

Pengelompokan Jumlah Pengangguran Di Sumatera Utara Menggunakan Algoritma K-Means Dan K-Medoids.

Clustering the number of unemployed individuals in North Sumatra using the K-Means and K-Medoids algorithms.

Rika Setiana^{a,1}, Dira Widi Astuti^{b,2}, M.Safii^{c,3}

^{a, b, c} STIKOM Tunas Bangsa, Pematang Siantar, 21127, Indonesia
rikasetiana957@gmail.com ¹, *m.safii@amiktunasbangsa.ac.id ²

ABSTRAK

Pengangguran merupakan isu utama di negara-negara berkembang, dan salah satunya adalah Provinsi Sumatera Utara. Ketidakseimbangan antara jumlah pencari kerja dan peluang kerja yang tersedia adalah salah satu penyebab utama tingginya tingkat pengangguran di daerah ini. Permasalahan pengangguran ini berdampak besar pada masyarakat, termasuk meningkatnya kejahatan, konflik sosial, dan kemiskinan. Oleh karena itu, penelitian ini bertujuan untuk menganalisis tingkat pengangguran berdasarkan kabupaten/kota di Provinsi Sumatera Utara menggunakan algoritma K-Means dan k-medoids, dengan fokus pada variabel jenjang pendidikan Diploma III dan S1. Pada penelitian ini untuk melakukan sebuah pengelompokan jumlah pengangguran maka diperlukan sebuah data, dan data tersebut pada penelitian ini diperoleh berdasarkan badan pusat statistik provinsi sumatera utara dibidang tenaga kerja, data yang akan dikelompokkan merupakan rata rata dari data jumlah pengangguran tahun 2017-2022. Penelitian ini mengidentifikasi dua kelompok utama pengangguran di Provinsi Sumatera Utara melalui penggunaan algoritma K-Means dan K-Medoids. Melalui pengujian menggunakan algoritma K-Means dan K-Medoids, teridentifikasi dua cluster utama, di mana Cluster 2 mencakup kota Medan. Kedua algoritma memberikan kesimpulan serupa mengenai tingkat pengangguran yang lebih tinggi di kota tersebut. Hasil ini menunjukkan konsistensi dalam identifikasi area dengan tantangan pengangguran, akan tetapi dalam algoritma K-Means menghasilkan 2 kota dan dalam algoritma k-medoids menghasilkan 1 kota. Penelitian ini mengungkapkan bahwa tingkat pengangguran ini terutama berdampak pada individu dengan pendidikan Diploma III dan S1. Analisis data menunjukkan bahwa tingkat pengangguran yang tinggi dalam kelompok ini mungkin disebabkan oleh ketidaksesuaian antara kebutuhan pasar tenaga kerja dan kualifikasi pendidikan yang dimiliki oleh individu. Hasil penelitian ini memiliki implikasi penting bagi pengambil kebijakan dan pembuat keputusan dalam upaya mengurangi tingkat pengangguran di Provinsi Sumatera Utara. Menyadari perbedaan antara kelompok pengangguran ini dapat membantu dalam perencanaan program pelatihan dan pengembangan keterampilan yang lebih efektif, serta pengembangan peluang kerja yang sesuai dengan kualifikasi pendidikan masyarakat setempat.

Kata Kunci : Pengelompokan, Pengangguran, Sumut, K-means, kemiskinan

ABSTRACT

Unemployment is a major issue in developing countries, and one of them is North Sumatra Province. The mismatch between the number of job seekers and available job opportunities is one of the main causes of the high attrition rate in this area. This unemployment problem has a major impact on society, including increasing crime, social conflict and poverty. Therefore, this research aims to analyze poverty levels based on districts/cities in North Sumatra Province using the K-Means and k-medoids algorithms, with a focus on the Diploma III and S1 education level variables. In this research, to group the number of unemployed, data is needed, and the data in this research was obtained based on the Central Statistics Agency of North Sumatra Province in the field of work energy, the data that will include is the average of the data on the number of unemployed for 2017-2022. This research identifies two main groups of poverty in North Sumatra Province through the use of the K-Means and K-Medoids algorithms. Through testing using the K-Means and K-Medoids algorithms, two main clusters were identified, where Cluster 2 includes the city of Medan. The second algorithm provided similar conclusions regarding higher poverty rates in the city. These results show consistency in identifying areas with reduced poverty, but in the K-

Means algorithm it produces 2 cities and in the k-medoids algorithm it produces 1 city. This research reveals that this level of poverty especially impacts individuals with Diploma III and S1 education. Data analysis suggests that high poverty rates in this group may be due to a mismatch between labor market needs and the educational qualifications possessed by individuals. The results of this research have important implications for policy makers and decision makers in efforts to reduce poverty levels in North Sumatra Province. Recognizing the differences between these poverty groups can help in planning more effective training and skills development programs, as well as developing employment opportunities that match the educational qualifications of local people.

Keywords : Clusterring, Unemployment, North Sumatra, K-means, poverty

Disubmit: 09 Oktober

*Info Artikel :
Direview: 10 November*

Diterima : 20 Desember

Copyright © 2022 - Journal UPU. All rights reserved.

1. PENDAHULUAN

Pengangguran adalah istilah yang diberikan kepada seseorang atau sekelompok orang yang sedang tidak bekerja atau tidak memiliki penghasilan tetap. Pengangguran mencakup orang-orang yang tidak memiliki pekerjaan tetapi sedang aktif mencari pekerjaan, atau tengah menyiapkan diri untuk memulai usaha, atau merasa putus asa karena sulitnya mendapatkan pekerjaan, atau bahkan sudah diterima bekerja namun belum memulai tugasnya di suatu wilayah. Masalah pengangguran menjadi salah satu perhatian utama dalam negara-negara berkembang (Akramunnisa & Fajriani, 2020). Di antara jenis-jenis pengangguran yang ada, terdapat pengangguran terbuka. Biasanya, pengangguran terbuka ini banyak terjadi pada generasi muda yang baru saja menyelesaikan pendidikan menengah dan tinggi. Ada kecenderungan bagi mereka yang baru lulus pendidikan untuk mencari pekerjaan yang sesuai dengan keinginan mereka (Pastia & Dikananda, 2023). Terjadinya tingkat pengangguran disebabkan oleh beberapa faktor, salah satunya adalah ketidakseimbangan antara jumlah pencari kerja dan kesempatan kerja yang tersedia, yang mengakibatkan ketidak mungkinan tercapainya tingkat kesempatan kerja yang seimbang (Agustina et al., 2023).

