



Penerapan Algoritma K-Means Untuk Mengelompokkan Kepadatan Penduduk Di Provinsi DKI Jakarta

Frisma Handayanna¹, Sunarti²

¹Informatika, Teknologi Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia

²Sistem Informasi Kampus Kota Bogor, Teknik dan Informatika, Universitas Bina Sarana Informatika

¹frisma.fha@nusamandiri.ac.id, ²sunarti.sni@bsi.ac.id

Abstract

DKI Jakarta Province is an attraction for immigrants. If the population increases, if it cannot be resolved and managed well, it will result in bad things such as increasing the number of unemployed and affecting economic growth. Population data is used to help group regions based on population density in DKI Jakarta Province in 2019-2022 using the K-Means clustering method. From the results of the research, it provides a solution for the government to pay attention to population groups with the aim of preventing population density because it causes bad effects, so that community welfare is more guaranteed, so grouping (clustering) of provinces in DKI Jakarta is needed to provide information for people who wish to live in the Province DKI Jakarta. The research proves that the test results carried out clustering iterations of population density data were obtained in three iterations. For the results obtained by calculations using the K-Means method and using the rapidminer application, the results obtained were of the same value, namely the cluster with the highest population density of three districts/cities, namely South Jakarta, East Jakarta and West Jakarta whose population density continues to increase.

Keywords: Population Density, Data Mining, Clustering, K-Means

Abstrak

Provinsi DKI Jakarta menjadi daya tarik bagi para pendatang. Bertambahnya jumlah penduduk apabila tidak dapat diselesaikan dan ditata dengan baik, mengakibatkan hal buruk seperti meningkatnya jumlah pengangguran dan berpengaruh terhadap pertumbuhan ekonomi. Dari data penduduk digunakan membantu pengelompokan wilayah berdasarkan jumlah kepadatan penduduk di Provinsi DKI Jakarta tahun 2019-2022 dengan metode K-Means clustering. Dari hasil penelitian memberikan solusi bagi pemerintah untuk memperhatikan kelompok jumlah penduduk bertujuan mencegah kepadatan jumlah penduduk karena mengakibatkan pengaruh buruk, sehingga diperoleh kesejahteraan masyarakat yang lebih terjamin, maka diperlukan pengelompokan (clustering) provinsi di DKI Jakarta untuk memberikan informasi bagi masyarakat yang berkeinginan berdomisili di Provinsi DKI Jakarta. Penelitian membuktikan hasil pengujian dilakukan iterasi clustering data jumlah kepadatan penduduk diperoleh tiga iterasi. Untuk hasilnya diperoleh perhitungan dengan metode K-Means dan penggunaan aplikasi rapidminer, diperoleh hasil bernilai sama yaitu cluster dengan jumlah kepadatan tertinggi penduduk sebanyak tiga Kab/Kota yaitu Jakarta Selatan, Jakarta Timur dan Jakarta Barat yang jumlah kepadatan penduduknya meningkat terus.

Kata kunci: Kepadatan Penduduk, Data Mining, clustering, K-Means

1. Pendahuluan

Pertumbuhan ekonomi yang tinggi menandakan bahwa suatu daerah memiliki kesejahteraan masyarakat yang baik[1]. Salah satu yang menjadi indikator untuk keberhasilan pembangunan ekonomi yaitu pertumbuhan ekonomi. Tapi, tidak semua negara mampu untuk memperoleh pertumbuhan ekonomi yang baik. Pertumbuhan ekonomi berpengaruh untuk besarnya penerimaan disuatu daerah tersebut karena penerimaan masyarakatnya akan meningkat. Bertambahnya jumlah penduduk pada kota atau pun kabupaten apabila tidak dapat diselesaikan dengan baik dan ditata dengan baik maka akan mengakibatkan hal yang buruk seperti meningkatnya jumlah pengangguran. Dan dengan

adaanya penambahan jumlah penduduk juga akan berpengaruh terhadap pertumbuhan ekonomi. [2].

