Journal Of Global Computer Science

Volume 1 ; Number 1 ; February 2024 ; Page 24-30
Doi : https://doi.org/10.59435/jgcs.v1i1.2025.24
Web : https://journal.padangtekno.com/index.php/jgcs
E-ISSN : XXXX-XXXX

Implementasi K-Means Clustring Untuk Mengelompokkan Provinsi Di Indonesia Berdasarkan Tingkat Pengangguran Terbuka

Muhammad Muhajir Saddami[1*], Zaehol Fatah[2]

[1.] Teknologi Informasi, Universitas Ibrahimy
[2.] Sistem Informasi, Universitas Ibrahimy
[1*]sadamimuhajir@gmail.com, [2*]zaeholfatah@gmail.com

ABSTRACT

Data mining atau penambangan data merupakan proses pengumpulan dan pengolahan data untuk mengekstrak informasi penting guna menghasilkan wawasan yang lebih dalam. K-Means Clustering digunakan untuk mengelompokkan provinsi di Indonesia berdasarkan Tingkat Pengangguran Terbuka (TPT) dari tahun 2020 hingga 2023 yang diterapkan melalui software RapidMiner. Data dari Badan Pusat Statistik (BPS) diproses melalui beberapa tahap, yaitu pembacaan data menggunakan operator Read Excel, pengelompokan dengan algoritma K-Means, dan evaluasi performa model dengan operator *performance*. Beberapa percobaan dengan berbagai jumlah cluster (1 hingga 3) dilakukan untuk menentukan hasil clustering optimal, menggunakan Davies-Bouldin Index (DBI) sebagai indikator kualitas. Hasil evaluasi menunjukkan pengelompokan terbaik pada dua cluster dengan nilai DBI 0.079, yang mengindikasikan pemisahan cluster yang baik. Visualisasi hasil clustering menggambarkan garis pola TPT tiap provinsi dari tahun ke tahun, di mana cluster 0 dan cluster 1 mencakup provinsi-provinsi dengan pola TPT yang mirip, sedangkan cluster 2 terdiri dari provinsi dengan karakteristik TPT yang berbeda. Analisis ini memberikan wawasan penting bagi pembuat kebijakan untuk memahami dinamika pengangguran di berbagai provinsi dan merancang strategi yang lebih efektif dalam menangani pengangguran.

Keywords: Pengangguran Terbuka, K-Means Clustering, Tingkat Pengangguran Terbuka(TPT)**.**

This Is Open Access Article Under The CC Attribution-ShareAlike 4.0 License.

PENDAHULUAN

Di Indonesia, pengangguran terbuka menjadi salah satu masalah sosial dan ekonomi yang cukup signifikan. Tingginya tingkat pengangguran di Indonesia dapat menghambat pertumbuhan ekonomi serta mempengaruhi kesejahteraan masyarakat. Untuk menangani masalah ini, penting untuk menganalisis pola dan karakteristik pengangguran di berbagai wilayah Indonesia. Pengangguran terbuka sendiri mengacu pada individu yang tidak memiliki pekerjaan, sedang aktif mencari pekerjaan, atau berusaha mendapatkan pekerjaan yang sesuai yang sesuai dengan kualifikasinya[1].

Pengangguran adalah kondisi di mana seseorang yang merupakan bagian dari angkatan kerja aktif mencari pekerjaan dengan tingkat upah tertentu, namun belum berhasil memperoleh pekerjaan yang diharapkan. Masalah pengangguran sangat kompleks karena melibatkan banyak faktor yang berinteraksi secara dinamis dan sulit untuk dipahami sepenuhnya. Di negara-negara berkembang, pengangguran menjadi isu yang semakin kompleks dan serius dibandingkan dengan persoalan ketimpangan distribusi pendapatan, yang sering kali merugikan kelompok berpenghasilan rendah[2].

