data dengan beberapa tahapan. *Data Selection* yang didapat dari *Open Data* Jabar, selanjutnya melakukan *data Preprocessing*. Pada tahap ini pembersihan data (*Data Cleansing*) yaitu proses pembersihan data dari *missing value* atau *noise*. Tahap Transformasi Data dengan cara normalisasi sehingga tidak ada perbedaan yang terlalu jauh pada skala data.
- d. *Modeling*: Tahap penting dalam menentukan teknik *data mining* yang digunakan dalam penelitian ini yaitu *Klasterisasi*, *tools* data mining yang digunakan Rapid Miner dan Microsoft Excel , algoritma data mining yang digunakan dalam penelitian ini menggunakan *K-Means*, serta parameter yang memiliki nilai optimal.
 - e. *Evaluation*: Pada tahap evaluasi, fokus utamanya adalah memastikan bahwa model yang dihasilkan sesuai dengan standar *K-Means Clustering* dan tidak ada tahap dari awal hingga pemodelan yang diabaikan atau terlewatkan. Langkah selanjutnya yang dipilih adalah memutuskan untuk melanjutkan ke tahap akhir [9]. Sebelum diolah dalam Rapid Miner untuk mendapatkan hasil *cluster*, data di Transformasi dalam Microsoft Excel.
 - f. *Deployment*: Adalah proses pembuatan laporan atau artikel jurnal menggunakan hasil penelitian [10]. Tahapan ini dilakukan setelah menyelesaikan tahapan diatas untuk menghasilkan laporan klasterisasi dalam menentukan Kabupaten/Kota mana saja yang termasuk *cluster 0* (tingkat pengangguran rendah), *cluster 1* (tingkat pengangguran sedang), *cluster 2* (tingkat pengangguran tinggi) di Provinsi Jawa Barat.



Gambar 1. CRISP-DM

3. HASIL DAN PEMBAHASAN

3.1 *Business Understanding*

Tahap pertama subjek pada penelitian ini adalah seluruh Kabupaten/ Kota di daerah Jawa Barat. Data yang diperoleh sebanyak 810 data pada tahun 2017 – 2022. Atribut dari data tingkat pengangguran di Provinsi Jawa Barat yaitu Kode Kabupaten/Kota, Nama Kabupaten/Kota, Tingkat Pendidikan, Jumlah Pengangguran, dan Tahun. Data tersebut didapatkan dari total seluruh data pada jenjang SD, SMP, SMA , dan Diploma. Oleh karena itu perlu adanya strategi dengan *clustering* atau pengelompokan data seluruh Kabupaten/Kota di Provinsi Jawa Barat untuk mengetahui Kabupaten/Kota mana saja yang termasuk dalam tingkat pengangguran rendah, sedang dan tinggi.

3.2 *Data Understanding*

Peneliti melakukan pemahaman terhadap kebutuhan data terkait dengan Tingkat Pengangguran di Provinsi Jawa Barat. Data yang digunakan pada penelitian ini didapatkan dari Open Data Jabar pada kurun waktu 2017-2022. Data Pengangguran meliputi 27 Kabupaten/Kota Provinsi di Jawa Barat yang meliputi data: Kode Kabupaten/Kota, Nama Kabupaten/Kota, Tingkat Pendidikan, Jumlah Pengangguran, dan Tahun. Data tersebut didapatkan dari total seluruh data pada jenjang SD, SMP, SMA , dan Diploma.

3.3 *Data Preparation*

Pada tahapan ini data yang didapat sebanyak 810 *record* jumlah pengangguran Provinsi Jawa Barat tahun 2017-2022. *Data preprocessing* pada penelitian ini ialah *Data Transformation*. Pada langkah ini, basis data akan disusun sedemikian rupa untuk mempermudah proses penambahan data. Terdapat beberapa teknik untuk melakukan transformasi data, seperti normalisasi, pemilihan atribut, dan diskritisasi. Variabel yang digunakan

dalam penelitian ini meliputi Nama Provinsi, Nama Kabupaten/Kota, Tingkat Pendidikan, Jumlah Pengangguran, Satuan, dan Tahun. Berikut merupakan hasil transformasi data dengan label *encoding*, seperti nama provinsi terlihat pada tabel 1, nama kabupaten/kota pada tabel 2, tingkat pendidikan pada tabel 3, satuan pada tabel 4, tahun pada tabel 5.