Apabila pertumbuhan ekonomi suatu negara atau wilayah tidak dapat berkembang dengan baik akan menyebabkan dampak terburuk yang akan muncul salah satunya adalah pengangguran. Pengangguran yang tinggi termasuk ke dalam masalah ekonomi dan sosial, orang-orang yang menganggur suatu saat bisa kehilangan kepercayaan dirinya sehingga dapat menimbulkan tindakan kriminal, perselisihan dengan masyarakat dan sebagainya[3]. Karena apabila pertumbuhan ekonomi tidak disertai dengan lapangan usaha, kesempatan kerja dengan kapasitas yang kecil serta jumlah penduduk yang selalu meningkat setiap



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

tahunnya maka akan menyebabkan jumlah pengangguran mengalami kenaikan [4]. Dengan demikian akan terjadi ketidak seimbangan pembagian penduduk [5]. Sehingga dapat mempengaruhi atau menyebabkan permasalahan yang cukup besar baik masalah perseorangan maupun kelompok yang berkeinginan tinggal di suatu wilayah [6].

Provinsi DKI Jakarta merupakan salah satu dari ibu kota di Indonesia dimana menjadi daya pikat bagi para pendatang diluar Propinsi DKI Jakarta. Dimana Pertambahan jumlah penduduk akan mempengaruhi pertumbuhan ekonomi dan kepadatan penduduk [7]. Karena di Jakarta begitu banyak kegiatan baik pendidikan, kesehatan, wisata, perdagangan dan lainnya. Penelitian ini dilakukan terhadap data penduduk Kab/Kota di DKI Jakarta yang digunakan yaitu diklasifikasikan berdasarkan jumlah penduduk per Kab/Kota dari tahun 2019-2021.

Untuk mengetahui jumlah penduduk yang terbanyak di wilayah Provinsi DKI Jakarta. Teknik *clustering* dapat membantu untuk mengelompokkan data secara otomatis tanpa perlu diberitahu label kelasnya [7]. Para ahli banyak mengusulkan metode *clustering*, salah satunya adalah *K-Means*. Metode *K-Means* digunakan dalam berbagai aplikasi kecil hingga menengah dan merupakan algoritma klasterisasi yang paling banyak karena kemudahan dalam mengaplikasikannya [8][9]. Algoritma *K-Means*, *cluster* yang dihasilkan cukup baik, sehingga metode ini bisa direkomendasikan sebuah *clustering* yang baik [10][11]. Metode *K-Means* metode *not hierarchical clustering* yang berusaha membantu penyertaan variabel kelompok dalam kelas hasil perhitungan akhir [12].

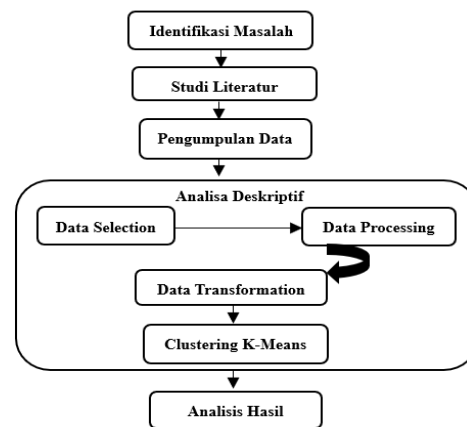
Meningkatnya jumlah penduduk dikota Medan sehingga menyebabkan tingkat kepadatannya paling tinggi di Sumatera Utara. Sehingga akan memicu permasalahan yang besar, Oleh karena, perlu dilakukan pemetaan penduduk setiap kelurahan kedalam beberapa kelompok (*cluster*), sehingga bisa memberikan informasi maupun pengetahuan baru bagi masyarakat dalam memilih lokasi tempat tinggalnya di kota Medan [6].

Pada penelitian ini data jumlah penduduk yang digunakan dididapatkan dari data Badan Pusat Statistik Jakarta. Data jumlah penduduk yang didapatkan yaitu data Kota/Kabupaten DKI Jakarta dari tahun 2019 yaitu sampai tahun 2021 yang terdiri dari enam Kota/Kabupaten. Data jumlah penduduk khususnya wilayah di DKI Jakarta yang diperoleh akan dilakukan penelitian dalam mengelompokkan jumlah penduduk berdasarkan Kota/Kabupaten menggunakan algoritma *K-Means*. Oleh karena itu diperlukan pemetaan penduduk setiap provinsi di DKI Jakarta menjadi beberapa kelompok (*cluster*), sehingga dapat memberikan informasi bagi masyarakat yang ini memilih tempat tinggal atau yang berkeinginan berdomisili di Provinsi DKI Jakarta.