Masalah pengangguran memiliki dampak yang signifikan terhadap kemiskinan, kriminalitas, dan ketimpangan taraf hidup masyarakat. Oleh karena itu, pemerintah perlu mengambil langkah-langkah antisipatif melalui berbagai kebijakan yang tepat. Pengetahuan memainkan peran penting dalam mendukung pengambilan keputusan dan perumusan kebijakan yang berkaitan dengan pengangguran. Beberapa penelitian telah berusaha untuk menggali data Tingkat Pengangguran Terbuka (TPT) Indonesia guna memperoleh wawasan baru. Data yang ada perlu terus dianalisis untuk menemukan pengetahuan yang lebih dalam. Penelitian ini bertujuan untuk menganalisis data TPT Indonesia dari tahun 2016 hingga 2021. Secara khusus, penelitian ini membandingkan perubahan klaster data TPT antara periode 2016-2018 dengan 2019-2021. Teknik yang digunakan dalam penelitian ini adalah analisis klaster dengan algoritma K-Means. Hasil penelitian menunjukkan bahwa dari analisis clustering terhadap data TPT, Provinsi Riau naik ke cluster 1 (TPT rendah), sementara Provinsi Sumatera Barat turun ke cluster 2 (TPT tinggi)[3].

Menurut data Badan Pusat Statistik (BPS), tingkat pengangguran terbuka di indonesia dari tahun ke tahun menunjukkan variasi yang cukup signifikan antarprovinsi[4]. Data ini memberikan gambaran komprehensif mengenai situasi ketenagakerjaan di setiap wilayah, yang penting untuk di analisis lebih dalam guna memahami tren dan pola distribusi pengangguran. Pengguna data dari BPS yang valid dan terpercaya memberikan dasar yang kuat bagi penelitian ini dalam menilai permasalahan pengangguran di Indonesia.

K-Means Clustering adalah algoritma data mining yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaan antar data. K-Means diaplikasikan untuk menganalisis data penduduk penyandang disabilitas di Jawa Timur, mengelompokkan wilayah berdasarkan jumlah penyandang disabilitas[5]. Seiring dengan kemajuan teknologi dan ketersediaan data yang lebih baik, metode pengolahan data seperti K-Means Clustering menjadi semakin relevan dalam menganalisis fenomena sosial ekonomi. K-Means Clustering memungkinkan pengelompokan provinsi-provinsi di Indonesia berdasarkan tingkat pengangguran terbuka, sehingga pola tersembunyi dapat teridentifikasi. Analisis ini di harapkan memberikan wawasan yang lebih baik tentang distribusi pengangguran di Indonesia serta menjadi dasar yang kuat bagi pengambilan kebijakan untuk mengurangi tingkat pengangguran di masa mendatang[6].

Metode ini berfokus pada penerapan algoritma K-Means Clustering untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan variasi tingkat pengangguran terbuka. K-Means Clustering, sebagai salah satu teknik pengelompokan non-hierarkis, memiliki kemampuan untuk membagi data ke dalam kelompok-kelompok (clusters) yang memiliki karakteristik serupa, berdasarkan parameter tertentu. Dalam konteks penelitian ini, setiap provinsi di Indonesia akan dikelompokkan berdasarkan tingkat pengangguran terbuka, sehingga wilayah-wilayah dengan pola pengangguran yang mirip akan dikelompokkan bersama[7].

Metode K-Means Clustering telah terbukti menjasdi alat yang efektif dalam mengelompokkan wilayah berdasarkan karakteristik sosial-ekonomi, termasuk tingkat pengangguran terbuka di berbagai provinsi. Dalam penelitian yang dilakukan oleh Muthmainnah dkk. (2020), K-Means digunakan untuk mengelompokkan data wilayah dengan tujuan memahami perbedaan karakteristik antarprovinsi secara lebih komprehensif[8]. Penggunaan metode ini memungkinkan pengidentifikasian pola dan tren yang tidak terlihat dengan analisis konvensional, membantu pengambil kebijakan untuk mengatasi perbedaan tingkat pengangguran dengan lebih efisien.

Selain itu, hasil clustering yang diperoleh dapat memberikan wawasan yang berharga untuk merumuskan kebijakan yang tepat guna mengatasi masalah pengangguran, terutama dalam konteks variasi antarprovinsi. Dengan mempertimbangkan perbedaan karakteristik wilayah yang ditemukan melalui proses clustering, hasil penelitian ini dapat berkontribusi dalam mendukung pemerintah dalam menanggulangi masalah pengangguran secara lebih efektif dan komprehensif.