- a. Nama Provinsi: Pada tabel 1 terdapat hasil transformasi Nama Provinsi, untuk Provinsi Jawa Barat ditransformasi menjadi angka 1.

Tabel 1. Hasil Transformasi Data Nama Provinsi

Nama Provinsi	Kode
Jawa Barat	1

- b. Nama Kabupaten/Kota: Pada tabel 2 terdapat hasil transformasi nama kabupaten dengan kode 1-18 dan nama kota dengan kode 19-27.

Tabel 2. Hasil Transformasi Data Kabupaten/Kota

Kabupaten/Kota	Kode
Bandung	1
Bandung Barat	2
Bekasi	3
Bogor	4
Ciamis	5
Cianjur	6
Cirebon	7
Garut	8
Indramayu	9
Karawang	10
Kuningan	11
Majalengka	12
Pangandaran	13
Purwakarta	14
Subang	15
Sukabumi	16
Sumedang	17
Tasikmalaya	18
Bandung	19
Banjar	20
Bekasi	21
Bogor	22
Cimahi	23
Cirebon	24
Depok	25
Sukabumi	26
Tasikmalaya	27

- c. Tingkat Pendidikan: Pada tabel 3 terdapat hasil transformasi tingkat pendidikan yaitu 1-6.

Tabel 3. Hasil Transformasi Data Tingkat Pendidikan

Nama Tingkat Pendidikan	Kode
Diploma I/II/III	1
SD dan SD ke bawah	2
SMA, SLTA, SMA UMUM	3
SLTP,SMP	4
SMK(SMA Kejuruan)	5
Tidak/Belum Pernah Sekolah	6

- d. Satuan: Pada tabel 4 terdapat hasil transformasi satuan orang yaitu kode 1.

Tabel 4. Hasil Transformasi Satuan

Satuan	Kode
Orang	1

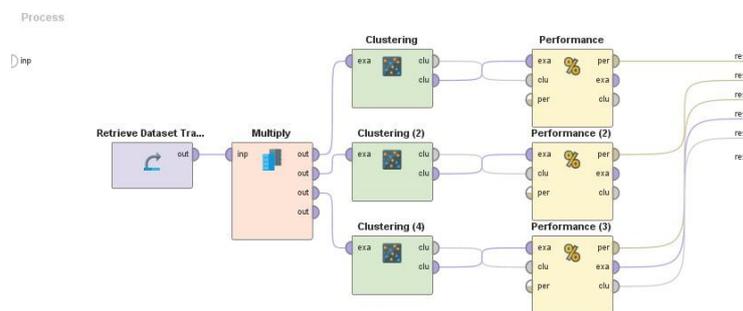
e. Tahun: Pada tabel 5 terdapat hasil transformasi tahun2017-2022 dengan kode 1-6.

Tabel 5. Hasil Transformasi Satuan

Tahun	Kode
2017	1
2018	2
2020	3
2020	4
2021	5
2022	6

3.4 Modeling

Pemodelan pertama yaitu menggunakan algoritma *K-Means*, yang berguna untuk mengetahui nilai Davies Bouldin Index paling kecil. Berikut Gambar 2 proses yang dihasilkan menggunakan *tools Rapid miner*.



Gambar 2. Proses Dari Algoritma K-Means

Berikut hasil yang telah diolah maka didapat nilai Davies Bouldin Index seperti terlihat pada tabel 6. Nilai Davies Bouldin Index paling kecil dengan penentuan jumlah *cluster* terbaik adalah berjumlah 3 *cluster* yaitu 0.503, Davies Bouldin Index yang didapat semakin kecil maka semakin optimal *cluster* yang dihasilkan.