Salah satu isu utama dalam pembangunan ekonomi daerah adalah pengangguran. Salah satu indikasi kegagalan pembangunan di suatu wilayah adalah tingginya tingkat pengangguran. Hal ini terjadi akibat ketidakseimbangan antara jumlah tenaga kerja yang ada dengan lapangan kerja yang tersedia. Permasalahan sering muncul di beberapa daerah di Indonesia, termasuk di Provinsi Sumatera Utara, dan berdampak signifikan terhadap masyarakat lainnya, antara lain meningkatnya kejahatan, konflik antar kelompok, dan meningkatnya angka kemiskinan. Penyebab dari keadaan ini adalah tingkat pengangguran (Sutris et al., 2023). Badan Pusat Statistik mengumpulkan data sosial ekonomi nasional atau biasa disebut Susenas setiap tahunnya. Pengumpulan data dilakukan untuk mengetahui status masyarakat. Pengumpulan data dimulai dari data diri, status, nama orang tua, golongan darah hingga pekerjaan masing-masing warga (Listiani et al., 2017). Berdasarkan data yang telah dikelola dari Badan Pusat Statistik Sumatera Utara bagian tenaga kerja bahwasannya kabupaten/kota yang terdapat pada Sumatera Utara memiliki tingkat pengangguran terbuka yang cukup tinggi pada jenjang Pendidikan S1 dan jenjang Pendidikan Diploma I/II/III. Berdasarkan Badan Pusat Statistik Sumatera Utara dilihat bahwasannya tingkat pengangguran terbuka di Sumatera Utara pada tahun terakhir yaitu tahun 2022 memiliki tingkat pengangguran yang cukup tinggi ketika sebanyak 10709 jiwa Pendidikan Diploma I/II/III dan 48527 jiwa pada jenjang Pendidikan S1. Jika tidak ada langkah signifikan untuk mengurangi persentase pengangguran di Indonesia, tingkat pengangguran akan terus meningkat setiap tahunnya. Pekerja memiliki peran kunci dalam pertumbuhan ekonomi di Indonesia, oleh karena itu pemerintah perlu mengimplementasikan upaya nyata untuk menyediakan lapangan kerja sesuai dengan pemetaan, pengelompokan, dan karakteristik wilayah Indonesia.

Tingkat pengangguran yang tinggi mempengaruhi tingkat kemiskinan di beberapa daerah. Oleh karena itu, mengingat pengangguran merupakan salah satu faktor utama yang menentukan tingkat kemiskinan, maka perlu dilakukan penelitian lebih lanjut mengenai tingkat pengangguran. Salah satu solusinya adalah dengan menggunakan data mining menggunakan teknik clustering dengan algoritma k-means untuk mengelompokkan data pengangguran (Andryadi & Nisa, 2023). Pemerintah terus berupaya untuk membuka lapangan pekerjaan sebanyak mungkin agar kebutuhan tenaga kerja terpenuhi bagi para pencari kerja. Setiap provinsi di Indonesia memiliki kebutuhan lapangan pekerjaan yang berbeda, hal ini bergantung pada tingkat pengangguran yang ada di masing-masing provinsi. Untuk memenuhi kebutuhan lapangan kerja, perlu dilakukan pengelompokan provinsi berdasarkan tingkat penganggurannya. Hal ini bertujuan agar dapat digunakan sebagai acuan dalam pengambilan keputusan terkait pemenuhan lapangan kerja (Almeira et al., 2021). Dalam dunia data mining, ada teknik yang disebut dengan clustering. Teknik ini termasuk dalam kategori unsupervised learning. Tujuan clustering adalah untuk mengelompokkan data yang tidak memiliki label berdasarkan karakteristiknya ke dalam beberapa kelompok. Dalam data mining terdapat dua metode clustering yaitu K-Means dan K-medoids. Metode K-means Clustering memiliki keunggulan yang sangat mudah diimplementasikan dan merupakan metode yang paling banyak digunakan dalam pengklasteran (Wahyuni & Utamajaya, 2022). K-Means, dengan pendekatan centroid, dapat memberikan hasil yang efisien dalam hal komputasi, sementara K-Medoids, yang berfokus pada medoid sebagai pusat klaster, cenderung lebih tahan terhadap outliers (Sihombing & Sihombing, 2021).