METODE PENELITIAN

2.1 Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia. Data yang digunakan adalah Tingkat Pengangguran Terbuka (TPT) dari tahun 2020 hingga 2023 yang mencakup seluruh provinsi di Indonesia. Data ini dipublikasikan secara resmi oleh BPS dan tersedia secara online melalui situs web resmi BPS. Penggunaan data dari tahun-tahun tersebut memberikan gambaran yang

komprehensif mengenai tren pengangguran selama beberapa tahun terakhir, sehingga dapat digunakan untuk analisis yang lebih mendalam. Pemanfaatan data dari sumber resmi ini penting untuk memastikan validitas dan keandalan hasil penelitian.

2.2 Pengumpulan Data

Proses pengumpulan data dilakukan dengan mengunduh dataset Tingkat Pengangguran Terbuka (TPT) dari tahun 2020 hingga 2023 yang tersedia di situs Badan Pusat Statistik (BPS)[9]. Data yang terkumpul mencakup informasi mengenai tingkat pengangguran di masing-masing provinsi di Indonesia selama periode tersebut. Proses pengunduhan dan pemrosesan data dilakukan secara manual untuk memastikan kelengkapan dan konsistensi data. Data ini digunakan sebagai variabel utama dalam pengelompokan provinsi menggunakan metode K-Means Clustering, yang bertujuan untuk mengidentifikasi pola-pola pengangguran yang tersembunyi di seluruh wilayah Indonesia.

2.3 Pengolahan Data

Pengolahan data dilakukan dengan menggunakan perangkat lunak Rapid Miner, yang memfasilitasi analisis clustering secara efektif tanpa memerlukan perhitungan manual yang kompleks. Tahapan dalam pengolahan data meliputi pembersih data, pemilihan fitur, dan nomalisasi data untuk memastikan hasil clustering yang akurat. Dalam penelitian ini, provinsi-provinsi di Indonesia akan dikelompokkan berdasarkan kemiripan tingkat pengangguran terbuka menggunakan K-Means Clustering[10].

Setelah data diperoses, hasil dari pengelompokan provinsi akan dianalisis lebih lanjut untuk mengidentifikasi pola-pola tersembunyi yang mungkin membantu dalam penyusunan kebijakan ekonomi di Indonesia. Hasil ini diharapkan memberikan wawasan baru terkait pengelompokan wilayah berdasarkan tingkat pengangguran.

2.4 Teknik Data Mining

Data mining digunakan dalam penelitian ini untuk mengekstraksi pola dan informasi penting dari data pengangguran yang tersedia. Teknik yang digunakan dalam penelitian ini adalah Clustering. Khususnya metode K-Means Clustering. Teknik ini berfungsi untuk mengelompokkan provinsi-provinsi di indonesia ke dalam beberapa. Clustering dipilih karena kemampuanya dalam mengidentifikasi kelompok yang memiliki karakteristik serupa tanpa memerlukan label data sebelumnya[11].

2.5 Tahapan Penelitian

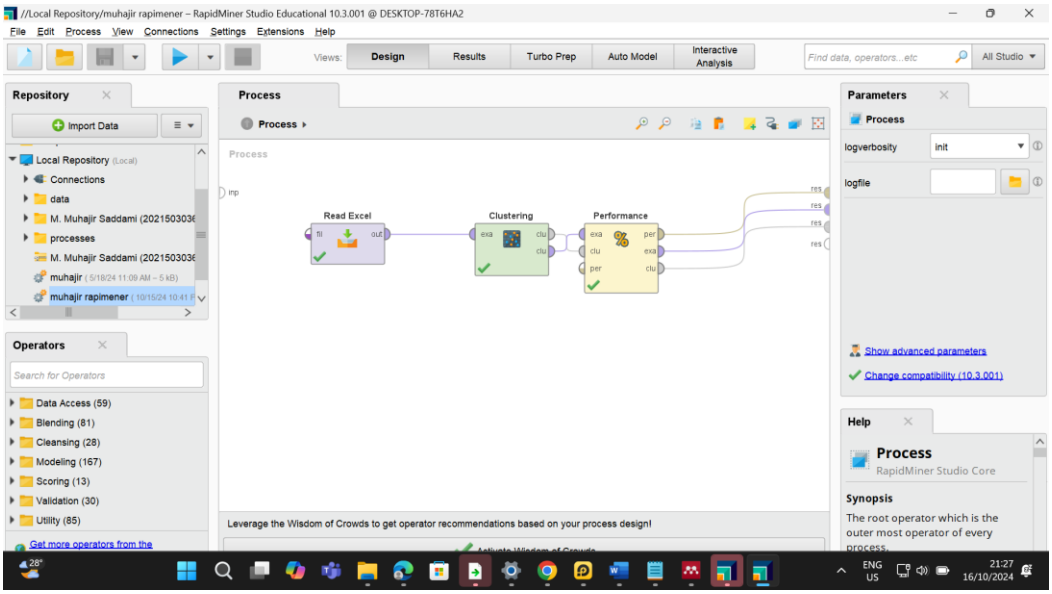
Tahapan penelitian ini terdiri dari beberapa langkah utama sebagai berikut :