Tabel 5. Nilai Davies Bouldin Index K-Means

Penentuan Jumlah <i>Cluster</i> Terbaik	Hasil Davies Bouldin Index
2 <i>Cluster</i>	0.577
3 <i>Cluster</i>	0.503
4 <i>Cluster</i>	0.504

Pada tabel 7, merupakan percobaan dengan 3 metode dan hasil yang terbaik berdasarkan nilai DBI terdapat pada algoritma *K-Means* dan *x-means* dengan nilai DBI sebesar 0,503. Nilai tersebut dikatakan baik karena hasil tersebut merupakan nilai terkecil yang didapat, karena semakin kecil nilai DBI atau semakin mendekati nilai nol maka semakin optimum *cluster* yang dihasilkan.

Tabel 7. Perbandingan Index Nilai Davies Bouldin

Algoritme	Hasil Davies Bouldin Index
K-Means	0,503
K-Medoids	0,543
X-Means	0,503

Gambar 3, merupakan hasil 3 *cluster* berdasarkan tingkat pengangguran yang didapat. Dengan menggunakan pemodelan *K-Means clustering* seperti pada gambar 3, dengan jumlah data sebanyak 810 dan inisialisasi jumlah *cluster* sebanyak 2 buah, *cluster*, sesuai dengan pendefinisian nilai k dengan jumlah *cluster* 0 (tingkat pengangguran rendah) : 558 *items*, *cluster* 1 (tingkat pengangguran sedang) : 31 *items*, *cluster* 2 (tingkat pengangguran tinggi) : 221 *items*.

Cluster Model

Cluster 0: 558 items
Cluster 1: 31 items
Cluster 2: 221 items
Total number of items: 810

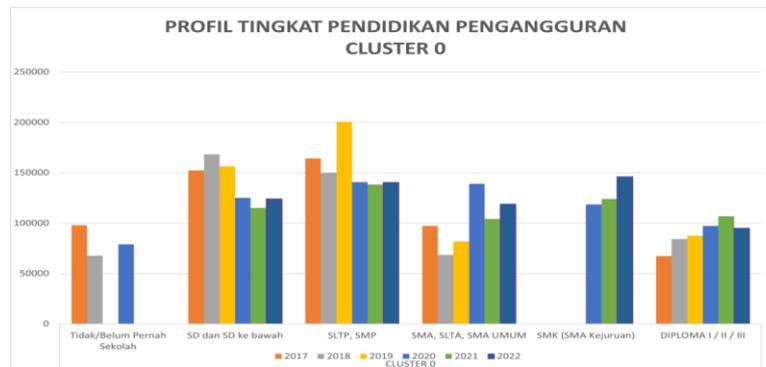
Gambar 3. Hasil Jumlah Cluster

Tabel 8 merupakan hasil dari Kabupaten/Kota yang masuk kedalam cluster 0, cluster 1 dan cluster 2.