Berdasarkan penelitian sebelumnya (Mutmainah et al., 2023) tentang pengelompokan data pengangguran di Jawa Barat menggunakan algoritma k-means, disimpulkan bahwa algoritma tersebut mampu mengklasifikasikan data pengangguran di Jawa Barat berdasarkan beberapa atribut. Dalam penelitian selanjutnya, yang merupakan penelitian (Aziz et al., 2023), telah dibahas mengenai klasterisasi tenaga kerja terbuka dengan menggunakan algoritma k-means. Dari penelitian ini, dapat disimpulkan bahwa penggunaan metode k-means clustering pada data mining berhasil diterapkan dengan menggunakan data dari 34 provinsi. Dalam penelitian (Harits et al., 2022) mengenai pengelompokan tingkat pengangguran dengan menggunakan algoritma k-means di wilayah Kalimantan Timur, ditemukan bahwa proses pengelompokan ini dapat memberikan perspektif awal mengenai karakteristik tingkat pengangguran terbuka di Kalimantan Timur. Dalam penelitian ini, telah diungkapkan adanya dua klaster tingkat pengangguran yang memiliki karakteristik yang cukup berbeda. Dalam penelitian ini (Tanjung et al., 2022), dibahas mengenai pengelompokan pengangguran di Indonesia dengan menggunakan algoritma k-means. Hasil penelitian menyimpulkan bahwa metode K-means dapat digunakan untuk memproses data persentase pengangguran berdasarkan provinsi. Pertumbuhan ekonomi yang pesat juga membawa dampak signifikan terhadap tingkat pengangguran di daerah ini. Pengukuran dan pemahaman yang baik tentang pola pengangguran menjadi krusial untuk menginformasikan kebijakan ekonomi yang efektif. Dalam penelitian mengenai pengelompokan jumlah pengangguran dengan menggunakan algoritma K-Medoids menyimpulkan bahwa metode K-Medoids Clustering dapat dianggap sebagai pendekatan yang efektif dalam mengelompokkan dan menganalisis data tingkat pengangguran, khususnya dalam konteks Sumatera Utara. Pada penelitian (Naya et al., 2023) mengenai pengelompokan pengangguran terbuka di Indonesia menyimpulkan bahwa Cluster 1 mencakup 9% dari total wilayah Indonesia. Cluster 2 mencakup 29% dari total wilayah Indonesia dan Cluster 3 mencakup 62% dari total wilayah Indonesia.

Oleh karena itu, penelitian ini bertujuan untuk melakukan perbandingan antara dua metode pengelompokan data, yaitu algoritma K-Means dan K-Medoids, dalam mengkategorikan jumlah pengangguran di Sumatera Utara. Dengan memanfaatkan keunggulan masing-masing algoritma, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam memberikan wawasan lebih mendalam terkait karakteristik pengangguran di wilayah tersebut. Dengan melakukan perbandingan antara algoritma K-Means dan K-Medoids, penelitian ini bertujuan untuk memberikan kontribusi terhadap pemahaman yang lebih baik tentang pola pengangguran di Sumatera Utara. Hasil penelitian

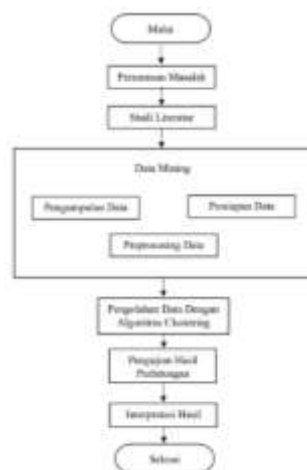
diharapkan dapat menjadi dasar untuk pengambilan keputusan yang lebih tepat dalam merancang kebijakan ekonomi di tingkat regional. Melalui penerapan metode ini, diharapkan dapat ditemukan pola-pola yang dapat membantu pihak berwenang untuk merancang langkah-langkah intervensi yang lebih efektif guna mengurangi tingkat pengangguran di Sumatera Utara.

2. METODE

Kerangka Kerja Penelitian

Dalam setiap penelitian membutuhkan sebuah kerangka kerja penelitian. Pada penelitian ini dilakukan sebuah proses pengumpulan data yang di peroleh dari Badan Pusat Statistik sumut pada bagian tenaga kerja mengenai pengangguran yang diambil berdasarkan link <https://sumut.bps.go.id/indicator/6/587/1/penduduk-sumatera-utara-berumur-15-tahun-ke-atas-yang-termasuk-pengangguran-terbuka-menurut-kabupaten-kota-dan-pendidikan-tertinggi-yang-ditamatkan-daerah-perkotaan-perdesaan-laki-laki-perempuan-.html> Dalam melakukan penelitian langkah awal yang di lakukan adalah menentukan perumusan masalah yang terjadi di masa kini, setelah mendapatkan perumusan masalah maka langkah selanjutnya melakukan studi Pustaka atau pengumpulan berbagai referensi dari jurnal atau internet. Langkah selanjutnya yang dilakukan adalah persiapan data dan melakukan standarisasi data. Pada penelitian ini dalam pengolahan data menggunakan algoritma K-means dan k-medoids untuk melakukan pengelompokan data .

Dalam algoritma K-means langkah awal yang perlu dilakukan adalah penentuan jumlah kluster yang akan digunakan, lalu menghitung jarak cluster dan menganalisis cluster yang telah didapatkan. Setelah mendapatkan iterasi 1 dalam algoritma K-means melanjutkan iterasi 2 berdasarkan pengelompokan cluster yang telah didapatkan dan melakukan langkah tersebut berulang kali sehingga mendapatkan persamaan terhadap cluster dan melakukan interpretasi hasil. Setelah mendapatkan hasil dari pengelompokan k-means maka perlu dilakukan pengelompokkan dengan algoritma yang kedua yaitu dengan menggunakan algoritma k-medoids dengan langkah pertama menentukan jumlah cluster yang akan digunakan dengan cara mengambil nilai secara acak dari data, langkah kedua dalam melakukan pengelompokkan dengan menggunakan algoritma k-medoids dengan menghitung jarak terdekat dan setelah itu pemilihan jumlah k-medoids baru seperti langkah sebelumnya, setelah setelah maka lakukan perhitungan selisih kedekatan antara iterasi 1 dan iterasi 2 jika selisih kedekatan yang terdapat pada iterasi 2 lebih besar maka perhitungan di berhentikan dan jika jumlah selisih kedekatan iterasi 2 lebih kecil maka perhitungan dilanjut pada iterasi berikutnya.