Pengelompokan yang dilakukan yaitu dengan membagi menjadi tiga *cluster* yaitu *cluster* berdasarkan jumlah penduduk rendah, *cluster* dengan jumlah penduduk sedang, dan *cluster* dengan jumlah penduduk tinggi. Pengelompokan ini dilakukan untuk memberikan sebuah solusi supaya pemerintah dapat melihat dan memperhatikan kelompok jumlah penduduk yang ada dengan ini dapat mencegah kepadatan jumlah penduduk dan dampak-dampak serta pengaruh buruk lainnya agar diperoleh kesejahteraan masyarakat yang lebih terjamin [13].

2. Metode Penelitian

Metode penelitian yang dilakukan terdiri dari lima tahap yaitu: a). Teknik Pengumpulan Data. Pengumpulan Data Jumlah Penduduk per Kota/Kabupaten DKI Jakarta 2019-2021 data ini diperoleh dari website Biro Pusat Statistik <https://jakarta.bps.go.id/> mengenai data jumlah penduduk provinsi DKI Jakarta dimana data terdiri dari enam Kab/Kota di DKI Jakarta. b). Tahapan Penelitian Adapun tahapan penelitian yang akan dilakukan penulis dalam proses penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

Adapun tahapan penelitian yang dilakukan penulisan adalah: (1). Identifikasi masalah. Dimana tahapan ini penulis melakukan identifikasi permasalahan dan melakukan perumusan masalah dimana tujuannya melakukan pemetaan penduduk setiap provinsi DKI Jakarta menjadi beberapa kelompok (*cluster*), sehingga dapat memberikan informasi bagi masyarakat yang ini memilih tempat tinggal atau yang berkeinginan berdomisili di Provinsi DKI Jakarta. (2). Studi Literatur Pada tahapan ini penulis melakukan kajian pustaka dengan mempelajari buku, artikel dan hasil penelitian yang sama dan sesuai dengan permasalahan yang diteliti. (3). Pengumpulan data. Dimana tahap ini data diperoleh di website bps yaitu di link <https://jakarta.bps.go.id/> mengenai data jumlah penduduk di provinsi DKI Jakarta. (4). Analisis Deskriptif. Untuk proses analisis deskriptif yaitu dengan melakukan beberapa tahapan, yaitu pengelompokkan, asosiasi dan skuenial mining. (5). Analisis Hasil. Untuk melakukan Analisa dari hasil

yang akan dinilai keakuratan data dengan menggunakan metode Daves Bouldin Index (DBI). c). Data Mining. Data mining, merupakan suatu langkah ekstraksi untuk memperoleh informasi yang penting yang sifatnya implisit dan belum diketahui. Selain itu, data mining mempunyai keterkaitan dengan berbagai bidang yaitu statistik, *machine learning* (pembelajaran mesin), *pattern recognition*, *computing algorithms*, *database technology*, dan *high performance computing* [14]. d). *Clustering*. Teknik klusterisasi berbeda dengan teknik klasifikasi yang kelas data telah ditentukan sebelumnya. *Clustering* merupakan teknik dengan cara mengelompokkan data secara otomatis tanpa diberitahukan label kelasnya [14]. *clustering* adalah salah satu metode untuk mengelompokkan data, dan sebagai salah satu metode dalam data mining, dimana tujuan *clustering* adalah untuk mengelompokkan data dengan karakteristik yang sama dan data dengan karakteristik yang berbeda ke ‘wilayah atau kelompok’ yang lain [15]. e). *K-Means*. Data *clustering* ini secara umum dilaksanakan menggunakan algoritma dasar seperti berikut ini [16]: 1). Menentukan jumlah *cluster*. Setiap variabel dianggap sebagai *cluster*. jika variabel dan k = jumlah *cluster*, maka k = n. 2). Memilih titik pusat atau *centroid* secara random, Setelah kita menentukan jumlah *cluster*nya maka langkah selanjutnya kita akan memilihkan titik pusat atau *centroid* secara random. 3). Menghitung jarak. kepusat kelompok dengan teori jarak euclidean. Eucclidean distance adalah suatu garis lurus yang mempertemukan antar tujuan yang diteliti. eucclidean distance biasanya dipakai karna perkiraan jarak dalam distance space ini ialah jarak terpendek yang diperoleh antar dua titik yang diperkirakan. untuk mengukur tingkat kemiripan data dengan rumus eucclidean distance digunakan rumus 1.