Tabel 8. Data Hasil Cluster

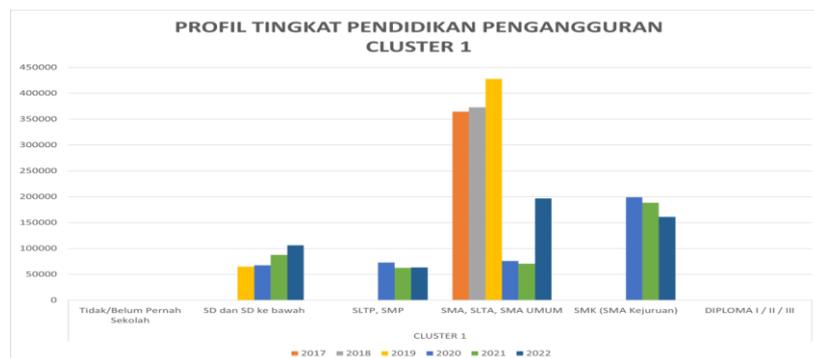
Cluster	
Cluster 0	Bandung, Bandung Barat, Bekasi, Bogor, Cianjur, Cirebon, Garut, Indramayu, Karawang, Kuningan, Majalengka, Pangandaran, Purwakarta, Subang, Sukabumi, Sumedang, Tasikmalaya, Kota Bandung, Kota Banjar, Kota Bekasi, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Depok, Kota Sukabumi, Kota Tasikmalaya.
Cluster 1	Bekasi, Bogor, Cirebon, Karawang, Kota Bandung, Kota Bekasi.
Cluster 2	Bandung, Bandung Barat, Bekasi, Bogor, Cianjur, Cirebon, Garut, Indramayu, Karawang, Kuningan, Purwakarta, Subang, Sukabumi, Sumedang, Tasikmalaya, Kota Bandung, Kota Bekasi, Kota Bogor, Kota Depok.

- a. Cluster 0: Dengan hasil grafik pada gambar 4 menggambarkan jumlah pengangguran yang ada pada tiap tingkat pendidikan dalam 5 tahun pada setiap kabupaten/kota di provinsi Jawa Barat yang termasuk dalam cluster 0.



Gambar 4. Tingkat Pendidikan Pengangguran Cluster 0

- b. Cluster 1: Dengan hasil grafik pada gambar 5 menggambarkan jumlah pengangguran yang ada pada tiap tingkat pendidikan selama 5 tahun di setiap kabupaten/kota di provinsi Jawa Barat yang termasuk dalam cluster 1.



Gambar 5. Tingkat Pendidikan Pengangguran Cluster 1

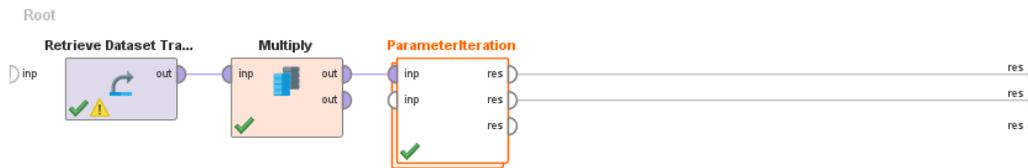
- c. Cluster 2: Dengan hasil grafik pada gambar 6 menggambarkan jumlah pengangguran yang ada pada tiap tingkat pendidikan selama 5 tahun di setiap kabupaten/kota di provinsi Jawa Barat yang termasuk dalam cluster 2.



Gambar 6. Tingkat Pendidikan Pengangguran Cluster 2

3.5 Evaluation

Pada bagian ini disampaikan mekanisme pengukuran performa model terbaik. Metode uji dari penelitian ini menggunakan Davies-Bouldin Index untuk menguji jumlah cluster yang optimal menggunakan tools Rapid miner. Proses Gambar 7 menggunakan operator loop parameters. Lalu dari hasil processLog menunjukkan nilai DB terkecil ialah 0,505 dengan jumlah 3 cluster yang optimal. Gambar 8 yang menunjukkan hasil pengujian dengan nilai DB terkecil.



Gambar 7. Operator Pengujian

ProcessLog (11 rows, 3 columns)

k	DB	W
2	0.594	1380233...
3	0.505	6001099...
4	0.507	3579680...
5	0.524	2518423...
6	0.482	1463316...
7	0.477	1113960...
8	0.478	9258998...
9	0.479	8431203...
10	0.492	7172148...
11	0.514	7605693...
13	0.516	6039969...

Gambar 8. Hasil Pengujian

3.6 Deployment

Pada tahap akhir ini, CRISP-DM menyebarkan pengetahuan atau informasi yang diperoleh yang diimplementasikan dalam laporan dan analisis setiap kluster yang diperoleh.