Gambar 1. Kerangka Kerja Penelitian

Data Mining

Salah satu cabang kecerdasan buatan yang mampu memproses data dalam jumlah besar dan mencarinya secara otomatis disebut dengan data mining. Oleh karena itu, peranannya sangat penting dalam beberapa bidang kehidupan, antara lain industri, keuangan, cuaca, ilmu pengetahuan dan teknologi (Damanik et al., 2020). Data mining merupakan teknik yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengidentifikasi data yang ada dengan tujuan menghasilkan informasi yang bermanfaat dan berpengetahuan yaitu data mining (Safira et al., 2020). Ada beberapa metode yang dapat diterapkan dalam data mining. Metode tersebut meliputi model Forecasting, Classification, Association, dan Clustering (Maulidiah et al., 2023). Data mining adalah proses mencari pola atau tren yang diinginkan dalam kumpulan data besar dengan tujuan membantu pengambilan keputusan di masa depan (Wahyudi & Pujiastuti, 2020).

K-means

Algoritma K-Means ialah satu algoritma pengkelompokan untuk menggolongkan data ke dalam beberapa kumpulan atau kelompok. Metode pengelompokan data non-hierarkis yang digunakan dalam algoritma K-Means adalah partition clustering yang berdasarkan titik pusat (Anggraini et al., 2022). Suatu teknik yang membagi sekumpulan data ke dalam kelompok-kelompok berdasarkan kesamaan yang telah ditentukan disebut clustering (Putrie et al., 2021). Dalam mendefinisikan cluster berdasarkan data yang tersedia, diperlukan diagram untuk membantu mengidentifikasi alur komputasi yang menemukan hasil penerapan cluster pada data yang akan diolah (Pasaribu et al., 2021). Langkah-langkah algoritma K-Means Clustering adalah sebagai berikut (Badri & Habibi, 2022):

1. Pertimbangkan jumlah kluster yang ingin diatur dengan menetapkan pusat kluster k.
2. Hitung setiap titik data hingga ke pusat cluster

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

3. Kelompokkan data ke dalam cluster dengan jarak terpendek menggunakan persamaan

$$\text{Min } \sum_k^k = d_{ik} = d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (2)$$

4. Hitunglah pusat cluster dengan persamaan berikut

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

5. Lakukan pengulangan langkah dua hingga empat agar tidak ada data yang berpindah ke cluster lain.

K-Medoids

K-Medoids merupakan suatu teknik klastering yang digunakan untuk mengelompokkan objek-objek ke dalam kumpulan cluster (Luchia et al., 2022). Berikut adalah langkah-langkah utama dalam algoritma k-medoids (Mulyawan et al., 2023):

1. Lakukan Inisialisasi

Pilih secara acak k titik data sebagai medoids awal. Setiap titik data akan diatributkan ke salah satu medoids sesuai dengan jarak terkecil.

2. Lakukan Perhitungan Jarak

Gunakan metrik jarak, seperti jarak Manhattan atau jarak Euclidean, untuk mengukur jarak antara setiap titik data dengan setiap medoid.

3. Pemilihan Medoids Baru

Untuk setiap klaster, pilih titik data yang memiliki total jarak terkecil dari klaster tersebut sebagai medoid baru. Iteratif update posisi medoid untuk meminimalkan total jarak antara titik data dalam klaster dengan medoidnya.

4. Perubahan Cluster

Hitung ulang jarak antara setiap titik data dengan medoid terdekat setelah perubahan medoid. Lakukan klasifikasi ulang titik data ke klaster yang memiliki medoid terdekat.

5. Iterasi

Ulangi langkah 3 dan 4 sampai tidak ada perubahan pada medoid atau iterasi mencapai batas tertentu.

3. HASIL DAN PEMBAHASAN

Pengangguran merupakan salah satu masalah sosial yang sering menjadi perhatian utama dalam pembangunan suatu daerah, dengan pengelompokan jumlah pengangguran terhadap masing-masing kota atau kabupaten maka akan pihak pemerintah akan mengetahui kota mana yang memiliki jumlah pengangguran yang paling tinggi. Dalam penelitian ini pengelompokan dilakukan dengan menggunakan algoritma k-means yang dilakukan seperti implementasi berikut. Berikut ini adalah beberapa contoh metode perhitungan manual yang digunakan dalam menggunakan algoritma k-means clustering, kami akan melakukan pengelompokan data pengangguran terbuka per provinsi. Pada penelitian ini menggunakan metode k-means untuk melakukan clustering berdasarkan tingkat kab/kota di provinsi sumut berdasarkan indikator tingkat pengangguran dengan menggunakan langkah-langkah berikut :

1. Pada penelitian ini data diperoleh dari tahun 2017-2022 dan dicari nilai rata-ratanya sehingga didapatkan fitur x dan y. Fitur x pada penelitian ini merupakan jenjang Pendidikan Diploma I/II/III dan fitur y pada penelitian ini merupakan jenjang Pendidikan S1. Langkah awal yang dapat dilakukan setelah mendapatkan rata rata maka melakukan penentuan jumlah cluster, pada penelitian ini jumlah cluster yang akan digunakan adalah 2 cluster. Pada fitur x memiliki cluster 1 = 326 dan fitur y = 1021, pada cluster 2 fitur x memiliki cluster = 259 dan fitur y=567. Kedua cluster tersebut diambil dari data pengangguran secara acak.
2. Melakukan pencarian nilai C1 dan C2 antara setiap pusat massa objek dan titik yang sesuai dengan menggunakan Jarak Euclidean pada iterasi 1.
3. Langkah berikutnya setelah melakukan perhitungan jumlah cluster setiap objek maka perlu dilakukan perhitungan jarak terdekat dengan cara mencari nilai terkecil dari cluster 1 dan cluster 2, apabila nilai terkecil terdapat pada cluster 1 maka cluster 1 mendapatkan nilai 1 dan cluster 2 mendapatkan nilai 0 dan begitu juga sebaliknya.