$$D_{xy} = \sqrt{(X_1 - Y_1)^2 + (X_2 + Y_2)^2 + \dots (X_n - Y_n)^2} \quad (1)$$

Dalam D_{xy} adalah Jarak data ke i ke pusat *cluster* I, untuk X_i ialah Data ke i pada atribut data ke k, sedangkan Y_i ialah Titik pusat ke y pada atribut ke k. 4). Mengelompokkan setiap data kejarak terdekat pusatnya Pembagian ulang data kedalam tiap-tiap kelompok k kedalam *K-Means* dilandaskan dalam perbandingan jarak diantara data *centroid* pada setiap kelompok yang da. 5). Menentukan posisi *cluster* baru dengan metode menjumlahkan atau melakukan perhitungan pada nilai rata-rata dari data yang sudah dimiliki pada *cluster* yang mirip. 6). Jika pusat *cluster* tidak berpindah-pindah lagi maka proses *cluster* kita akhiri, atau diulang ke langkah 3 apabila terdapat data yang masih berpindah-pindah kelompoknya.

3. Hasil dan Pembahasan

Pada penelitian yang sudah dilakukan maka hasil dan pembahasan akan dijelaskan yaitu terdiri atas tujuh:

Menentukan Jumlah *cluster*. Sampel data jumlah kepadatan penduduk yang akan digunakan dalam proses *clustering* adalah data jumlah penduduk yang terdapat pada Badan Pusat Statistik pada tahun 2019-2021 yang terdiri dari kota/kabupaten di DKI Jakarta dan jumlah data sebanyak 6 provinsi dijelaskan pada Tabel 1.

Tabel 1. Data Jumlah Penduduk per Provinsi/Kabupaten/Kota DKI Jakarta 2019-2021 (ribu)

Kab/Kota	2019	2020	2021
Kep Seribu	24.30	27.75	28.24
Jakarta Selatan	2264.70	2226.81	2233.86
Jakarta Timur	2937.86	3037.14	3056.30
Jakarta Pusat	928.11	1056.90	1066.46
Jakarta Barat	2589.93	2434.51	2440.07
Jakarta Utara	1812.91	1778.98	1784.75

Menetapkan nilai k jumlah *cluster*. Melakukan penetapan jumlah *cluster* kepadatan jumlah penduduk sebanyak 3 (tiga) *cluster* (k-3). *cluster* yang dibentuk yaitu *cluster* tertinggi, *cluster* sedang dan *cluster* rendah.

Menentukan nilai *centroid* (pusat *cluster*) awal. Melakukan penentuan nilai *centroid* (pusat *cluster*) awal yang telah ditentukan secara acak berdasarkan nilai variabel data yang di *cluster* sebanyak k yang ditentukan sebelumnya yang dijelaskan pada Tabel 2.

Tabel 2. *centroid* (pusat *cluster*) awal

Kab/Kota	2019	2020	2021
C1 (Kepulauan Seribu)	24.30	27.75	28.24
C2 (Jakarta Selatan)	2264.70	2226.81	2233.86
C3 (Jakarta Timur)	2937.86	3037.14	3056.30

Menghitung *centroid* terdekat. Untuk menghitung *centroid* terdekat yaitu setiap *record* akan ditentukan dimana yang menjadi pusat kelompok terdekatnya. Kemudian *record* tersebut akan ditetapkan menjadi kelompok yang terdekat dari pusat kelompoknya. Kemudian dilakukan perhitungan jarak untuk setiap data yang ada terhadap setiap pusat *cluster* dengan rumus 2.