1. Pengumpulan Data: data tingkat pengangguran terbuka (TPT) tahun di unduh dari situs resmi badan pusat statistik (BPS). Data yang diambil mencakup tingkat pengangguran di seluruh provinsi di Indonesia pada tahun 2023.
2. Pembersihan Data: Data diperiksa untuk memastikan konsistensi dan kelengkapan. Jika terdapat missing values, maka dilakukan penanganan seperti interpolasi atau penghapusan data yang tidak relevan.
3. Normalisasi Data: Dilakukan normalisasi data untuk menghindari bias dalam pengelompokan, terutama jika terdapat perbedaan skala yang signifikan antara provinsi-provinsi.
4. Clustering: Penerapan metode K-Means Clustering menggunakan software RapidMiner untukmengelompokkan provinsi berdasarkan tingkat pengangguran terbuka. K-Means akan diterapkan dengan menetukan jumlah cluster yang optimal, yang dapat ditentukan menggunakan metode Elbow atau Silhouette Score untuk mendapatkan hasil yang akurat.
5. Interpretasi Hasil: Setelah proses Clustering, hasil dari pengelompokan dianalisis untuk mengidentifikasi pola-pola penting mengenai pengangguran di provinsi-provinsi yang dikelompokkan dalam cluster. Hasil ini digunakan untuk memberikan rekomendasi dalam penyusunan kebijakan pengangguran di Indonesia.

HASIL DAN PEMBAHASAN

Pembuatan model clustering dilakukan dengan menggunakan K-Means Clustring untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan Tingkat Pengangguran Terbuka (TPT) tahun 2020 hingga 2023. Algoritma ini diterapkan melalui software RapidMiner, menggunakan data yang telah diimpor dari laporan BPS. Proses dimulai dengan membaca data menggunakan operator Read Excel dan

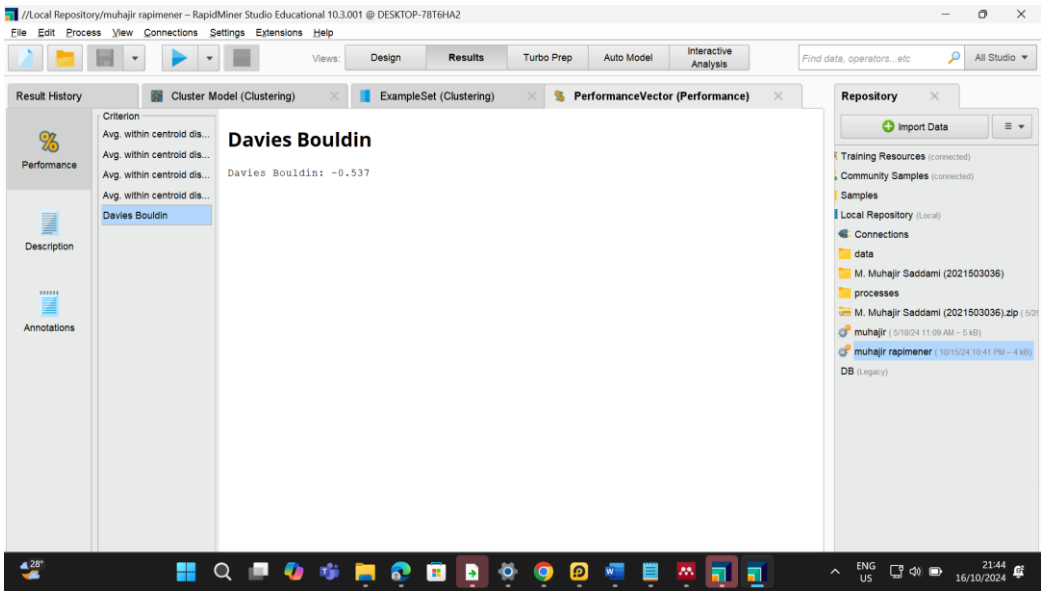
diikuti oleh tahapan clustering untuk mengelompokkan provinsi ke dalam beberapa cluster berdasarkan kesamaan karakteristik tingkat pengangguran di Indonesia.