Tabel 1. Hasil Iterasi 1

No	Kab/Kota	DI/II/III	S1	c1	c2	min
1	Nias	178	417	364260	22407	22407
2	Mandailing Natal	326	1021	0	206182	0
3	Tapanuli Selatan	99	417	364640	22560	22560
4	Tapanuli Tengah	235	1143	15016	331993	15016
5	Tapanuli Utara	259	567	206182	0	0
6	Toba	170	192	686844	140464	140464
.....						
32	Gunungsitoli	281	756	70093	35869	35869

Dapat dilihat pada tabel 1 disimpulkan bahwa pada iterasi 1, c1 memiliki 13 data yaitu (2,4,7,8,9,12,13,17,21,28,29,30,31) dan c2 memiliki 19 data yaitu (1,3,5,6,10,11,14,15,16,18,19,20,22,23,24,25,26,28,32). Langkah selanjutnya untuk mencari iterasi kedua maka di perlukan cluster 1 dan cluster 2. Dalam mencari cluster satu dan cluster 2 untuk iterasi 2 pada fitur x dan fitur y maka caranya jumlah cluster 1 dibagi dengan jumlah data cluster 1 dan jumlah nilai cluster 2 dibagi jumlah data cluster 2. Setelah dilakukan pencarian nilai C1 dan C2 pada iterasi kedua maka di hasilkan jumlah C1 pada fitur x yaitu 1094 dan y(jenjang s1)=3309, jumlah C2 fitur x yaitu = 232 dan C2 pada fitur y yaitu 441. Untuk pencarian nilai cluster 1 dan cluster 2 maka di lakukan perhitungan manual berikut ini:

Tabel 2. Hasil Iterasi 2

No	Kab/Kota	DI/II/III	S1	c1	c2	min
1	Nias	178	417	8364810	630	630
2	Mandailing Natal	326	1021	5238564	335817	335817
3	Tapanuli Selatan	99	417	8366335	720	720
4	Tapanuli Tengah	235	1143	4694393	492221	492221
5	Tapanuli Utara	259	567	7522816	15757	15757
6	Toba	170	192	9718419	62187	62187
.....						
32	Gunungsitoli	281	756	6520101	99117	99117

Dapat dilihat pada tabel 3.2 bahwasannya disimpulkan iterasi mendapatkan nilai cluster 1 sebanyak 3 data dan nilai cluster 2 sebanyak 29 data, sehingga disimpulkan bahwa iterasi 1 dan iterasi 2 belum memiliki nilai yang sama dan harus melanjutkan iterasi 3. Pada iterasi 3 dalam mencari nilai C1 dan C2 sama dengan mencari C1 dan C2 pada iterasi 2. Nilai C1 pada fitur x iterasi 3 yaitu 3097 dan nilai y yaitu 10010, nilai C2 pada fitur x yaitu 322 dan nilai y yaitu 737. Dalam melakukan pencarian nilai cluster 1 dan cluster 2 menggunakan perhitungan di bawah ini :

Tabel 3. Hasil Iterasi 3

No	Kab/Kota	DI/II/III	S1	c1	c2	min
1	Nias	178	417	92017377	102440	102440
2	Mandailing Natal	326	1021	80804892	80421	80421
3	Tapanuli Selatan	99	417	92022252	102678	102678
4	Tapanuli Tengah	235	1143	78623596	164717	164717
5	Tapanuli Utara	259	567	89173087	29106	29106
6	Toba	170	192	96389505	297272	297272
.....						
32	Gunungsitoli	281	756	85633162	398	398

Berdasarkan perhitungan yang telah dilakukan maka disimpulkan bahwa iterasi 3 mendapatkan cluster 1 sebanyak 2 data dan cluster 2 sebanyak 30 data, sehingga disimpulkan jumlah cluster 3 dan cluster 2 tidak sama dan harus melanjutkan pada iterasi ke 4. Setelah melakukan perhitungan diatas maka didapatkan nilai C1 pada fitur x yaitu 3839 dan fitur y yaitu 13913, nilai C2 pada fitur x yaitu 365 dan fitur y yaitu 786. Dalam mencari nilai cluster 1 dan cluster 2 pada iterasi ke empat maka dilakukan perhitungan berikut ini :

Tabel 4. Hasil Iterasi 4

No	Kab/Kota	DI/II/III	S1	c1	c2	min
1	Nias	178	417	182147927	136111	136111
2	Mandailing Natal	326	1021	166224367	55141	55141
3	Tapanuli Selatan	99	417	182154753	136374	136374
4	Tapanuli Tengah	235	1143	163089275	127511	127511
5	Tapanuli Utara	259	567	178137091	48182	48182
6	Toba	170	192	188278657	352946	352946
.....						
32	Gunungsitoli	281.4	756	173118978	979.2212	979.2212

Berdasarkan perhitungan yang telah dilakukan disimpulkan bahwa iterasi ke 4 memiliki jumlah cluster yang sama seperti iterasi ke 3 yaitu C1 2 data dan C2 30 data, apabila nilai cluster sudah memiliki jumlah cluster yang sama pada objek yang sama perhitungan iterasi dapat dihentikan. Disimpulkan bahwa c1 yaitu kota medan dan kota deli Serdang memiliki jumlah nilai pengangguran yang sangat tinggi baik jenjang Pendidikan diploma III maupun jenjang Pendidikan S1. Maka diambil kesimpulan kota Medan dan kota Deli Serdang memiliki jumlah nilai pengangguran yang sangat tinggi, baik pada jenjang

Pendidikan Diploma III maupun Pendidikan S1. Hal ini menunjukkan adanya masalah serius terkait tingkat pengangguran di kedua kota tersebut, independen dari tingkat pendidikan. Oleh karena itu, perlu dilakukan analisis lebih lanjut untuk memahami faktor-faktor yang menyebabkan tingkat pengangguran yang tinggi dan menyusun strategi atau kebijakan yang dapat membantu mengatasi masalah tersebut di kedua kota. Penelitian ini memberikan gambaran yang kuat tentang masalah pengangguran di wilayah ini, khususnya di dua kota tersebut. Hasil ini mengindikasikan bahwa langkah-langkah perbaikan dan kebijakan yang lebih konkret perlu diterapkan untuk mengatasi masalah ini, terutama dalam hal peningkatan peluang kerja dan pendidikan di kedua jenjang.