$$D_{xy} = \sqrt{(X_1 - Y_1)^2 + (X_2 - (X_n - Y_n))^2} \quad (2)$$

Menghitung *centroid* terdekat:

Kota/Kabupaten Kepulauan Seribu

D_{11}

$$D_{11} = \sqrt{(24.30 - 24.30)^2 + (27.75 - 27.75)^2 + (28.24 - 28.24)^2}$$

$D_{11} = 0,00$

D_{12}

$$D_{12} = \sqrt{(2264.70 - 24.30)^2 + (2226.81 - 27.75)^2 + (2233.86 - 28.24)^2}$$

$D_{12} = 3836,67$

D_{13}

$$D_{13} = \sqrt{(2937.86 - 24.30)^2 + (3037.14 - 27.75)^2 + (3056.30 - 28.24)^2}$$

$D_{13} = 5168.60$

Hasil dari keseluruhan perhitungan dapat dihasilkan pengelompokan berdasarkan iterasi ke-1, dijelaskan pada Tabel 3.

Menentukan posisi *cluster*. Setelah itu menentukan posisi *cluster* dari masing-masing data dari jumlah penduduk provinsi DKI Jakarta berdasarkan jarak

minimum data terhadap pusat *cluster*. Data yang memiliki jarak terkecil dari *centroid* akan menjadi anggota pada kelompok tersebut. Penjelasan terdapat pada Tabel 4 dimana posisi data dengan tiap *cluster* pada iterasi ke-1 dengan menggunakan bantuan tanda (1) yang menjelaskan bahwa data yang menjadi anggota pada *cluster*.

Tabel 3. Hasil Perhitungan Jarak Pusat *cluster* Iterasi 1

No	Kab/Kota	<i>cluster</i>			Jarak Terpendek
		C1	C2	C3	
1.	Kepulauan Seribu	0.00	3836.67	5168.60	0,00
2.	Jakarta Selatan	3836.67	0.00	1336.48	0,00
3.	Jakarta Timur	5168.60	1336.48	0.00	0,00
4.	Jakarta Pusat	1718.70	2125.56	3452.52	1718.70
5.	Jakarta Barat	4265.19	437.53	929.49	437.53
6.	Jakarta Utara	3057.98	778.69	2113.13	778.69

Tabel 4. Posisi Data Dengan Tiap *cluster* Pada Iterasi Ke-1

No	Kab/Kota	<i>cluster</i>		
		C1	C2	C3
1.	Kepulauan Seribu	1		
2.	Jakarta Selatan		1	
3.	Jakarta Timur			1
4.	Jakarta Pusat	1		
5.	Jakarta Barat		1	
6.	Jakarta Utara		1	

Menghitung *centroid* Baru. 1. *Centroid* Data Iterasi 2 Untuk menentukan *centroid* baru yaitu dengan menghitung jumlah yang terpilih pada *cluster* kemudian membagikannya berdasarkan banyaknya jumlah *cluster* yang terpilih. Demikian di peroleh dari hasil jarak dari setiap objek pada iterasi ke-1 maka dilanjutkan ke iterasi ke-2 dijelaskan pada Tabel 5.

Tabel 5. *centroid* Data Iterasi 2

	2019	2020	2021
C1 (Kepulauan Seribu)	24.30	27.75	28.24
C2 (Jakarta Timur)	2937.86	3037.14	3056.30
C3 (Jakarta Pusat)	928.11	1056.90	1066.46

Hasil keseluruhan perhitungan dapat diperoleh hasil pengelompokan menurut iterasi 1, dijelaskan pada Tabel 6.

Tabel 6. Hasil Perhitungan Jarak Pusat *cluster* Iterasi 2

No	Kab/Kota	<i>cluster</i>			Jarak Terpendek
		C1	C2	C3	
1.	Kepulauan Seribu	0,00	5168.60	1718.70	0
2.	Jakarta Selatan	3836.67	1336.48	2125.56	1336,48
3.	Jakarta Timur	5168.60	0.00	3452.52	0
4.	Jakarta Pusat	1718.70	3452.52	0.00	0
5.	Jakarta Barat	4265.19	929.49	2558.57	929.49
6.	Jakarta Utara	3057.98	2113.13	1349.15	1349.15

Dengan menentukan posisi *cluster* dari data jumlah penduduk provinsi DKI Jakarta berdasarkan jarak minimum data terhadap pusat *cluster*. Dimana data yang mempunyai jarak terkecil dari *centroid* akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan Tabel 7 posisi data dengan tiap *cluster* pada iterasi ke-2.