Tabel 9. Data Jumlah Minimal, Maksimal, dan Rata – Rata Pada Setiap Atribut Data

	NILAI	DIPLOMA I / II / III	SD dan SD ke bawah	SMA, SLTA, SMA UMUM	SLTP, SMP	SMK (SMA Kejuruan)	Tidak/Belum Pernah Sekolah
Cluster 0	Min	0	0	812	432	1942	0
	Rata - rata	3843	7509	8709	8267	9252	3020
	Max	17436	17657	17882	17927	17188	14788
Cluster 1	Min	19009	18581	18237	18226	18112	0
	Rata - rata	29768	27283	32754	26284	30042	0
	Max	51149	47812	55342	43728	56194	0
Cluster 2	Rata - rata	0	64836	57059	62464	66724	0
	Min	0	81429	83770	66158	91352	0
	Max	0	105933	136782	72817	119828	0

Dari hasil data yang ada pada tabel 9 dalam 5 Tahun diperoleh hasil yaitupada *cluster* 0 atribut data tingkat pendidikan Tidak sekolah memiliki nilai minimal 0 dan nilai maksimal 14.768 serta memiliki nilai rata – rata 3.020, atribut data SD memiliki nilai minimal 0 dan nilai maksimal 17.657 serta memiliki nilai rata – rata 7.509, atribut data SMP memiliki nilai minimal 432 dan nilai maksimal 17.927 serta nilai rata – rata 8.267, atribut data SMA memiliki nilai minimal 812 dan nilai maksimal 17.882 serta memiliki nilai rata – rata 8.709, atribut data SMK memiliki nilai minimal 1.942 dan nilai maksimal 17.188 serta memiliki nilai rata – rata 9.252, atribut data DIPLOMA memiliki nilai minimal 0 dan nilai maksimal 17.436 serta memiliki nilai rata – rata 3.843.

Pada *cluster* 1 atribut data tingkat pendidikan Tidak sekolah memiliki nilai minimal 0 dan nilai maksimal 0 serta memiliki nilai rata – rata 0, atribut data SD memiliki nilai minimal 18.581 dan nilai maksimal 47.812 serta memiliki nilai rata – rata 27.283, atribut data SMP memiliki nilai minimal 18.226 dan nilai maksimal 43.728serta memiliki nilai rata – rata 26.284, atribut data SMA memiliki nilai minimal 18.237dan nilai maksimal 55.342 serta memiliki nilai rata – rata 32.754, atribut data SMK memiliki nilai minimal 18.112 dan nilai maksimal 56.194 serta memiliki nilai rata – rata 30.042, atribut data DIPLOMA memiliki nilai minimal 19.009 dan nilai maksimal 51.149 serta memiliki nilai rata – rata 29.786.

Selanjutnya, pada *cluster* 2 atribut data tingkat Pendidikan tidak Sekolah memiliki nilai minimal 0 dan nilai maksimal 0 serta memiliki nilai rata-rata 0, atribut data SD memiliki nilai minimal 81.429 dan nilai maksimal 105.933 serta memiliki nilai rata-rata 64.836, atribut data SMP memiliki nilai minimal 66.158 dan nilai maksimal 72.917 serta memiliki nilai rata-rata 62.464, atribut data SMA memiliki nilai minimal 83.770 dan nilai maksimal 136.782 serta memiliki nilai rata – rata 57.059, atribut data SMK memiliki nilai minimal 66.724 dan nilai maksimal 119.828 serta memiliki nilai rata – rata 66.724, atribut data DIPLOMA memiliki nilai minimal 0 dan nilai maksimal 0 serta memiliki nilai rata – rata 0.