Dalam penelitian ini pengelompokan kedua menggunakan metode yang berbeda yaitu dengan menggunakan metode *k-medoids*. Dalam melakukan pengelompokan menggunakan *k-medoids* menggunakan data yang sama seperti melakukan pengelompokan pada algoritma *k-means*. Dalam melakukan proses perhitungan *k-medoids* memiliki kemiripan rumus dalam pencarian cost dan jarak terdekat, hanya saja dalam *k-medoids* melakukan pencarian selisih kedekatan. Dalam melakukan perhitungan *k-medoids* dapat dilakukan dengan langkah-langkah berikut :

1. Melakukan penentuan cost yang akan digunakan, dalam penelitian ini dalam melakukan pengelompokan menggunakan algoritma *k-medoids* maka dipakai 2 cluster. Dalam *K-Medoids* jumlah nilai cluster diambil secara acak dari data yang digunakan, Nilai cluster 1 diambil dari data ke 3 dan nilai cluster ke 2 diambil dari data ke 12.
2. Setelah mendapatkan nilai cluster pada iterasi 1 cluster 1 dan cluster 2 maka diperlukan penentuan nilai jarak terdekat dan lakukan kelompok setiap cluster. Dapat dilihat pada tabel 5 bahwasanya algoritma *k-medoids* dalam melakukan pengelompokan pada iterasi 1 menghasilkan 30 data pada cluster 1 dan 2 data pada cluster 2 dan menghasilkan jumlah kedekatan 25823.08.

Tabel 5. Hasil *K-Medoids* Iterasi 1

No	Kab/Kota	DI/II/III	S1	c1	c2	min
1	Nias	178	417	78	9386	78
2	Mandailing Natal	326	1021	645	8765	645
3	Tapanuli Selatan	99	417	0	9403	0
4	Tapanuli Tengah	235	1143	738	8665	738
5	Tapanuli Utara	259	567	219	9223	219
.....						
32	Gunungsitoli	281	756	385	9033	385

Tabel 6. Hasil *K-Medoids* Iterasi 2

No	Kab/Kota	DI/II/III	S1	c1	c2	min
1	Nias	178	417	1242	18602	1242
2	Mandailing Natal	326	1021	658	17982	658
3	Tapanuli Selatan	99	417	1281	18625	1281
4	Tapanuli Tengah	235	1143	644	17892	644
5	Tapanuli Utara	259	567	1072	18436	1072
.....						
32	Gunungsitoli	281	756	897	18248	897

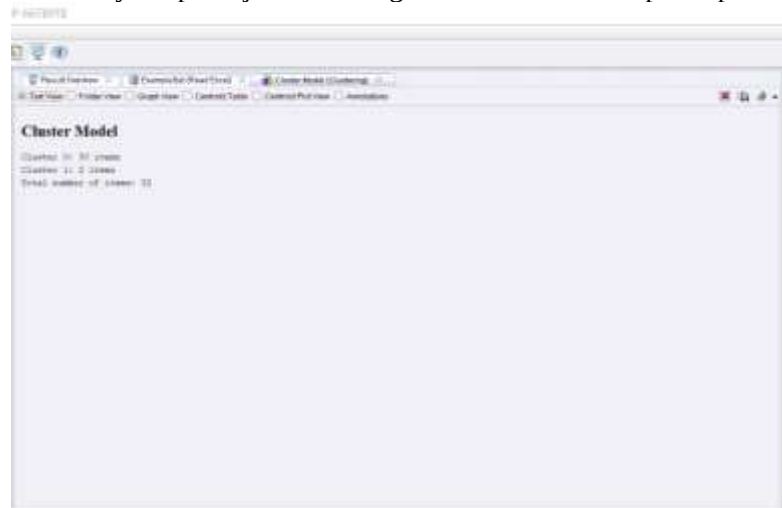
Dapat dilihat pada tabel 6 menghasilkan jumlah jarak kedekatan sebesar 3.6773,2. Berdasarkan hasil pengujian yang telah dilakukan maka dilihat bahwasannya algoritma *k-means* menghasilkan 30:2, yaitu 30 cluster 1 dan 2 data pada cluster 2 yaitu kota medan dan deli serdang, sementara algoritma *k-medoids* menghasilkan 30:1 yaitu 30 cluster 1 dan 1 data pada cluster 2 yaitu kota medan. Dalam penelitian ini Cluster 2 merupakan kab/kota yang memiliki jumlah tingkat pengangguran yang lebih tinggi diantara lainnya, dan cluster 1 merupakan jumlah nilai yang memiliki nilai lebih rendah diantara yang lainnya. Berdasarkan informasi di atas, kedua algoritma (*K-Means* dan *K-Medoids*) menghasilkan kesimpulan

yang serupa, yaitu bahwa Kota Medan cenderung memiliki tingkat pengangguran yang lebih tinggi dibandingkan dengan daerah lainnya. Jika fokus pada keberagaman dalam cluster, K-Means mungkin memberikan gambaran yang lebih terperinci dengan menunjukkan dua kota dalam cluster yang sama sehingga pihak pemerintah akan lebih luas dalam memerhatikan kota yang memiliki jumlah pengangguran yang tinggi. Dengan menggunakan hasil clustering sebagai dasar untuk analisis lebih lanjut dan pengambilan keputusan yang lebih mendalam. Selain itu, interpretasi hasil clustering harus mempertimbangkan konteks sosioekonomi dan politik yang relevan.