Tabel 7. Posisi Data Dengan Tiap *cluster* Pada Iterasi Ke-2

No	Kab/Kota	<i>cluster</i>		
		C1	C2	C3
1.	Kepulauan Seribu	1		
2.	Jakarta Selatan		1	
3.	Jakarta Timur		1	
4.	Jakarta Pusat			1
5.	Jakarta Barat		1	
6.	Jakarta Utara			1

2. *Centroid* Data Iterasi 3. Dengan menentukan *centroid* baru dengan menghitung jumlah yang terpilih pada *cluster* kemudian membagi data sebanyak jumlah *cluster* yang terpilih. Demikian di peroleh hasil jarak dari setiap objek pada iterasi ke-1 maka dilanjutkan ke iterasi ke-2 menurut perhitungan di bawah ini pada Tabel 8.

Tabel 8. *centroid* Data Iterasi 3

<i>cluster</i>	2019	2020	2021
C1 (Kepulauan Seribu)	24.30	27.75	28.24
C2 (Jakarta Timur)	2937.86	3037.14	3056.30
C3 (Jakarta Pusat)	928.11	1056.90	1066.46

Hasil dari keseluruhan perhitungan mampu di amati hasil pengelompokan menurut iterasi 3 pada Tabel 9.

Tabel 9. Hasil Perhitungan Jarak Pusat *cluster* Iterasi 3

No	Kab/Kota	<i>cluster</i>			Jarak Terpendek
		C1	C2	C3	
1.	Kepulauan Seribu	0,00	5168.60	1718.70	0
2.	Jakarta Selatan	3836.67	1336.48	2125.56	1336,48
3.	Jakarta Timur	5168.60	0.00	3452.52	0
4.	Jakarta Pusat	1718.70	3452.52	0.00	0
5.	Jakarta Barat	4265.19	929.49	2558.57	929.49
6.	Jakarta Utara	3057.98	2113.13	1349.15	1349.15

Menentukan posisi data dengan tiap *cluster*.

Melakukan penentuan posisi dari data jumlah penduduk provinsi DKI Jakarta dengan jarak minimum data dari pusat *cluster*. Dimana data yang memiliki jarak terkecil dari *centroid* maka akan menjadi anggota pada kelompok tersebut. Penjelasan pada Tabel 10 merupakan posisi data dengan tiap *cluster* pada iterasi ke-3.

Berdasarkan posisi *cluster* masing-masing data penduduk dan nilai *cluster* hasil iterasi ketiga maka dapat disimpulkan bahawa *Cluster* kesatu dengan jumlah penduduk terendah sebanyak satu kab/kota yaitu kepulauan seribu, *Cluster* kedua dengan jumlah penduduk tertinggi sebanyak tiga kab/kota yaitu Jakarta

Selatan, Jakarta Timur dan Jakarta Barat dan *cluster* ketiga dengan jumlah penduduk sedang sebanyak dua kab/kota yaitu Jakarta Pusat dan Jakarta Utara.

Tabel 10. Posisi Data Dengan Tiap *cluster* Pada Iterasi Ke-3

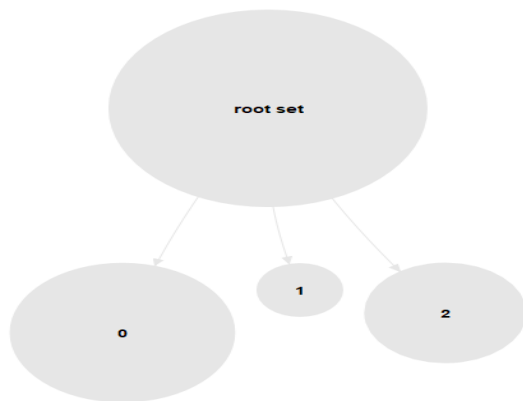
No	Kab/Kota	<i>cluster</i>		
		C1	C2	C3
1.	Kepulauan Seribu	1		
2.	Jakarta Selatan		1	
3.	Jakarta Timur		1	
4.	Jakarta Pusat			1
5.	Jakarta Barat		1	
6.	Jakarta Utara			1

Data hasil *clustering* menggunakan tools rapidminer menjelaskan bahwa terdapat tiga *cluster* yang dimulai dari *cluster* 0, *cluster* 1 dan *cluster* 2. Dapat dilihat pada Gambar 2.