Gambar 1: Pembuatan Model K-Means Clustering

Gambar 1 menunjukkan proses yang terdiri dari tiga tahapan utama, yaitu pembacaan data (Read Excel), pengelompokan dengan algoritma K-Means, dan evaluasi performa model menggunakan operator Performance. Dalam tahapan pengelompokan, dilakukan percobaan dengan beberapa jumlah cluster (dari 1 hingga 3) untuk mendapatkan hasil cluster yang paling optimal. Evaluasi dilakukan dengan menghitung jarak antar cluster sebagai indikator kualitas pengelompokan. Hasil evaluasi ini kemudian dianalisis untuk mengetahui pola dan perbedaan karakteristik pengangguran antarprovinsi.

3.1 Evaluasi Model



Gambar 2: Hasil dari Proses yang dibuat

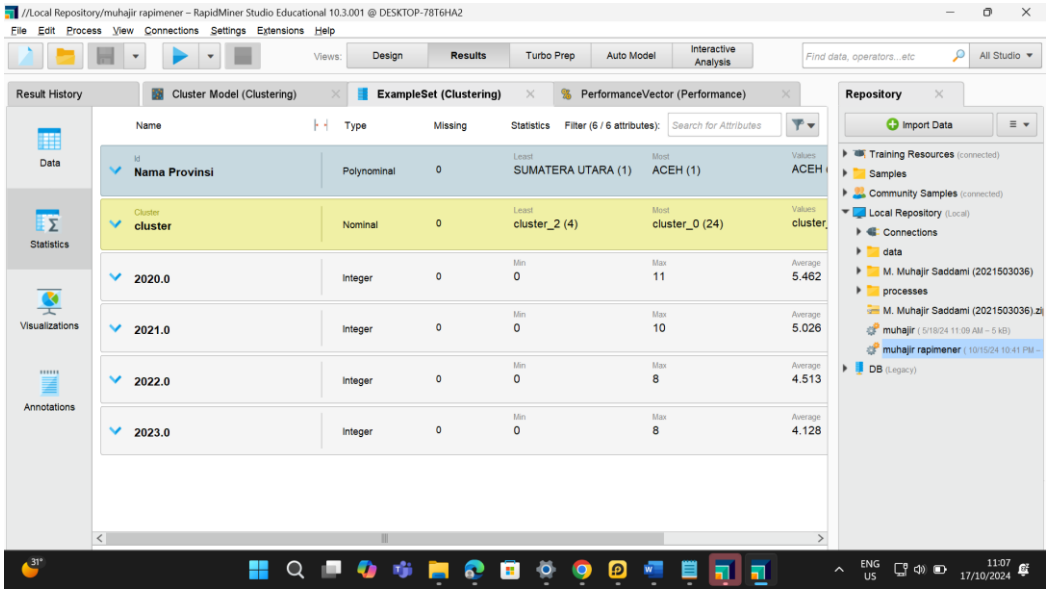
Hasil analisis menunjukkan nilai Davies Bouldin untuk model pengelompokan yang diterapkan. Nilai Davies Bouldin yang diperoleh adalah -0.537. Nilai ini menunjukkan kualitas dari pengelompokan yang dilakukan, dengan semakin kecil nilai ini, menunjukkan semakin baik pemisahan antar cluster. Hasil ini memberikan gambaran tentang seberapa baik provinsi-provinsi dikelompokkan berdasarkan Tingkat Pengangguran Terbuka (TPT) yang dianalisis.

Hasil evaluasi performa model clustering menggunakan Davies-Bouldin Index pada gambar di atas menunjukkan bahwa kualitas pengelompokan bervariasi berdasarkan jumlah cluster yang dipilih. Tabel 1 menampilkan perbandingan nilai Davies-Bouldin untuk jumlah cluster K dari 1 hingga 3.