Pemerintah dapat mengetahui tingkat pengangguran rendah, sedang dan tinggi dalam 3 *cluster* yang ada pada setiap Kabupaten/Kota di Provinsi Jawa Barat berdasarkan tingkat pendidikan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang ada dapat disimpulkan bahwa terdapat 3 *Cluster*, yaitu *Cluster* 0 (Tingkat Pengangguran Rendah) yang terdiri dari 27 Kabupaten/Kota seperti Kuningan, Majalengka, Pangandaran, Purwakarta, Subang, Sukabumi, Kota Bandung, Kota Banjar, Kota Bekasi dan Kota Bogor. *Cluster* 1 (Tingkat Pengangguran Sedang) terdapat 6 Kabupaten/Kota seperti Bekasi, Bogor, Cirebon, Karawang, Kota Bandung dan Kota Bekasi. *Cluster* 2 (Tingkat Pengangguran Tinggi) terdiri dari 19 Kabupaten/Kota seperti Bandung, Bandung Barat, Bekasi, Bogor, Cianjur, Cirebon, Garut, Indramayu, Kota Bandung, Kota Bekasi, Kota Bogor. Nilai Davies Bouldin Index menggunakan *tools* Rapid Miner dengan algoritma *K-Means* paling kecil dalam penentuan *cluster* terbaik adalah berjumlah 3 *cluster* yaitu 0,503. Banyaknya total pengangguran yang ada di Provinsi Jawa Barat berdasarkan tingkat pendidikan berada pada pendidikan SMA. Pada tingkat pendidikan SMA di Jawa Barat harus diberikan perhatian khusus seperti membuat balai pelatihan untuk mengembangkan skill dan bakat sesuai dengan tingkat pengangguran di setiap pendidikan pada wilayahnya. Berdasarkan pembekalan sesuai skill yang dimilikinya maka mereka bisa mendapatkan kesempatan untuk bekerja atau membuka usaha.

DAFTAR PUSTAKA

- [1] Putrie, A. A., & Sanjaya, R., “Pengelompokan Kabupaten/Kota Berdasarkan Indikator Tingkat Pengangguran Menggunakan Algoritma K-Means Clustering (Studi Kasus: Provinsi Jawa Barat)”, *E-Prosiding Sistem Informasi (POTENSI)*, 2(2), 111–121, 2021, <http://eprosiding.ars.ac.id/index.php/psi/article/view/595>.
- [2] Kompas.com. (2021). *Jawa Barat Jadi Provinsi dengan Penduduk Terbanyak, Bagaimana dengan DKI Jakarta?* Kompas.Com. <https://regional.kompas.com/read/2021/12/31/184223678/jawa-barat-jadi-provinsi-dengan-penduduk-terbanyak-bagaimana-dengan-dki?page=al>.
- [3] Wahyudi, M., & Pujiastuti, S., “Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma K - Means”, *Prosiding Seminar Nasional Riset Dan Informatika Science (SENARIS)*, 2, 432–440, 2020.
- [4] Harits, D. P., “Analisis Klaster Tingkat Pengangguran di Provinsi Kalimantan Timur Menggunakan Algoritma K-Means”. *Jurnal Surya Teknika*, 9(2), 456–460, 2022. <https://doi.org/10.37859/jst.v9i2.4430>.
- [5] Akramunnisa, A., & Fajriani, F., “K-Means Clustering Analysis pada Persebaran Tingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan”, *Jurnal Varian*, 3(2), 103–112, 2020. <https://doi.org/10.30812/varian.v3i2.652>.
- [6] Nurseptia, Bella Voutama, Apriade Haeryana, Nono HSRonggo Waluyo, J., “Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru”, *Jurnal Teknologi Informasi*, 6(2), 181–186, 2022.
- [7] Kamila, I., Khairunnisa, U., & Mustakim, M., “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, 2019. <https://doi.org/10.24014/rmsi.v5i1.7381>.
- [8] I. A. Salam, K. W. Putra, S. Yuliatina, and B. N. Sari, “Application of Naïve Bayes for Classification of Criteria for Potable Water with the CRISP-DM Method”, *Journal of Computer Science and Informatics*, vol. 25, no. 1, Mar. 2023.
- [9] Astuti, D. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 60–72. <https://doi.org/10.20895/inista.v1i2.71>.
- [10] Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77. <https://doi.org/10.34010/jati.v12i1.6674>.