3.2 Hasil Pengujian

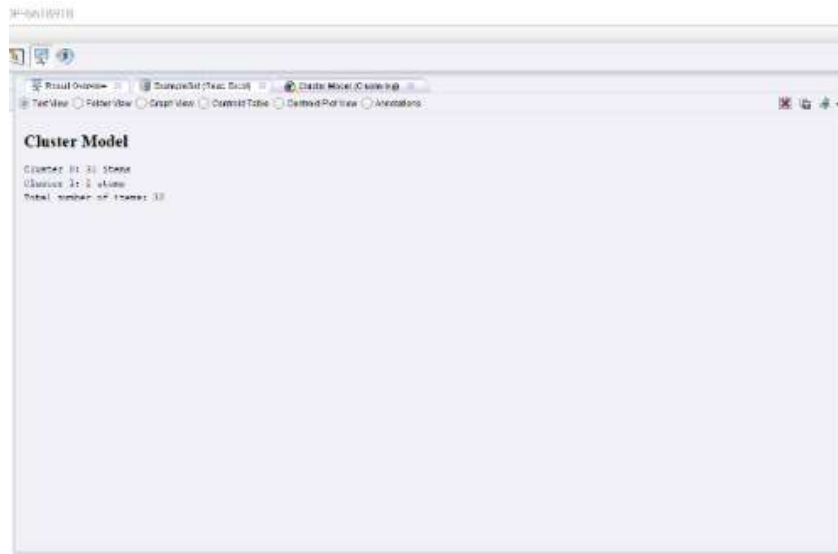
Dalam penelitian ini, aplikasi RapidMiner digunakan untuk menyelesaikan proses pengelompokan data pengangguran di provinsi Sumatera Utara. Setelah jumlah pengangguran data siap diolah, data tersebut diimpor ke dalam rapidminer menggunakan operator ReadExcel. Pada bagian konfigurasi impor wizard, pilih file excel yang akan digunakan dan tentukan lembar tempat data tersimpan. Setelah data diimpor ke RapidMiner, langkah selanjutnya adalah menghubungkannya dengan operator clustering. Di tahap ini, kita perlu menentukan algoritma yang akan digunakan, yaitu algoritma K-means. Data ini terhubung dengan node-node pada operator dan kemudian hanya perlu terhubung ke node akhir untuk menunjukkan hasil akhirnya. Dari hasil ini, dapat disajikan dalam berbagai grafik yang mudah dipahami. Dengan menggunakan konfigurasi data dan operator, penulis menggunakan tools RapidMiner sebagai alternatif untuk proses ekstraksi data menggunakan teknik data mining.

Berdasarkan hasil pengujian dengan menggunakan perangkat lunak RapidMiner 5.3 dan perhitungan manual menggunakan metode K-means pada aplikasi Microsoft Excel 2010, dapat disimpulkan bahwa hasil perhitungan manual dan hasil yang dihasilkan oleh RapidMiner 5.3 sama. Hal ini menunjukkan bahwa perhitungan manual yang dilakukan dalam penelitian ini telah diverifikasi dan divalidasi dengan baik, sehingga dapat diandalkan sebagai representasi yang tepat dari data. Hasil ini memberikan kepercayaan bahwa data yang digunakan dalam penelitian memiliki keakuratan yang baik, dan bahwa algoritma K-means yang digunakan di kedua platform memberikan hasil yang konsisten. Dengan demikian, penelitian ini memberikan bukti yang kuat bahwa perhitungan manual yang dilakukan dalam analisis data menggunakan Microsoft Excel 2010 sesuai dengan metode K-means pada RapidMiner 5.3, dan hasilnya dapat dijadikan sebagai dasar untuk kesimpulan penelitian lebih lanjut.



Gambar 2. Hasil Pengujian K-Means

Dapat dilihat pada gambar 4 bahwasannya hasil perhitungan manual pada algoritma *k-medoids* yang telah dilakukan sama dengan pengujian validasi yang dilakukan rapid miner yaitu menghasilkan nilai cluster 1 30 data dan nilai cluster 2 yaitu 1 data.

Gambar 3. Hasil Pengujian *K-Medoids*

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terkait analisis algoritma *K-Means* dan *K-Medoids* dalam mengelompokkan jumlah pengangguran berdasarkan kabupaten/kota di Provinsi Sumatera Utara, dapat diambil beberapa kesimpulan yang memberikan wawasan penting terkait pengangguran di wilayah tersebut :

1. Penelitian ini mengidentifikasi dua kelompok utama pengangguran di Provinsi Sumatera Utara melalui penggunaan algoritma *K-Means* dan *K-Medoids*. Melalui pengujian menggunakan algoritma *K-Means* dan *K-Medoids*, teridentifikasi dua cluster utama, di mana Cluster 2 mencakup kota Medan. Kedua algoritma memberikan kesimpulan serupa mengenai tingkat pengangguran yang lebih tinggi di kota tersebut. Hasil ini menunjukkan konsistensi dalam identifikasi area dengan tantangan pengangguran.
2. Penelitian ini menggarisbawahi nilai algoritma *K-Means* dan *K-Medoids* sebagai alat analisis yang efektif dalam mengelompokkan data sosial-ekonomi. Ini menunjukkan bahwa penggunaan analisis *cluster* dapat membantu memahami pola pengangguran yang ada dan memberikan landasan ilmiah bagi pengambilan keputusan terkait kebijakan ekonomi dan tenaga kerja.
3. Dalam menghadapi permasalahan pengangguran ini, penelitian ini memberikan landasan untuk menyusun rekomendasi kebijakan yang dapat diimplementasikan oleh pemerintah daerah Sumatera Utara. Langkah-langkah selanjutnya melibatkan pemahaman lebih mendalam terhadap faktor-faktor determinan tingkat pengangguran di kedua kota tersebut, serta penyusunan strategi intervensi yang sesuai.