Table 1: Perbandingan Nilai Davies Bouldin

Banyak Cluster	Nilai Davies Bouldin
1	0.448
2	0.079
3	0.511

Nilai Davies-Bouldin Index mengukur seberapa baik pemisahan antar cluster, di mana semakin rendah nilainya menunjukkan cluster yang lebih baik. Pada model K-Means ini, nilai Davies-Bouldin terbaik diperoleh pada jumlah cluster 2 dengan nilai 0.079, yang menunjukkan bahwa pembagian data ke dalam dua kelompok menghasilkan pemisahan antar cluster yang paling optimal. Sebaliknya, nilai Davies-Bouldin meningkat pada jumlah cluster 3, menandakan penurunan kualitas pengelompokan.

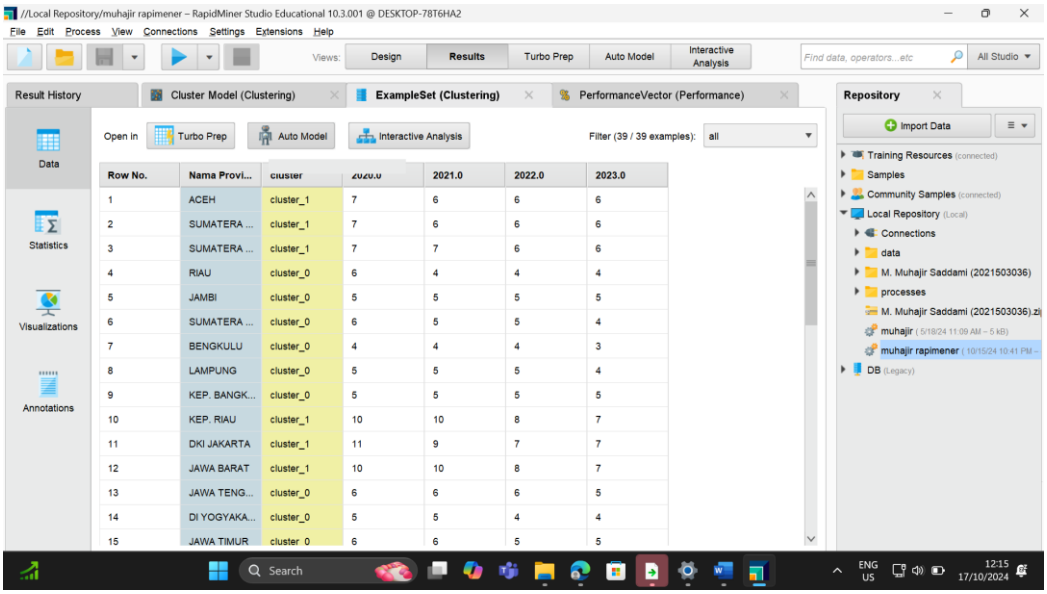


Gambar 3: Hasil Data Pengelompokan Data Provinsi Dengan K-Means

Gambar di atas menunjukkan hasil pengelompokan provinsi di Indonesia berdasarkan Tingkat Pengangguran Terbuka (TPT) untuk tahun 2020 hingga 2023 menggunakan algoritma K-Means dengan jumlah cluster sebanyak 3. Pada bagian "Cluster", terlihat bahwa data provinsi terbagi ke dalam dua cluster utama, yaitu cluster_0 yang terdiri dari 24 provinsi dan cluster_2 yang terdiri dari 4 provinsi. Pengelompokan ini dilakukan berdasarkan data dari tahun 2020 hingga 2023, di mana variabel yang digunakan adalah nilai TPT pada tiap tahunnya.

Nilai minimum dan maksimum untuk masing-masing tahun juga ditampilkan dalam tabel, dengan rata-rata nilai TPT di seluruh provinsi berkisar antara 4.128 hingga 5.462 selama periode 2020 hingga 2023. Dari hasil ini, terlihat bahwa cluster_0 memiliki jumlah provinsi yang lebih banyak dibandingkan cluster_2, menandakan bahwa sebagian besar provinsi memiliki pola yang serupa dalam hal tingkat pengangguran terbuka selama kurun waktu tersebut.