Gambar 2. Folder Fiew

Pada Gambar 3 dapat dijelaskan provinsi mana saja yang termasuk *cluster* jumlah tertinggi, jumlah *cluster* sedang dan jumlah *cluster* terendah. Dimana diperoleh *cluster* 0 sebagai *cluster* tertinggi yaitu kab/kota Jakarta Selatan, Jakarta Timur dan Jakarta Barat, *cluster* 1 sebagai *cluster* terendah yaitu kab/kota Kepulauan Seribu dan *cluster* 2 sebagai *cluster* sedang yaitu kab/kota Jakarta Pusat dan Jakarta Utara.



Gambar 3. Root Set

Pada Gambar 3 diatas Hasil *tree* menjelaskan seberapa besar perbandingan antar *cluster* dengan semua data, dimana diperoleh hasil sebuah pemecahan *root set* atau data induk yang menghasilkan turunan yang disebut dengan *cluster*.

Pada Gambar 4 diatas menunjukkan bahwa *cluster* 1 jumlah kepadatan pendudukan dari tahun 2019-2021 ada peningkatan namun tidak signifikan relatif lebih sedikit sehingga rata-rata kepadatan jumlah penduduknya menjadi yang terendah. Untuk *cluster* 0 jumlah kepadatan penduduknya mengalami peningkatan yang sangat signifikan dari tahun 2019-2021 sehingga rata-rata kepadatan jumlah penduduknya menjadi yang tertinggi. Sedangkan untuk *cluster* 2 rata-rata kepadatan jumlah penduduknya relatif sedang dari tahun 2019-2021.



Gambar 4. Visualisasi Rata-rata jumlah penduduk Provinsi DKI Jakarta

4. Kesimpulan

Kesimpulan yang dihasilkan dari hasil penelitian yaitu dengan mengelompokkan jumlah penduduk berdasarkan kepadatan jumlah penduduk yang ada Kab/Kota DKI Jakarta dapat diterapkan dengan Algoritma *K-Means clustering*. Oleh karena itu diperlukan pemetaan penduduk setiap provinsi di DKI Jakarta menjadi beberapa kelompok (*cluster*). Sehingga dapat memberikan informasi bagi masyarakat yang ini memilih tempat tinggal atau yang berkeinginan berdomisili di Provinsi DKI Jakarta. Dimana hasil pengujian yang dilakukan dalam penelitian ini, iterasi *clustering* pada data jumlah kepadatan penduduk diperoleh tiga iterasi. Untuk hasil yang diperoleh yaitu perhitungan dengan metode *K-Means* dan penggunaan aplikasi rapidminer dimana diperoleh hasil yang bernilai sama yaitu *cluster* dengan jumlah kepadatan tertinggi penduduk sebanyak tiga Kab/Kota yaitu Jakarta Selatan, Jakarta Timur dan Jakarta Barat. *cluster* dengan jumlah kepadatan penduduk terendah sebanyak satu Kab/Kota yaitu Kepulauan Seribu. Serta *cluster* dengan jumlah kepadatan penduduk sedang sebanyak dua Kab/Kota yaitu Jakarta Pusat dan Jakarta Barat. Penelitian ini membuktikan bahwa adanya kepadatan jumlah penduduk di Kab/Kota Provinsi DKI Jakarta yang jumlah kepadatan penduduknya meningkat terus dari tahun 2019-2021 yaitu *cluster* yang terdiri dari tiga Kab/kota yaitu Jakarta Selatan, Jakarta Timur dan Jakarta Barat.