REFERENSI

- [1] Agustina, D., Mukti, Y. I., & Muntari, S. (2023). Integrasi Particle Swam Optimization Menggunakan K-Means untuk Klasterisasi Pengangguran di Kota Pagar Alam. *Junal Pusat Akses Kajian*, 3(1), 34–41.
- [2] Akramunnisa, & Fajriani. (2020). K-MEANS Clustering Analysis Pada Persebaran Tingkat Pengangguran Kabupaten / Kota Di Sulawesi Selatan. *Jurnal Varian*, 3(2), 103–112. <https://doi.org/10.30812/varian.v3i2.652>

- [3] Almeida, D., Juanda, G. G., Sani, S. R., & Fachrin, A. A. (2021). Analisis Multidimensional Scaling dan k-Means Clustering untuk Pengelompokan Provinsi Berdasarkan Tingkat Pengangguran. *SEMINAR NASIONAL STATISTIKA X*.
- [4] Andryadi, A. A., & Nisa. (2023). PENERAPAN METODE ALGORITMA K-MEANS UNTUK DATA PENGANGGURAN JAWA BARAT. *Jurnal Wahana Informatika*, 2(2), 277–283.
- [5] Anggraini, R., Haerani, E., & Afrianty, I. (2022). Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 1840–1849. <https://doi.org/10.30865/jurikom.v9i6.5145>
- [6] Aziz, A., Norhasanah, S., Ali, U. D., Hulu, M. B., Timur, K. K., Tengah, K., Informasi, J. S., Darwan, U., & Sampit, A. (2023). Klastisasi Data Tenaga Kerja Terbuka Menurut Provinsi Dengan Penggunaan Algoritma K - Means. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(3), 444–456.
- [7] Badri, F., & Habibi, A. (2022). Implementasi Metode K-Mean Clustering Analysis pada Pengelompokan Pengangguran di Indonesia sebagai Dampak dari Pandemi Covid-19. *Journal of Computer Science and Applied Informatics*, 4(2), 171–179.
- [8] Damanik, A. R., Defit, S., Hartama, D., Ilmi, P. P. P. A. N. W. F., & Zer, R. H. (2020). Implementasi Metode K-Means Dalam Perbandingan Tingkat Pengangguran Dengan Tenaga Kerja Berdasarkan Provinsi. *SENARIS*, 2, 24–31.
- [9] Harits, D., Puji, A. A., Andivas, M., Dermawan, D., & Ahmed, E. (2022). Analisis Klaster Tingkat Pengangguran di Provinsi Kalimantan Timur Menggunakan Algoritma K-Means. *Teknika*, 9(2), 456–460.
- [10] Listiani, L., Agustin, Y. H., & Ramdhani, M. Z. (2017). Implementasi algoritma k-means cluster untuk rekomendasi pekerjaan berdasarkan pengelompokan data penduduk. *Sensitif*, 761–769.
- [11] Luchia, N. T., Handayani, H., Hamdi, F. S., Erlangga, D., & Octavia, S. F. (2022). Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 35–41. <https://doi.org/10.57152/malcom.v2i2.422>
- [12] Maulidiah, R., Muchtar, M., Fitri, N. A., Asriani, I., & Yasmine, M. P. (2023). Pengelompokan Data Pertumbuhan dan Kontribusi Ekonomi Indonesia Menurut Provinsi Menggunakan Metode K-Means Clustering. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 6, 436–444.
- [13] Mulyawan, A. R., Gunawan, D., Basri, H., Alfarizi, S., & Ichsan, N. (2023). Penerapan K-Medoids Clustering Dan Silhouette Method Untuk Strategi Pemasaran Program Donasi Pada Lembaga Amil Zakat. *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS: Journal of Information System*, 8(1), 107. <https://doi.org/10.51211/isbi.v8i1.2468>
- [14] Mutmainah, T., Indriyana, S., & Enri, U. (2023). Penerapan Algoritme K-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 5(2), 122–129.
- [15] Naya, C., Siswandi, A., Studi, P., Informatika, T., Teknik, F., Pelita, U., & K-means, A. (2023). IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKAN PENGANGGURAN TERBUKA DI INDONESIA DENGAN METODE CLUSTERING. *Jurnal Teknologi Pelita Bangsa*, 14(2), 99–104.
- [16] Pasaribu, D. F., Damanik, I. S., Irawan, E., & Tambunan, H. S. (2021). Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat. *Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 2(1), 11–20.
- [17] Pastia, N. I., & Dikananda, F. N. (2023). Pengelompokan Data Pengangguran Terbuka Menggunakan Algoritma K-Means Berdasarkan Provinsi Jawa Barat. *Jurnal Dinamika Informatika*, 12(1), 59–69.
- [18] Putrie, A. A., Sanjaya, R., Studi, P., Informasi, S., Adhirajasa, U., & Sanjaya, R. (2021). Pengelompokan Kabupaten / Kota Berdasarkan Indikator Tingkat Pengangguran Menggunakan Algoritma K-Means Clustering. *E-Prosiding Sistem Infomarsi*, 2(2), 111–121.
- [19] Safira, D., Lestari, E. D., Iffa, M., & Annisa, S. (2020). BERDASARKAN ANGKATAN KERJA MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 6(1), 26–31.
- [20] Sihombing, S. C., & Sihombing, D. A. (2021). Pengelompokan Tingkat Kesejahteraan Masyarakat di Sumatera Utara dengan Metode K-Means Clustering. *Jurnal Matematika Integratif*, 17(2), 127–135. <https://doi.org/10.24198/jmi.v17.n2.2021.127-135>
- [21] Sutris, D., Simanjuntak, M., Gunawan, I., & Sari, I. P. (2023). Penerapan Algoritma K-Medoids Untuk Pengelompokan Pengangguran Umur 25 tahun Keatas Di Sumatera Utara. *Jurnal Krisnadana*, 3(1).
- [22] Tanjung, F., Windarto, P. A., & Fauzan, M. (2022). Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, 6, 61–74.
- [23] Wahyudi, M., & Pujiastuti, L. (2020). Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma K-Means Mochamad. *Senaris*, 2, 432–440.

- [24] Wahyuni, A. S., & Utamajaya, J. N. (2022). Penerapan Data Mining Untuk Menentukan Strategi Penjualan Pada Toko Raja Komputer Menggunakan Metode Clustering. *Jurnal Riset Komputer*, 9(2), 281–288. <https://doi.org/10.30865/jurikom.v9i2.4003>