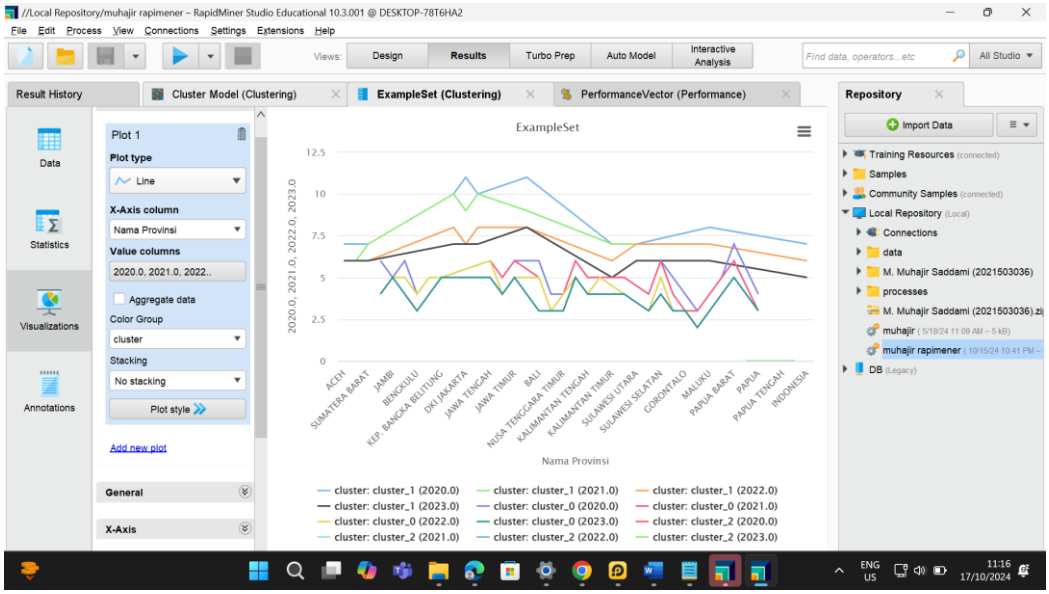
Hasil pengelompokan ini memberikan wawasan yang penting bagi para pembuat kebijakan untuk memahami dinamika pengangguran di Indonesia. Dengan mengetahui bagaimana provinsi dikelompokkan, pemerintah dapat merumuskan strategi yang lebih efektif dalam penanganan pengangguran, terutama di provinsi yang tergabung dalam cluster_2, yang menunjukkan tingkat pengangguran yang lebih tinggi. Analisis ini juga menyoroti perlunya pendekatan yang berbeda dalam penanganan pengangguran di setiap provinsi, mengingat karakteristik dan pola yang berbeda-beda. Selain itu, pentingnya melakukan monitoring dan evaluasi secara berkala terhadap tingkat pengangguran di seluruh provinsi dapat membantu mengidentifikasi tren serta mengantisipasi potensi masalah di masa depan.



Gambar 4: Hasil Pengelompokan Provinsi Dengan K-Means

Gambar di atas menampilkan hasil pengelompokan provinsi di Indonesia berdasarkan Tingkat Pengangguran Terbuka (TPT) untuk tahun 2020 hingga 2023 menggunakan algoritma K-Means. Data ini dikelompokkan ke dalam tiga cluster berbeda, yaitu cluster_0, cluster_1, dan cluster_2. Pengelompokan ini mengkategorikan provinsi-provinsi berdasarkan kesamaan pola pengangguran dari tahun ke tahun. Sebagai contoh, provinsi seperti Aceh, Sumatera Utara, dan Riau dikelompokkan dalam cluster_1, sedangkan provinsi lainnya seperti DKI Jakarta dan Jawa Barat termasuk dalam cluster_0. Pengelompokan ini memberikan gambaran yang lebih jelas mengenai tren pengangguran di berbagai provinsi dan membantu dalam analisis pola pengangguran antar wilayah selama periode tersebut.

3.2 Visualisasi Hasil



Gambar 5: Visualisasi Pengelompokan Provinsi Dengan K-Means

Gambar di atas menampilkan visualisasi garis yang menunjukkan pola pengelompokan Tingkat Pengangguran Terbuka (TPT) dari berbagai provinsi di Indonesia selama periode 2020 hingga 2023, berdasarkan hasil clustering menggunakan algoritma K-Means. Sumbu X merepresentasikan nama provinsi, sedangkan sumbu Y menunjukkan nilai TPT dari tahun 2020 hingga 2023. Setiap garis warna melambangkan cluster yang berbeda, dengan provinsi yang dikelompokkan berdasarkan kesamaan pola TPT. Cluster 0 dan cluster 1 mendominasi provinsi-provinsi dengan tren TPT yang mirip, sementara beberapa provinsi dengan karakteristik pengangguran yang berbeda dikelompokkan dalam cluster 2. Grafik ini membantu memahami fluktuasi TPT di berbagai provinsi dan bagaimana tren pengangguran di masing-masing wilayah berubah dari tahun ke tahun.