Daftar Rujukan

- [1] I. Bawinti, G. M. V Kawung, and A. Y. Luntungan, "Pengaruh Pengeluaran Pemerintah Dan Investasi Swasta Terhadap Pertumbuhan Ekonomi Di Kabupaten Kepulauan Talaud," *J. Berk. Ilm. Efisiensi*, vol. 18, no. 4, p. 27, 2018.
- [2] S. S. S, A. T. Purba, and F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode *K-Means*," *J. Tek. Inf. dan Komput.*, vol. 2, no. 2, p. 166, 2019, doi: 10.37600/tekinkom.v2i2.115.
- [3] D. S. M. Simanjuntak, I. Gunawan, S. Sumarno, P. Pongingsih, and I. P. Sari, "Penerapan Algoritma *K-Medoids* Untuk Pengelompokan Pengangguran Umur 25 tahun Keatas Di Sumatera Utara," *J. Krisnadana*, vol. 2, no. 2, 2023, doi: 10.58982/krisnadana.v2i2.264.
- [4] M. A. Muminin and W. Hidayat, "Pengaruh Pertumbuhan Ekonomi Dan Jumlah Penduduk Terhadap Tingkat Pengangguran Terbuka Di Kabupaten/Kota Provinsi Jawa Timur Tahun 2011-2015," *J. Ilmu Ekon.*, vol. 1, pp. 374–384, 2017.
- [5] D. Gultom, I. Gunawan, I. Purnamasari, S. R. Andani, and Z. A. Siregar, "Penerapan Algoritma *K-Means* Dalam Pengelompokan Kepadatan Penduduk Menurut Kecamatan di Kabupaten Simalungun," *TIN Terap. Inform. Nusant.*, vol. 2, no. 10, pp. 622–628, 2022, doi: 10.47065/tin.v2i10.1375.
- [6] P. Marpaung and R. F. Siahaan, "Penerapan Algoritma *K-Means clustering* Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan," *J. Sains Komput. Inform.*, vol. 5, no. 1, pp. 503–521, 2021.
- [7] A. Fitria Yulia and H. Widi Nugroho, "Implementasi Algoritma *K-Means* Classifier Sebagai Pendukung Keputusan Penerima Dana Bantuan Siswa Miskin (Studi Kasus: SMKN Sukoharjo)," in *Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2022*, 2022, pp. 48–57.
- [8] Y. R. Sembiring, Saifullah, and R. Winanjaya, "Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)* Vol. 2, No. 2, vol. 2, no. 2, pp. 125–132, 2021.
- [9] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma *K-Means* Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.
- [10] V. Kurnia Bakti and A. Rakhman, "Klasterisasi Dokumen Penelitian Perguruan Tinggi Menggunakan *K-Means clustering*, Sebagai Analisa Penerapan Sistem Temu Kembali," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 10, no. 3, pp. 167–169, 2021, doi: 10.30591/smartcomp.v10i3.2941.
- [11] D. N. Alfiansyah, V. R. S. Nastiti, and N. Hayatin, "Penerapan Metode *K-Means* pada Data Penduduk Miskin Per Kecamatan Kabupaten Blitar," *J. Repos.*, vol. 4, no. 1, pp. 49–58, 2022, doi: 10.22219/repositor.v4i1.1416.
- [12] H. Sibarani, Solikhun, W. Saputra, I. Gunawan, and Z. M. Nasution, "Penerapan Metode *K-Means* Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 1, pp. 154–161, 2022, doi: 10.36040/jati.v6i1.4590.
- [13] L. Y. Hutabarat, "Kmeans-jumlah penduduk siantar," vol. 2, no. 2, pp. 20–26, 2021.
- [14] M. A. Muslim, B. Prasetyo, E. L. H. Mawarni, and A. J. Herowati, *Data Mining Algoritma C4.5, Pertama*, no. December. Semarang: ILKOM UNNES, 2019.
- [15] Mustika *et al.*, *DATA MINING DAN APLIKASINYA*, Pertama. Bandung: Widina Bhakti Persada Bandung, 2021.
- [16] R. Anand and D. U. Jeffrey, *Mining of Massive Datasets*. California: Cambridge University Press, 2011.