KESIMPULAN

Hasil analisis menggunakan K-Means Clustering menunjukkan bahwa pengelompokan provinsi di Indonesia berdasarkan Tingkat Pengangguran Terbuka (TPT) tahun 2020 hingga 2023 dapat menghasilkan wawasan penting mengenai pola pengangguran di berbagai wilayah. Penggunaan dua cluster utama menunjukkan bahwa provinsi-provinsi yang memiliki kesamaan dalam pola TPT dapat dikelompokkan bersama, sedangkan provinsi dengan karakteristik unik dikelompokkan dalam cluster lain. Visualisasi dan pemantauan rutin hasil clustering ini dapat mendukung pembuat kebijakan dalam memahami tren pengangguran dan menetapkan strategi penanganan yang lebih tepat dan efektif. Dengan demikian, analisis berbasis data seperti ini dapat menjadi alat penting dalam perumusan kebijakan pengurangan pengangguran di Indonesia.

DAFTAR PUSTAKA

[1] M. Nurfathullah and I. Purnamasari, "Implementasi K-Means Untuk Mengelompokkan Provinsi Di Indonesia Berdasarkan Indeks Jumlah Pengangguran Terbuka," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2277–2282, 2024, doi: 10.36040/jati.v8i2.9466.

[2] Nabilla Wardah Bonitta and A. H. Primandari, "Analisis Clustering Tingkat Pengangguran Terbuka di Provinsi DIY Tahun 2010-2022 dengan Dynamic Time Warping," *Emerg. Stat. Data Sci. J.*, vol. 2, no. 1, pp. 135–144, 2024, doi: 10.20885/esds.vol2.iss.1.art13.

[3] T. Dan, "3172-7028-1-Pb," vol. 22, no. 01, pp. 89–99, 2022.

[4] "No Title," *data badan Pus. Stat.*, 2024, [Online]. Available: https://www.bps.go.id/id/statistics-table/2/NTQzIzI=/tingkat-pengangguran-terbuka--februari-2024.html

[5] M. B. Uddin and Z. Fatah, "Gudang Jurnal Multidisiplin Ilmu Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas," vol. 2, no. November, pp. 86–94, 2024.

[6] D. Syaputri, P. H. Noprita, and S. Romelah, "Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 1–6, 2021, doi: 10.57152/malcom.v1i1.5.

[7] C. Naya, A. Siswandi, H. Rahendra Herlianto, and A. Fauzi, "Implementasi Data Mining Untuk Pengelompokan Pengangguran Terbuka di Indonesia dengan Metode Clustering," *SIGMA-Jurnal Teknol. Pelita Bangsa*, vol. 14, no. 2, pp. 99–104, 2023.

[8] T. N. Muthmainnah, S. Indriyana, and U. Enri, "Penerapan Algoritme K-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 5, no. 2, p. 122, 2023, doi: 10.36499/jinrpl.v5i2.8736.

[9] J. V. Lenda and M. A. I. Pakereng, "Analisis Tingkat Pengangguran Di Kota Palopo Menggunakan Metode K-Means," *Zo. J. Sist. Inf.*, vol. 6, no. 2, pp. 400–408, 2024, doi: 10.31849/zn.v6i2.20061.

[10] S. Kasus, W. Jawa, W. P. Priyadi, J. D. Irawan, and A. Faisol, "PENERAPAN DATA MINING UNTUK CLUSTERING WILAYAH PRODUKSI PADI MENGGUNAKAN METODE K-MEANS," vol. 8, no. 5, pp. 8381–8388, 2024.

[11] W. An-naziz Safaat, R. Kurniawan, and Y. Arie Wijaya, "Penerapan Data Mining Clustering Menggunakan Algoritma Fuzzy C-Means Pada Data Pencari Kerja Di Kabupaten Kuningan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1507–1511, 2024, doi: 10.36040/jati.v8i2.8411.